Proteomic analysis of neonatal meningitis-causing *Escherichia coli*

Karoline Marie Hennum

This thesis is submitted in partial fulfilment of the requirements for the degree of Master in Medical Biology

Department of Biomedicine and Department of Clinical Science

University of Bergen

September 2017

Acknowledgements

The practical work for this thesis was performed January 2016 – December 2016 at the Gade Research group for Infection and Immunity, Department of Clinical Science, University of Bergen. This work is part of a larger project that was funded by The Research Council of Norway Grant No. 204743. We thank Ulrich Dobrindt of University of Münster for donating the *E. coli* strain IHE3034, and Ørjan Samuelsen of UiT – The Arctic University of Norway for providing the strain J53.

I would like to acknowledge my outstanding supervisors, Professor Harald G. Wiker and Dr Veronika Kuchařová Pettersen. Without them, this thesis would not have become a reality.

I am deeply grateful towards Harald for inviting me to write my thesis at the Gade Research Group for Infection and Immunity. He has been an invaluable resource, always willing to share his knowledge. His feedback and suggestions in the writing process have given this thesis great improvements, while always allowing it to be my own work.

My sincerest thanks go to Veronika for all her help, her expert guidance in the lab, and for teaching me the intricacies of Perseus and MaxQuant. Her encouragement and proficiency has made me work even harder, and helped widen my knowledge about *E. coli*.

To the colleagues in the research group, it has been an absolute joy working with you all, and a special mention goes to Sonja Ljostveit for her practical help and support.

I am grateful towards P. Christoffer Lindemann for his guidance in assessing microbial resistance, and towards Eirik Jovall Nybakken for assisting me with the blood culture work. Olav Mjaavatten at the Proteomic Unit has been of tremendous importance for the experiments relying on LC-MS/MS.

My friends and family, especially my brother Kristoffer, have encouraged me and helped me keep my focus.

And finally, the one who has been my rock through it all, my Ari.

Bergen, September 2017

Karoline Marie Hennum

Table of Contents

Acknowledgements 3
Summary 6
Abbreviations 7
1. Introduction 9
1.1 – *Escherichia coli* – the organism and its characteristics 9
1.1.1 – *E. coli* as a commensal and pathogenic species 9
1.1.2 – *E. coli* genomics 10
1.1.3 – Extraintestinal pathogenic *E. coli* 11
1.2 – Neonatal meningitis-causing *E. coli* 11
1.2.1 – Pathophysiology and infection 12
1.2.2 – NMEC genomics 12
1.2.3 – Other characteristics 13
1.3 – Proteomics 14
1.3.1 – Proteomics as a field in microbiology using current techniques 14
1.3.2 – *E. coli* proteomics 16
1.3.3 – Bioinformatics 16
2. Aims 18
3. Materials and Methods 19
3.1 – Bacterial cell culture 19
3.1.1 – Streaking of bacteria from cold storage 19
3.2 – Phenotypic characteristics assay 20
3.3 – Multilocus sequence typing 20
3.4 – Phylogenetic grouping protocol 23
3.5 – Antibiotic susceptibility testing by disc diffusion 25
3.6 – Growing bacterial cells in blood culture, purification and lysis 27
3.6.1 – Preparation of bacterial samples for MS/MS analysis 27
3.6.2 – Purification of bacterial cells from blood culture 28
3.6.3 – Pre-enzymatic and enzymatic steps in preparation of MS/MS analysis 29
3.6.4 – Stop and Go (STAGE) Extraction Tips 31
3.7 – LC-MS/MS 32
3.8 – Pulsed-Field Gel Electrophoresis 33
3.9 – Investigating the presence of smaller plasmids in H622 35
3.10 – Computational analysis of MS results 35
4. Results 40

4.1 – Phenotypic characteristics 40
4.2 – Multi-locus sequence typing 40
4.3 – Phylogenetic grouping 41
4.4 – Antibiotic susceptibility 41
4.5 – Pulsed-field gel electrophoresis 43
4.6 – Presence of small plasmids 43
4.7 – Mass spectrometry results of bacteria grown in blood cultures and blood agar 45
4.7.1 – Detection of proteins using LC-MS/MS analysis 47
4.7.2 – Quantitative analysis of proteins detected in LC-MS/MS analysis using normalized LFQ intensities 50
4.7.2.1 – Proteins upregulated in the NMEC strains when grown in clinical blood cultures 56
4.7.2.2 – Proteins expressed at a higher level in J53 than in H622 and IHE3034 when grown in clinical blood cultures 63
4.7.2.3 – Analysis of IHE3034 and J53 cultured under different conditions 68
4.7.2.4 – Other MS results 71
5. Discussion 73
5.1 – Phenotypic characterisation and genetic analysis 73
5.2 – Apart fromH622 exhibiting resistance towards ampicillin, the strains were susceptible to all tested antimicrobials 74
5.3 – H622 carries a large plasmid 75
5.4 – Mass spectrometry results of bacteria grown in blood cultures and blood agar 76
5.4.1 – Protein detection and prediction of cellular localization of proteins 76
5.4.2 – Validity and reliability of biological replicas 77
5.4.3 – Virulence factors are found at higher levels in the NMEC strains 78
5.4.4 – Culturing conditions have an impact on protein expression 82
Concluding remarks 83
Future perspectives 83
References 85
Appendix 94

Summary

Meningitis in newborns is a serious infection that causes mortality and neurological injury worldwide. The infection progresses from sepsis, and is dependent on the pathogen being able to cross the blood-brain barrier and invade the spinal fluid. One of the most common gram-negative organisms to cause neonatal meningitis is *Escherichia coli*. In this work, neonatal meningitis-causing *E. coli* strains H622 and IHE3034 were grown in clinical blood cultures alongside commensal *E. coli* K12 derivative J53. The bacteria were purified, lysed, and digested with proteases, before being analysed using mass spectrometry. The mass spectrometry results were quantified in order to create quantitative protein profiles of each bacterium. By using statistical and computational analysis, we compared the strains and identified proteins that were differentially expressed between the pathogenic strains and J53. The results indicate that the pathogenic strains share a number of regulatory mechanism, and demonstrate a higher expression than J53 of virulence factors, motility proteins, and proteins involved in capsule synthesis. In addition to the mass spectrometric analysis, the bacteria were characterised using genetic and phenotypic methods. The results indicate that the pathogenic strains H622 and IHE3034 share a closer evolutionary relationship than either does with J53.

The mass spectrometry raw files have been deposited to the ProteomeXchange Consortium with the dataset identifier PXD005779.

Abbreviations

2DE	Two-dimensional electrophoresis
ACN	Acetonitrile
AmBic	Ammonium Bicarbonate
ANOVA	Analysis of variance
APEC	Avian pathogenic *E. coli*
AST	Antimicrobial Susceptibility Testing
ATCC	American Type Culture Collection
BBB	Blood-brain barrier
BLAST	Basic Local Alignment Search Tool
BMEC	Brain microvascular endothelial cell
Bp	Base pairs
CSF	Cerebrospinal fluid
Da	Daltons
DTT	Dithiothreitol
EPEC	Enteropathogenic *E. coli*
EUCAST	The European Committee on Antimicrobial Susceptibility Testing
ExPEC	Extraintestinal pathogenic *E. coli*
FA	Formic acid
FDR	False Discovery Rate
Kb	Kilobase
LFQ	Label-Free Quantification
MALDI-TOF	Matrix-assisted laser desorption/ionization time-of-flight
MIC	Minimum inhibitory concentration
Min	Minutes
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NeuNAc	N-acetylneuraminic acid
NMEC	Neonatal Meningitis-causing *E. coli*
HUH	Haukeland University Hospital
HPLC	High-Performance Liquid Chromatography

IPEC	Intestinal pathogenic *E. coli*
kb	Kilobase, one thousand basepairs
LC	Liquid chromatography
MCPs	Methyl-accepting chemotaxis proteins
MID	Department of Microbiology
MLST	Multilocus sequence typing
MNEC	Meningitis-associated *E. coli*
NMEC	Neonatal meningitis-causing *E. coli*
MS	Mass spectrometry
PFGE	Pulsed-field gel electrophoresis
PCR	Polymerase Chain Reaction
RCF	Relative Centrifugal Force, G-force
RT	Room temperature
SDS	Sodium Dodecyl Sulphate
SEPEC	Septicaemia-associated *E. coli*
Sec	Seconds
ST	sequence type
TBS	Tris buffered saline
TCA	Tricarboxylic acid
UniProt	The Universal Protein Resource
UniProtKB	UniProt KnowledgeBase
UPEC	Uropathogenic *E. coli*
VESPA	Visual Evaluation and Statistics to Promote Annotation

1. Introduction

1.1 – *Escherichia coli* – the organism and its characteristics

Escherichia coli is a gram-negative, rod-shaped bacterium, with peritrichous flagella to propel it for movement, and short, "hair-like" pili for attachment to different substrates [1, 2]. It was first described by Theodor Escherich, a German paediatrician and microbiologist in 1884, while he was conducting research on the bacteria found in the gut of infants [3, 4]. A common place to find the bacterium is in faeces, as it colonizes the lower intestine of virtually all mammals and a number of other warm-blooded animals. This bacterium is a facultative anaerobe; able to utilize oxygen in aerobic respiration, but also perfectly capable of growing without it using mixed-acid fermentation. Many traits have made it a favourite among researchers, and the preferred bacterial work-horse in molecular biology. It grows quickly on many different substrates at 37 °C, and is generally easy to cultivate. As a member of the phylum *Proteobacteria,* class *Gammaproteobacteria,* and the family *Enterobacteriaceae, E. coli* is closely related to well-known pathogens such as *Klebsiella* and *Salmonella* [5].

1.1.1 – *E. coli* as a commensal and pathogenic species

Despite its more notorious peers, *E. coli* is most commonly found in humans as a commensal bacterium, hitching a ride in the intestinal tract. However, the relationship is mutually beneficial, as the bacterium produces vitamin B_{12} and vitamin K, which humans are unable to synthesize for themselves [6, 7]. In addition, by colonizing parts of the gut as a "nice" bacterium, it can keep pathogenic microorganisms at bay, simply by taking valuable space on the intestinal lining [1].

In the gut, microorganisms are subject to a myriad of physical forces pushing them around. *E. coli* is able to counteract this by using long, helical flagella moved by rotating motors [2]. They can swim in one direction by bundling the filaments and rotating them counterclockwise in a coordinated fashion. Change of direction - tumble - can be achieved quickly by rotating one or more motors in a clockwise direction [8]. Chemotaxis, the process where a bacterium moves towards or away from a chemical gradient is highly reliant on the bacterium's ability to change direction or to keep moving. The gradient is sensed by transmembrane chemoreceptors, methyl-accepting chemotaxis proteins (MCPs), and relayed to the flagellar motor through Che- and Fli-proteins [9]. Motility also enables the bacterium to create biofilms: A bacterial community of bacteria, colonizing surfaces by enmeshing themselves in a network of adhesins, polymers, and extracellular network proteins [10]. A biofilm enables the bacteria to better withstand stress, such as environmental fluxes and antibiotics. The cells in the deeper layers of a biofilm are more sheltered from the host's immune system. *E. coli* is a very versatile bacterium, and –

keeping in tune with the family relations – also include several strains with a pathogenic lifestyle [11]. Pathogenic *E. coli* can be divided into two main groups, on the basis of where the infection is found: intestinal pathogenic *E. coli* (IPEC), and extraintestinal pathogenic *E. coli* (ExPEC). These groups are further subdivided into pathotypes. A pathotype is defined as "a group of strains of a single species with certain pathogenic traits" [11]. The most commonly found pathotypes of ExPEC are *E. coli* causing urinary tract infections; uropathogenic *E. coli* (UPEC), bacteria found in sepsis (blood poisoning); septicemia-associated *E. coli* (SEPEC), *E. coli* creating illness in birds; avian pathogenic *E. coli* (APEC), and meningitis-associated *E. coli* (MNEC) [12, 13].

1.1.2 – *E. coli* genomics

Escherichia coli is a bacterium that includes a diverse array of strains, and consequently has a very broad genome. It is therefore useful to differentiate between different categories of genomes. The all-inclusive *pan-genome* describes all genes that have been discovered and sequenced in bacteria of a specific species. The pan-genome can be further divided into the *core genome,* which encompasses genes that all strains within a species carry, and the *accessory*, or *dispensable* genome (also known as the *flexible* genome), which entails genes that are found only in some strains [14, 15]. The dispensable genes lend attributes that are not strictly necessary for survival, but gives increased fitness under certain conditions and in niches, such as antibiotic resistance and virulence factors. The entire pan-genome of *E. coli* now includes upwards of 16 000 genes [1]. With such a wide variety in the genome, it is a waste of resources for one cell to carry the entire pan-genome. An average *E. coli* genome contains 5000 genes on a circular chromosome. Depending on the strain, about two thirds of these are shared with all other strains, while one third is part of the accessory genome, and only found in a few, or none, other strains [16]. Distantly related strains may share as little as 40 % of their genes [1]. One of the most common ways of categorizing *E. coli* strains have been through serotyping. Using this method, three specific surface antigens are used to determine the serotype of the strain; the outer core of the lipopolysaccharide layer, the capsular antigen, and proteins from the flagella (O:K:H) [17]. This has historically been the most used method, and is useful in vaccine design and epidemiological research. Techniques such as multilocus sequence typing (MLST), which is based on determining the different alleles of housekeeping genes, may be considered more reliable when studying evolutionary relationships, as housekeeping genes are part of the core genome, and surface antigens are encoded by the accessory genome [14].

1.1.3 – Extraintestinal pathogenic *E. coli*

As already established, pathogenic *E. coli* can be categorized in many different subtypes. The extraintestinal pathogenic *E. coli*(ExPEC), share the common core genome with all other *E. coli* strains, but have a number of accessory genes that make them well adapted to life as an infectious organism, known as virulence factors. These accessory genes enable the bacterium to infect the host and avoid the immune system, and are considered virulence factors. Examples of virulence factor genes often found in ExPEC strains include machinery for iron uptake, adhesins for attachment to cellular surfaces, capsules around the bacteria made of polysaccharides, and lipopolysaccharide layers [11, 18-20]. They can be encoded on the main chromosome or carried on extrachromosomal plasmids. Often, they are part of mobile genetic elements, or have been at some time in the evolutionary past [13]. Being mobile, these genes can be transferred between bacteria, creating new strains with a different combination of traits. Thus, elucidating the evolutionary relationship between pathogenic *E. coli* is dependent upon robust genetic methods, and cannot rely on a set of specific virulence genes [17]. The current methods for assigning *E. coli* strains to phylogenetic groups, are MLST and quadruplex phylotyping [21, 22]. The phylogenetic lineages currently consist of eight defined phylo-groups: Groups A, B1, B2, C, D, E, F, which are all a part of the *E. coli sensu stricto*, while the eight group, *Escherichia* cryptic clade I, is not [21]. ExPEC strains are more likely to be members of phylogroups B2 or D, than members of A or B1 [23, 24]. According to data obtained via MLST, phylogroup B2 is one of the oldest linages in the *E. coli* phylogenetic tree [25]. Like commensal, non-pathogenic strains, ExPEC strains are capable of colonizing the gut without causing illness. When they are transmitted to another part of the body, like the urinary tract, however, they possess the attributes necessary to invade and cause damage to human cells. Genetic analyses have shown that the most prominent sequence types (STs) among ExPEC isolates are 69, 73, 95, and 131 [26]. One ExPEC type that is capable of inflicting long-term damage, due to the site of infection and the particular vulnerability of the patients, is neonatal meningitis-causing *E. coli* (NMEC).

1.2 – Neonatal meningitis-causing *E. coli*

The spinal fluid circulates around and within the central nervous system, where the brain is surrounded by tough membranes known as the meninges: The *dura mater,* the arachnoid membrane, and the *pia mater.* When bacteria are isolated from the spinal fluid, it is known as an inflammation of the meninges – meningitis. Neonates are especially vulnerable to infection. Preterm neonates, or infants with a low birth weight, even more so. Preterm neonates are especially vulnerable due to an insufficiently developed blood-brain barrier (BBB), and an

underdeveloped immune system. Infants diagnosed with bacterial meningitis are at risk of death and neurological sequelae. After group B *Streptococcus, E. coli* is the second most common cause of bacterial meningitis in newborns [27]. *E. coli* causes approximately 20 - 25% of cases of bacterial meningitis in newborns [28, 29]. Of diagnosed cases, approximately 15% of patients die, while 30 - 50% suffer disabilities such as intellectual developmental disorders, epilepsy, and long-term problems with vision and hearing [29-35]. A study carried out in the United Kingdom and Republic of Ireland found that the incidence of neonatal meningitis caused by *E. coli* was 0.04 out of 1000 live births [27].

1.2.1 – Pathophysiology and infection

Before meningitis, the patient usually suffers bacteraemia, and a high number of colony-forming units in the blood are necessary for the bacteria to cross the BBB and invade the cerebrospinal fluid [13, 36, 37]. The infection starts by the bacteria colonizing a mucosal surface such as the nasopharynx or gastrointestinal tract, after which they cross the epithelial cell layer and enter the bloodstream [36, 38]. The BBB is weakened, and the bacteria induce accumulation rearrangement of actin cytoskeleton within the brain microvascular endothelial cells (BMECs) [13, 36, 39]. This leads to transcytosis of the bacteria in a closed vacuole through a "zipper-like mechanism". Proteins OmpA and CNF1, along with the K1 capsular antigen, have been demonstrated as being vital for attachment and invasion of BMECs [13, 36, 39]. As the bacterium usually starts in the blood, bacteremia is detected through blood samples using clinical blood cultures. Blood from the patient is drawn in specially made culture bottles containing nutrient broth and beads that neutralize antibiotics. The blood culture is incubated at 37 °C, and analysed for the presence of bacteria. In order to diagnose meningitis, a sample of CSF is acquired through a lumbar puncture. The spinal fluid is analysed, looking for low glucose levels and white blood cells. It is also cultivated, so that any bacteria present may grow and be identified through PCR. It has been suggested that a source for NMEC is faecal transmission of bacteria from healthy carriers, which can be transmitted from the mother to the infant during vaginal delivery [37]. Bacterial neonatal meningitis is treated with antibiotics and supportive care.

1.2.2 – NMEC genomics

The majority of NMEC belong to phylo-group B2, the same evolutionary group as a majority of ExPEC strains found in bacteremia and urinary tract infections [37, 40, 41]. When analysing a large set of NMEC strains, Bonacorsi and Bingen found that 68% of strains belonged to group B2, 25% to groups D and B1, with the smallest subset of 7% belonging to group A [27].

Wijetunge et al. found a similar distribution of NMEC phylogroup belonging [37]. They also found no statistical difference between NMEC and *E. coli* donated from faecal samples of healthy individuals in regards to belonging to group D. Serotype analysis have shown that NMEC strains are likely to be of oligoclonal origin [27]. It is not unusual for NMEC strains to carry large plasmids containing genes increasing their fitness in a host environment [42]. Genes found on 11 plasmids isolated from 9 NMEC strains were shown to carry many genes for iron uptake, resistance towards host immunity, and genes for the maintenance of the plasmid itself [42].

1.2.3 – Other characteristics

E. coli isolated from the spinal fluid of neonates share many characteristics. More than 80 % have the K1 capsular antigen, which is considered a virulence factor [13, 31, 37, 40, 43]. Many human cells have glycans – sugar conjugates – on their surface. These glycoconjugates are important in many processes, including cell-cell recognition. Sialic acids refer to a wide group of nine-carbon sugar acids that are often found in the terminal position of glycoconjugates on the surface of human cells, exposed to the environment [44]. Many bacteria have evolved to exploit this fact, and include a layer of sialic acid in their outer capsule, which can aid in avoiding detection by the innate immune system. *E. coli* synthesize a large number of capsular polysaccharide antigens, K antigens. The K1 capsular antigen is one the most thoroughly studied. Bacteria that are serotype K1 synthesize N-acetylneuraminic acid (NeuNAc) in a linear, α-2,8-linked homopolymer [45]. In addition to being anti-immunogenic, the K1 capsule confers some resistance towards the complement system, has antiphagocytic properties, and assists the bacteria in crossing the BBB [45, 46]. The main genes important for K1 polysialic acid synthesis are *neu*- and *kps*-genes of the kps gene cluster [47]. In a host environment, there is little free iron available, and iron acquisition systems such as *iroBCDN* and *iutA* increase the bacterium's fitness, and have been found in several NMEC strains [42, 48]. The bacterium may also use the host's heme proteins by expressing the hemin receptor molecule ChuA, giving it a an efficient way of obtaining this essential nutrient [49]. NMEC strains have also shown a higher frequency of adhesion genes, along with an increased ability to create biofilm when compared to non-pathogenic *E. coli* [24, 37]. Pathogenic *E. coli* also often secrete toxins such as heat-stable enterotoxin a and b, Shiga toxin, and cytolethal distending toxin [13].

Despite sharing many traits, it is not yet possible to categorize an *E. coli* strain as an NMEC based solely on analysis of the genome, as NMEC strains are varied in what virulence genes they possess [40]. However, recent research demonstrates that "virulence genotyping and phylogrouping may assist in defining the potential NMEC pathotype" [37].

1.3 – Proteomics

Genomics and proteomics are inseparably linked; without genes, no protein expression. However, it has long been clear that we cannot fully understand the proteins, their functions and expression solely through studying their genes of origin. And so, proteomics, the study of proteins come into play. Since the term was first introduced in 1995, the scientific field of proteomics has grown exponentially [50]. As the name implies, proteomics denotes the large-scale study of proteins, from 3D modelling of single proteins, to whole proteome analysis of organelles and cells. The field has branched into many sub-fields, such as the study of post-translational modifications, protein-protein interactions, structural and functional proteomics, and protein expression profiling [50]. All have an important role in biology and often demands the mixing of tools and expertise from different fields such as genetics and bioinformatics. The synergistic combination of genomics and proteomics has been aptly named proteogenomics, and is quickly leading to better annotation of sequenced genomes [51].

1.3.1 – Proteomics as a field in microbiology using current techniques

In microbiology, as in all the biological fields, proteomics has a large role to play. Proteins are, after all, how a microorganism interacts with its surroundings. Whether one studies pathogenic bacteria – host interactions or commensal microorganisms on plants, protein studies can be vital for one's work. Genomics and transcriptomics can yield valuable information, but do not show post-translational modifications, protein turnover, or relative levels of protein expression.

Microorganisms are extremely good at adapting to their environment and changing which and how much proteins they express. The object of one's study may be one single protein, or one can attempt to tie together information about all the proteins expressed by the organism. As protein expression can be regulated on the transcriptional, translational, and protein level, one cannot use only genetics to study this. One of the fields in proteomics is the study of protein expression levels – this involves research into the quantitative amounts of expressed proteins. Studying this relies heavily upon integrating earlier work in identifying proteins and their functions, and demands highly sensitive methods to detect thousands of proteins at once. One can create a "protein profile" for a bacterium. This is an overview of identified proteins, usually including their relative levels. A widely used method for large-scale protein analysis has for a long time been two-dimensional electrophoresis (2DE), where the proteins are separated based on isoelectric point and mass [52]. Other widely used analysis techniques is matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF), a very versatile method that can be used for bacterial identification, provided that the profile of the particular strain is present in a database [53].

Today, another method has taken centre stage; liquid chromatography-tandem mass spectrometry: LC-MS/MS. This method combines two powerful analysis techniques to create a very sensitive and broad detection method. LC, or high-performance liquid chromatography (HPLC) is used to separate the different substances suspended in a mixture, creating a fractionation of the entire sample. This keeps it from entering the mass spectrometer "all at once". The separation is achieved by pushing the sample through a column, where the substances, in this case peptides, are held back to a varying degree. The separation creates a precise and orderly elution of the substances, which can then be sent into the mass spectrometer. The mass spectrometric analysis is where the detection and measuring of the peptides' *m/z* takes place. Put simply, the mass spectrometer is able to measure the charge and the mass of a sample. It does this by ionizing the sample and turning it into a gas phase, whereupon it injects the ionized gas into a detection chamber. Depending on the ration between the ions' mass and charge (*m/z* ratio), the ions travel to different parts of the detector. The mass spectrometer employed in this thesis is an Orbitrap produced by Thermo Scientific with a nano-electrospray ion source. An electrospray ionisation uses a strong electrical current to charge the sample before it is sprayed in a thin mist and the solvent evaporated. This method of ionisation is well suited for the study of biological molecules, as the energy used on each molecule is relatively weak, and creates little fragmentation. Strong electrical fields are used to guide and accelerate the charged sample to the detection chamber. Orbitrap is an ion trap mass analyser based around the principle of confining ions in an oscillating trajectory using an electric field created by ring-shaped electrodes. The ions are ejected from the trapping cell "in order of increasing m/z ratio", and detected in the ion detector [54, 55]. The instrument also records the number of ions of a particular *m/z* ration per time unit, usually counts per second, with this particular ratio. The counts per time unit reflects the abundancy of this particular peptide, and is referred to as "intensity". With newer technology, tandem mass spectrometry – MS/MS – was introduced. Here, ions of interest are analysed more thoroughly by being analysed in sequence: The ion of interest, denoted the precursor ion, is analysed once, then fragmented through collision with a neutral gas, and the fragment – the product ion – undergoes MS analysis again [56]. This is known as collision-induced dissociation. This way, the peptide is analysed twice. The first analysis is very precise, and the second somewhat less. MS/MS is still more exact in identifying peptides than non-tandem MS. Results of MS analysis are given as mass spectrums, a "graphical display of the relative abundance of ion signals against the *m/z* ratios" [55]. The most intense ion is set as 100 % abundance, and the other ions acquire their abundance relative to this [55].

Limitations of MS analysis include a high sensitivity towards contaminants. The method is non-discriminatory and highly sensitive, so all non-peptide materials that make their way into the sample can give disturbances in the final spectra. Common contaminants are keratin from skin and hair, and residual chemicals from laboratory equipment. Proper technique, high-quality solvents and a clean working environment are all necessary when handling samples for MS analysis [57].

1.3.2 – *E. coli* proteomics

As a widely used lab organism and a common human pathogen, there have been many proteomic studies of *E. coli.* There are thousands of studies into the proteins of *E. coli*, and it has become one of the most intensively studied prokaryotes to date. From 3D-modelling to sequence analysis and comparison, these studies vary in aims, theme, and methodology. The entire *E. coli* genome has been sequenced, and is well annotated. There are entire databases devoted solely to proteins and genes detected in *E. coli.* A review by Han and Lee considers the proteome studies to have moved from gel-based approaches to predictive proteomics using bioinformatic tools [58]. They divide today's research into *E. coli* proteomics into "proteomics for biology and proteomics for biotechnology". This divides the research based on the intention of the studies; if it is done with the intention of improving current knowledge about the biological function of the bacteria and proteins, or if it is an attempt at improving production and synthesis of bioproducts.

The work done here falls into the category of "proteomics for biology". Earlier studies in this category include comparative studies growing the bacteria under different conditions, assessment of genome coverage using MS, and comparison with transcriptomic data [59-61]. These studies usually identify a number of unique proteins ranging from 2600 to 2800, depending on how many strains and growth conditions the study included. This constitutes an approximate 60 % coverage of the proteins of an annotated genome [60]. Typically, more proteins are identified with an increasing number of strains and conditions. A noteworthy study executed at Gades Research Group for Infection and Immunity included mass spectrometric analysis and comparison of several *E. coli* strains grown under aerobic and anaerobic conditions [59]. Pettersen et al. used LC-MS/MS to create protein profiles of six pathogenic *E. coli* strains and compared relative levels of protein expression in aerobic and anaerobic conditions. This thesis follows the same analytical methods and employs the same underlying principles.

1.3.3 – Bioinformatics

MS/MS analysis yields enormous amounts of data. The *m/z* spectra are used to identify specific peptide sequences by matching against six-frame translations of chosen genomes. The peptides

can in turn be mapped to proteins with known sequences. In this thesis, this work is done in the program MaxQuant, developed by Jürgen Cox and colleagues at the Max Planck Institute of München, Germany [62]. This program employs the search algorithm Andromeda to search for and identify the peptides [63]. MaxQuant maps the sequences of detected peptides, and uses the levels of these peptides to determine the detected level of the identified protein. As it is possible, in fact probable, that the levels of proteins vary between samples due to small differences in handling and analysis, a normalization of protein levels is necessary. In order to normalize the levels of proteins, a large group of proteins that vary minimally between the samples is used as a standard [64]. This is known as Label-Free Quantification (LFQ), as it uses no labels, radioactive or otherwise. LFQ uses the signal strength and number of observations of frequently detected peptides to create a relative scale. This is used along with an absolute scale of summed-up peptide intensities to assign peptides new, normalized intensities; LFQ intensities. The LFQ algorithm is integrated into the MaxQuant search engine, and creates two distinctive data outputs: the samples without normalized levels, and the same samples with levels corrected by LFQ [64]. The resulting data output can be further analysed in Perseus, a proteomic analysis program with a range of statistical analysis tools available [65].

When the proteins have been identified, there are several large databases dedicated to proteomic information available. The Universal Protein Resource (UniProt) is one of the decidedly largest, and the UniProt KnowledgeBase (UniProtKB) contains millions of protein sequences and annotated proteomes of more than 50 thousand species. The database is curated and constantly kept up to date [66]. EcoCyc is another database that will be used comprehensively throughout this work. It is a database dedicated to the *E. coli* K12 MG1655, containing experimental results, proteomic data, and known biochemical pathways [67]. EcoCyc is a member of the broader spanning BioCyc, a collection of species-specific databases dedicated to genomes and metabolic pathways [68].

2. Aims

The main aim of this thesis is to compare the proteomic expression of *E. coli* strains isolated from neonatal meningitis to that of a derivative of a commensal *E. coli* using mass spectrometry. This includes identifying the proteins that differ, what cellular processes they are involved in, and the magnitude of difference in expression.

Sub-aims

- Genetic and phenotypic characterisation of *E. coli* strain H622
- Determine the evolutionary relations between the *E. coli* strains used in this study

3. Materials and Methods

Bacterial strains

E. coli H622	– Isolated at Haukeland University Hospital (Not published)
E. coli IHE3034	– Kindly donated by Ulrich Dobrindt (University of Münster, Germany)
E. coli J53	– Kindly donated by Ørjan Samuelsen, (UiT The Arctic University of Norway).
E. coli BL21(DE3)pLysS	– Commercial product (Promega)
E. coli ATCC 25922	– Commercial product (American Type Culture Collection (ATCC))

H622 was isolated from a case of neonatal meningitis at Haukeland University Hospital in 2012. The strain has not been genome sequenced, nor has it undergone extensive analysis until now.

IHE3034 (O18:K1:H7) has been genome sequenced (GenBank accession no. CP001969), and belongs to multilocus sequence type 95 (ST95) and the EcoR B2 group [20].

The *E. coli* strain J53 (F^- *met pro*) is a derivative of *E. coli* K12 and has been genome sequenced, but the genome has not been assembled. The whole genome shotgun sequencing project is available in 42 contigs (GenBank accession no. AICK00000000). It exhibits resistance towards sodium azide [69].

3.1 – Bacterial cell culture

All work with living bacteria was performed in a Biosafety Level 2 (BSL-2) laboratory, in a disinfected class II laminar air flow (LAF) cabinet using gloves and sterile techniques, in accordance with UiB's safety regulations. For storage, the bacteria were kept at -70 °C in Greaves' solution (manufactured by the Substrate lab, Department of Microbiology (MID), 4th floor of the Laboratory Building, HUH).

3.1.1 – Streaking of bacteria from cold storage

Materials:

- Agar plates (All agar plates used in this thesis were manufactured by the Substrate lab, MID, 4th floor of the Laboratory Building, HUH)
    - Blood agar plate
    - Lactose agar plate (Corresponds to MacConkey agar)

Bacteria were collected from -70 °C and kept on wet ice. Tubes and plates were only opened inside the LAF bench to keep them from being contaminated. Each strain was streaked using

dilution streaking and a sterile 1 µl inoculation loop on blood and lactose agar plates. The plates were incubated at 37 ° overnight (16 – 18 hrs).

3.2 – Phenotypic characteristics assay

A phenotypic characteristics assay is used to determine key metabolic traits of a microorganism, such as motility and the ability to utilize certain nutrients.

The assay was carried out on strains H622, IHE3034 and J53. It was done at the MID at Haukeland University Hospital according to their standardized protocols.

3.3 – Multilocus sequence typing

Multilocus sequence typing (MLST) is a genetically based method of bacterial strain determination and categorization based on variations in the sequence of housekeeping genes. Housekeeping genes are involved in basic cellular functions, and are constituents in all strains of a bacterial species. These genes are quite stable, and occasional mutation in these genes create different versions, known as alleles. In MLST, short stretches of 450-500 base pairs (bp) 7 housekeeping genes are amplified using PCR and sequenced with Sanger sequencing to determine the allele version. The housekeeping genes and their products can be seen in *Table 1*. The combination of allele versions is a unique "fingerprint" for the strain, and the basis for determining the multilocus sequence type (ST). Determining the ST is important to characterise a bacterial strain, and to elucidate the evolutionary relations between different strains [22, 70].

Table 1 – Seven housekeeping genes used in multilocus sequence typing of E. coli*, the primer sequences, and the PCR product size [71, 72].*

Target	Primer	Primer sequence	Product size (bp)	Gene product
adk	adk-P1	5′-ATTCTGCTTGGCGCTCCGGG-3′	583	Adenylate kinase
	adk-P2	5′-CCGTCAACTTTCGCGTATTT-3′		
fumC	fumC-P1	5′-TCACAGGTCGCCAGCGCTTC-3′	806	Fumarase hydratase
	fumC-P2	5′-GTACGCAGCGAAAAAGATTC-3′		
gyrB	gyrB-P1	5′-TCGGCGACACGGATGACGGC-3′	911	DNA gyrase
	gyrB-P2	5′-ATCAGGCCTTCACGCGCATC-3′		
icd	icd-P1	5′-ATGGAAAGTAAAG TAGTTGTTCCGGCACA-3′	878	Isocitrate/ isopropylmalate dehydrogenase
	icd-P2	5′-GGACGCAGCAGGATCTGTT-3′		
mdh	mdh-P1	5′-ATGAAAGTCGCAG TCCTCGGCGCTGCTGGCGG-3′	932	Malate dehydrogenase
	mdh-P2	5′-TTAACGAACTCCT GCCCCAGAGCGATATCTTTCTT-3′		
purA	purA-P1	5′-CGCGCTGATGAAAGAGATGA-3′	816	Adenylosuccinate dehydrogenase
	purA-P2	5′-CATACGGTAAGCCACGCAGA-3′		

recA	recAF1	5′-ACCTTTGTAGCTGTACCACG-3´	780	ATP/GTP binding motif
	recAR1	5′-AGCGTGAAGGTAAAACCTGTG-3´		

Materials:

- Agar plates
  - Blood agar plate
  - Lactose agar plate
- Bacterial strains
  - IHE3034
  - H622
- $_dH_2O$ (Nuclease-free water, Promega)
- GoTaq Green Master Mix, 2X (Promega)
- Primers as described in Table 1, 10 µM (Custom made by Sigma-Aldrich)
- Tris, acetic acid and EDTA (TAE) buffer, pH 8.15 (Substrate lab, MID, 4th floor of the Laboratory Building, HUH)
- Agarose, analytical grade (Sigma-Aldrich)
- Gel loading dye (Sigma-Aldrich)
- GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific)
- ExoSAP-IT PCR Product Cleanup (Affymetrix)
- Gelred Nucleic Acid Gel Stain, 10 000X in water (Biotium)
- Milli-Q water, Type 1 H_2O (Purification system: Milli-Q, Merck)
- BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific)
- Sequencing buffer: BigDye Sequencing buffer 5X (Thermo Fisher Scientific)

Method

3.2.1 Amplification of housekeeping genes used in MLST via PCR

The bacterial strains IHE3034 and H622 were streaked on to blood agar plates and lactose agar plates and grown over night. An isolated colony was picked from the blood agar plate using a sterile 1 µl inoculation loop and transferred to an eppendorf tube with 50 µl $_dH_2O$. Lysis of the bacteria and pelleting of cell debris was achieved by boiling at 99°C for 15 minutes before centrifuging at 16 000 g for 20 minutes. For the PCR amplification, 5 µl of the supernatant from each strain was used along with primers and GoTaq Green Master Mix 2X PCR mix. The following volumes describe the PCR mix per strain. The given concentration is for the starting solutions.

Component	Volume
Forward primer, 10 µM	1 µl
Reverse primer, 10 µM	1 µl
GoTaq Green, 2X	25 µl
DNA	5 µl
$_dH2O$	18 µl

Total volume: **50 µl**

The PCR tubes were run in a thermal PCR cycler on the program described in Table 2

Table 2 – PCR program used in the amplification of seven housekeeping genes. (min: minutes, sec: seconds)

Step	Purpose	Temperature	Time	Number of cycles
1	Initialization step	94 °C	5 min	1
2	Denaturation	94 °C	45 sec	30
3	Annealing	58 °C	45 sec	
4	Elongation	72 °C	45 sec	
5	Final elongation	72 °C	7 min	1
6	Final hold	4 °C	∞	1

After PCR, the samples were stored at 4 °C over night.

3.2.2 Gel electrophoresis of PCR products

A 2 % (w/v) agarose gel was created by heating 3 gr agarose and 150 ml TAE buffer, and the liquid agar cast in a gel mould. To weight the sample and facilitate loading, 1 µl gel loading dye was mixed with 5 µl PCR product. The gel was submerged in TAE buffer, and the samples loaded along with a 1 kb DNA marker, before the gel was run at 120 V for 60 minutes. In order to visualize the bands, the gel was soaked in a solution of 50 µl GelRed in 100 ml MilliQ H_2O for 40 minutes using mild agitation. The gel was rinsed and destained in Milli-Q water 5 minutes, 3 times, with a change of water per destaining. The gel was visualized in a UV transilluminator.

3.2.3 Clean-up of PCR products

The resulting DNA was cleaned; single nucleotides and single-stranded DNA was removed, by using ExoSAP-IT; 5 µl PCR product was combined with 2 µl ExoSAP-IT and incubated at 37 °C for 15 minutes before inactivation of the ExoSAP-IT at 80 °C for 15 minutes.

3.2.3 Sanger sequencing and analysis

This procedure is based upon the one created by the Sequencing Facility at the University of Bergen [73]. In order to obtain precise sequencing results, the sanger dideoxy sequencing was carried out using both the reverse and the forward primer for all the seven genes. A Qubit Fluorometric Quantitation (Thermo Fisher Scientific) fluorometer was used to determine DNA concentration of the samples. The same primers were used as for the previous amplification,

and the primers were diluted to a concentration of 3.2 µM using $_dH_2O$. The samples were sequenced with forward and reverse primers separately, giving a total of 14 samples per bacterial strain. The sequencing mix described below was run in a PCR thermal cycler at the program described in Table 3. The given concentration is for the solutions before they are mixed. DNA implies inactivated PCR and ExoSAP-IT mix:

Component	Volume
BigDye v.3.1	1 µl
Sequencing buffer	1 µl
DNA	3 µl
Primer, 3.2 µM	1 µl
dH2O	4 µl
Total volume:	**10 µl**

Table 3 – PCR thermal cycler program for the sequencing of seven amplified housekeeping genes used to determine sequence type. (min: minutes, sec: seconds) [73].

Step	Purpose	Temperature	Time	Number of cycles
1	Initialization step	96 °C	5 min	1
2	Denaturation	96 °C	10 sec	25
3	Annealing	55 °C	5 sec	
4	Elongation	60 °C	4 min	
5	Final hold	4 °C	∞	1

When the cycler was finished, 10 µl $_dH_2O$ was added to each sample before they were sent for analysis at the sequencing unit at the University of Bergen [73]. The MLST database of University of Warwick was used to determine the allele version of each gene, and the ST [74].

3.4 – Phylogenetic grouping protocol

E. coli can be grouped and characterised in many ways. An important aspect of *E. coli* research is the categorization of different strains and determination of evolutionary relations. In 2000, Clermont et al. developed a phylo-typing method known as the triplex PCR method, which could group strains into subgroups A, B1, B2, or D. This was based on the presence or absence of three PCR products from the genes *chuA* and *yjaA*, and a DNA fragment known as TspE4.C2 [75]. All primers are mixed in with the sample before cycling, and the presence of DNA fragments of specific sizes is considered a positive (+) or a negative (-) presence of the gene. The pattern of which genes are present form the basis of phylo-typing the bacterial strain.

Extensive research has led to the establishing of a total of 8 phylo-groups of *E. coli*, of which seven (A, B1, B2, C, D, E, F) belong to the *E. coli sensu stricto*, and one is known as *Escherichia* clade I. Clermont et al. has developed a new method based upon the triplex PCR, with the inclusion of a fragment from the gene *arpA* [21]. This makes the method a quadruplex PCR method, and further refines its ability to discriminate between groups (Table 4).

Table 4 – Primer sequences and product size for the primers used in phylogenetic grouping [21, 75].

Target	Primer	Primer sequence	Product size (bp)
arpA	AceK.f	5′-AACGCTATTCGCCAGCTTGC-3′	400
	ArpA1.r	5′-TCTCCCCATACCGTACGCTA-3′	
chuA	ChuA.1	5´-GACGAACCA ACGGTCAGGAT-3´	279
	ChuA.2	5´-TGCCGCCAGTACC AAAGACA-3´	
yjaA	YjaA.1	5´-TGAAGTGTCAGGAGACGCT G-3´	211
	YjaA.2	5´-ATGGAGAATGCGTTCCTCAAC-3´	
TspE4.C2	TspE4C2.1	5´-GAGTAATGTCGGGGCATTCA-3´	152
	TspE4C2.2	5´-CGCGCCAACAAAGTATTACG-3´	

Materials

- Agar plates
  - Blood agar plate
  - Lactose agar plate
- Bacterial strains
  - J53
  - IHE3034
  - H622
- $_dH_2O$ (Nuclease-free water)
- GoTaq Green Master Mix, 2X
- Primers as described in Table 1, 10 μM (Custom made by Sigma-Aldrich)
- Tris, acetic acid and EDTA (TAE) buffer, pH 8.15
- Agarose, analytical grade
- 100 bp DNA Ladder (N32315, New England Biolabs)
- Gelred Nucleic Acid Gel Stain, 10 000X in water
- Milli-Q water, Type 1 H_2O

Method

The bacterial strains H622, IHE3034 and J53 were grown and lysed as described in section 3.2.1. From this sample, 5 μl of the supernatant from the sample (DNA) was transferred to a sterile eppendorf tube and mixed with primers, GoTaq Green Master Mix 2X, and nuclease-

free water as described below. The given concentration is for the solutions before they are mixed.

Component	Volume
Primer AceK.f, 10 µM	1 µl
Primer ArpA1.r, 10 µM	1 µl
Primer ChuA.1, 10 µM	1 µl
Primer ChuA.2, 10 µM	1 µl
Primer YjaA.1, 10 µM	1 µl
Primer YjaA.2, 10 µM	1 µl
Primer TspE4C2.1, 10 µM	1 µl
Primer TspE4C2.2, 10 µM	1 µl
GoTaq Green, 2X	25 µl
DNA	5 µl
$_{d}$H2O	12 µl
Total volume:	**50 µl**

The samples were amplified in a PCR thermal cycler using the program described in Table 5.

Table 5 – Table describing the PCR thermal cycler program used in the amplification of DNA fragments for use in phylo-typing of E. coli.

Step	Purpose	Temperature	Time	Number of cycles
1	Initialization step	94 °C	5 min	1
2	Denaturation	94 °C	45 sec	30
3	Annealing	58 °C	45 sec	
4	Elongation	72 °C	45 sec	
5	Final elongation	72 °C	7 min	1
6	Final hold	4 °C	∞	1

A 2 % (w/v) agarose gel was created as described earlier, and submerged in TAE buffer. The samples were loaded along with a 100 bp ladder and run at 120 V for 60 minutes. The gel was stained in GelRed, destained and visualized as described earlier.

3.5 – Antibiotic susceptibility testing by disc diffusion

A standard method for determining the antibiotic susceptibility of a strain is by disc diffusion [76, 77]. A paper disc saturated with a known quantity of antibiotics is placed on the surface of an agar plate inoculated with a bacterium, and the antibiotic diffuses through the agar. After incubation, an area with no growth around the disc can be seen if the bacteria have responded. The size of the zone of inhibition represents the *in vitro* efficacy of the tested drug. The

European Committee on Antimicrobial Susceptibility Testing (EUCAST) have standardized the methods and resistance determinants, including breakpoint tables for antibiotics and microorganisms [78, 79]. Bacterial resistance levels are based on the diameter of the inhibitory zone. Using clinical breakpoint tables devised for species and antimicrobial agent, one can categorize the bacteria as susceptible (S), intermediate resistant (I), or resistant (R) [80]. These categories describe the probability of therapeutic success when using the antibiotic to treat patients [78].

Materials:

- Agar plates (Substrate lab, MID, 4th floor of the Laboratory Building, HUH)
  - Blood agar plate
  - Lactose agar plate
  - Mueller-Hinton agar
- Antibiotics
  - MIC test strip (Liofilchem): Gradient:
    - Ampicillin 0.016 – 256 µg/ml
    - Cefotaxime 0.016 – 256 µg/ml
    - Ceftriaxone 0.016 – 256 µg/ml
    - Ciprofloxacin 0.002 – 32 µg/ml
    - Gentamicin 0.016 – 256 µg/ml
  - Discs (Oxoid): Antibiotic per disc:
    - Ampicillin 10 µg
    - Cefotaxime 5 µg
    - Gentamicin 10 µg
    - Ciprofloxacin 5 µg
    - Cefuroxime 30 µg
    - Amoxicillin-clavulanate 20/10 µg
    - Ceftazidime 10 µg/disc
    - Ertapenem 10 µg/disc
    - Meropenem 10 µg/disc
    - Imipenem 10 µg/disc
    - Trimethoprim-sulfamethoxazole 1.25/23.75 µg/disc
    - Piperacillin-tazobactam 30/6 µg/disc

Method

E. coli ATCC 25922 was chosen as quality control strain in accordance with EUCAST guidelines [81]. The bacteria were brought from – 70 °C storage, streaked onto blood agar and lactose agar plates and incubated overnight at 37 °C. Colonies were picked from the blood agar plate and suspended in physiological saline (NaCl 0.09 % (w/v)) to a density of 0.5 McFarland (± 0.05). Within 15 minutes, a sterile cotton wad was dipped in the solution and pressed against the side of the tube to remove excess inoculum. To create an even carpet growth of bacteria, an automatic rotator was used to spin the Mueller-Hinton agar plates and the cotton wad moved

from the edge of the plate to the centre. A maximum of 6 discs or 2 strips were applied to each agar plate before 15 minutes passed since streaking. All plates were incubated at 37 °C for 18 hours. No more than four plates were placed in each stack.

Reading the plates:

Discs: Inhibition zone diameter was measured using a slide calliper on the back of the plate, using the naked eye and held against a black background approximately 30 cm from the eyes. If the zone was divided by a weaker line, giving an inner and outer zone, the inner border was registered.

Strips: MIC was determined by finding where the zone of inhibition fell on the strip. If this point fell between two MIC concentrations, the most stringent concentration was chosen. I.e., if the point fell between 8 and 6, one would round up to 8.

3.6 – Growing bacterial cells in blood culture, purification and lysis

Before LC-MS/MS analysis, the bacteria were grown in clinical blood culture, lysed, and the proteins digested using proteases.

3.6.1 – Preparation of bacterial samples for MS/MS analysis

Materials

- Agar plates
  - Blood agar plate
  - Lactose agar plate
- Bacterial strains
  - J53
  - IHE3034
  - H622
- Tris-buffered saline (TBS) (Substrate lab, MID, 4th floor of the Laboratory Building, HUH)
- Blood lysis buffer
  - 0,6 % (w/v) Brij 97 (Polyoxyethylene (10) oleyl ether) (Sigma-Aldrich), 0.4 M CAPS (3-(Cyclohexylamino)-1-propanesulfonic acid) (Sigma-Aldrich), pH 11.7
    Filtered through a sterile 0.2 µm pore filter
- Bacterial washing buffer
  - 20 mM sodium phosphate dibasic (ICN Biomedicals), 0.05 % (w/v) Brij 97, 0.45 % (w/v) NaCl (Sigma-Aldrich), ph 7.2
    Filtered through a sterile 0.2 µm pore filter
- Bacterial extraction buffer
  - 2,5 % SDS (w/v) (Invitrogen), 10 mM Tris-HCl (Sigma-Aldrich)

The bacteria were streaked on blood agar plates and incubated overnight at 37 °C. A healthy volunteer donated 10 ml blood to each of six clinical blood culture bottles (aerobic BacT/ALERT FA Plus (Biomérieux #410851). These clinical blood cultures are the ones used

in hospital to screen for bacteremia. Biological replicas were created by inoculating each strain in two separate blood culture bottles. For the inoculum to be created, room temperature (RT) physiological saline was used. Three well-isolated, medium-sized colonies were picked and completely suspended in 3 ml saline. The blood cultures were inoculated with 0.5 ml of the suspension, before being incubated in the automated blood culture system BacT/ALERT 3D (Biomérieux) at 36 °C for 14 hours. When taken out of the incubator, the blood cultures sat at RT for approx. 2 hrs, and at 4 °C for 2 hrs, before treatment.

3.6.2 – Purification of bacterial cells from blood culture

As it is only the bacterial proteins that are of interest, it is important to extract and rinse the bacteria from the blood culture before lysis. This way, one lessens the risk of human proteins contaminating the sample.

The blood lysis buffer lyses human blood cells, but leaves bacterial cells intact. That way, the sample can be centrifuged, and the microorganisms collected. From each clinical blood culture, 30 ml of blood broth was taken. The broth was lysed by mixing 30 ml blood culture broth and 15 ml blood lysis buffer. The solution was incubated at RT for 5 minutes, before being centrifuged at 1800 RCF. Post centrifugation, the supernatant was removed and the pellet inspected. The pellet in all tubes had red discolouration, and was resuspended in 5 ml blood lysis buffer, incubated for 5 minutes and centrifuged at 4600 rpm in a Heraeus Multifuge 3SR with a swing-out rotor 4 place. A red-coloured edge could be seen in the top of the pellet, this was gently removed with a sterile pipette. The bacterial pellet was washed by resuspension in 10 ml washing buffer, and centrifuged at the same conditions for 4 minutes. This was done a total of 3 times. Following the washing step, the pellet was resuspended in 10 ml tris-buffered saline (TBS) and centrifuged at the same conditions. Post centrifugation, the supernatant was removed, and the process repeated. The pelleted bacterial cells were resuspended in 1 ml TBS and transferred to the eppendorf tube before being centrifuged for 2 minutes 18 000 RCF. The supernatant was removed, and the pellet frozen over night at -70 °C.

Extraction of bacterial proteins for MS/MS analysis

The samples were kept on wet ice when not handled.

The pellet was resuspended in 750 µl bacterial extraction buffer. The solution was transferred to a 2 ml screw cap tube (FastPrep Lysing Matrix A (#116910050, MP Biomedicals)) and the bacterial cells were disrupted in a ribolyser (FastPrep-24, QBiogene Inc.) for 60 seconds at maximum speed (6,5 m/s). The samples were incubated on ice for five minutes, before centrifugation for 45 minutes at 4 °C and 10 000 RCF. To clear the supernatant of particulates,

300 µl of the supernatant was transferred to a clean eppendorf tube and centrifuged in an eppendorf centrifuge for 5 minutes at 4 °C and 20 000 RCF. Direct Detect (EMD Millipore) was used to determine the protein concentration of the supernatant. From this point onwards, all samples were treated in triplicate. At the end of the procedure, two of the replicas were sent to the MS analysis, while a third was kept at – 70 °C, as a backup.

3.6.3 – Pre-enzymatic and enzymatic steps in preparation of MS/MS analysis

The proteomic sample preparation was done according to the Multi-Enzyme Digestion, Filter-Aided Sample Preparation (MED-FASP) procedures, as described by Wiśniewski et al. [82, 83]. In this procedure, the samples are digested using two proteases consecutively. The FASP procedure is carried out within the confines of a membrane, which acts as a miniature reactor for the digestion. The membrane retains larger molecules, such as DNA and undigested proteins, while letting peptides and smaller contaminants through. This enables the digestion of proteins directly on top of the membrane. By using the FASP method, one bypasses the in-solution or in-gel digestion methods usually used for digestion of proteins before MS-analysis. This method has also been demonstrated as very robust when dealing with smaller amounts of protein. It typically converts approximately 50 % of the available proteins to peptides, and facilitates easy rinsing of the samples to remove detergents [82, 83].

In bottom-up proteomic experiments such as this, the goal of the enzymatic digestion is to reduce the proteins to peptides that are small enough to ionize and analyse via MS. In addition to digesting the proteins on a membrane, they were digested using two enzymes. Done correctly, this creates a broader coverage of digested peptides than digestion with only one enzyme, or digestion with several enzymes separately [83]. The use of different enzymes creates peptides cut in different places, and a more precise identification of proteins using overlapping peptide sequences. According to Wiśniewski et al., the choice of enzymes that gave the highest number of unique identifications of peptides were digestion using LysC followed by digestion with trypsin. The LysC used here is an endoproteinase that cleaves at lysine residues, including those followed by a proline. Trypsin is a serine endoproteinase with cleavage action on the C-terminal side of arginine and lysine residues.

All work pertaining enzymes and preparation for enzymatic digestion were done in a clean and semi-closed cabinet, protected from atmospheric dust and contamination. All reagents were of MS-grade.

Materials

- Microcon®-30 centrifugal filter units with a molecular cutoff of 30 000 Dalton (EMD Millipore)
- Chromasolv MS grade H_2O (Sigma-Aldrich)
- DTT-AmBic solution
    - 10 mM Dithiothreitol (DTT) (Sigma-Aldrich), 100 mM ammonium bicarbonate (Sigma-Aldrich)
      Kept on wet ice
- UA solution
    - 8 M urea (Sigma-Aldrich), 10 mM HEPES (Sigma-Aldrich), pH 8
- IAA solution
    - 0.05 M iodoacetamide in UA solution
      Kept dark at all times
- Enzymes
    - LysC
        - Lysyl endopeptidase, MS grade, 20 µg (Wako, Catalogue # 121-05063)
    - Trypsin
        - Pierce trypsin protease, MS grade, 20 µg (Thermo Fisher Scientific, Catalogue # 1862748)
- Tris-HCl (Sigma-Aldrich)

Denaturation of proteins

In order to achieve efficient enzymatic cleavage of proteins into peptides, the proteins were denatured using sodium dodecyl sulphate (SDS) and dithiothreitol (DTT). SDS has strong denaturing effects due to its numerous negative charges. DTT reduces disulphide bonds, and keeps the protein from reforming its secondary structure. As the digestive enzymes are unable to cleave proteins that are tightly folded, denaturing is used to cause the proteins to lose their tertiary structures, thus exposing restriction sites for the enzymes to cut. Strong denaturing conditions also render the proteins stable, and keep them from refolding. Thiol groups were alkylated to prevent reformation of disulphide bonds by adding iodoacetamide.

The FASP method was followed [82], using Microcon®-30 centrifugal filter units with a molecular cutoff of 30 000 Dalton.

For each sample, 50 µg was used, and diluted to 50 µl with Chromasolv MS grade water. To reduce the disulphide bonds and raise the pH, 5 µl DTT-AmBic solution was added and the samples incubated for 45 minutes at 56 °C with mixing of 350 rpm. The spin filters were assembled and recieved 100 µl UA solution before being spun at 14 000 RCF for five minutes at RT. The protein sample received 200 µl UA solution, was transferred to the spin filter and centrifuged at 14 000 g for 15 minutes. Another 200 µl UA solution was added to the filter before centrifuging at the same conditions, and the flow-through discarded. The filter unit received 100 µl IAA solution and was incubated in darkness for 20 minutes with 300 rpm

shaking at RT. The samples were centrifuged at 14 000 g for 10 minutes. 100 µl UA solution was added, and the samples were centrifuged at 14 000 g for 15 minutes. This was repeated twice.

Enzymatic digestion of protein samples using LysC

Each filter unit received 100 µl 50 mM Tris-HCl and centrifuged at 14 000 g for 10 minutes. This was done a total of three times. The collection tube was changed to a fresh one. To resuspend the lyophilized enzyme, 1020 µl 50 mM Tris-HCl pH 8 was added and the solution vortexed. 50 µl enzyme suspension was added to each filter unit and the samples were placed in a thermo mixer for 60 seconds at 600 rpm. The samples were transferred to an incubator along a bowl of water to keep the humidity high. They were incubated at 30 °C for 13 hours. The samples were centrifuged at 14 000 g for 15 minutes. 50 µl MS grade H_2O was added, and the samples centrifuged using the same conditions. This was repeated once. The filter units with the remaining proteins were transferred to new collection tubes and placed on wet ice. The collection tubes containing eluate were placed in a SpeedVac (Eppendorf Concentrator Plus), and run at 30 °C until approx. 20 µl solution was left. The samples were treated using STAGE tips, see section 3.6.4.

Enzymatic digestion of protein samples using trypsin

The trypsin enzyme was resuspended using 1020 µl 40 mM AmBic and vortexed before each filter received 50 µl of the resuspended enzyme. The samples were mixed in a thermo mixer for 60 seconds at 600 rpm. The filter units were incubated on wet ice for 2.5 hours. The same wet chamber was used as previously, including a bowl of water to keep the humidity high. The samples were incubated at 37 °C for 12 hours. In order to collect the peptides, the samples were centrifuged at 14 000 g for 15 minutes. The remaining peptides were eluted by adding 50 µl H_2O and centrifuging for 15 minutes at 14 000 g. This step was repeated once. The collection tubes were placed in a SpeedVac, and run at 30 °C until approx. 20 µl solution was left. The samples were treated using STAGE tips, see section 3.6.4.

3.6.4 – Stop and Go (STAGE) Extraction Tips

Materials:

- Micropipette tip, 2-200 µl (Eppendorf)
- Cutting tool
- Syringe, 1 ml
- C_{18} Empore filters (Varian, 3M)
- LoBind tubes for proteins (Eppendorf)
- Acetonitrile, LC-MS Chromasolv (Fluka) (ACN)

- Formic acid (Sigma-Aldrich) (FA)
    - 0.1 % FA, diluted in Chromasolv H_2O (v/v)
    - 0.1 % FA in 80 % ACN, diluted with Chromasolv H_2O (v/v)
- Methanol, Chromasolv (Sigma-Aldrich)

Method

The method of Rappsilber et al. [84] was followed to create micro tips with stacked filter material for cleaning peptides. The membrane consists of reverse-phase beads immobilized in a polytetrafluoroethylene mesh. The chromatographic beads are able to bind the peptides, which can be washed of all contaminants using the correct solvents, before being eluted. Small, circular pieces of membrane can be cut out and stacked in micropipette tips using an appropriate cutting tool, like a blunt-ended syringe. Each circular disc is able to hold 2 to 4 µg of peptide [84]. Because of the tapering of the micropipette tip, the membrane disks are immobilized and do not fall through. This makes them able to withstand relatively high pressure. Exploiting this, an air-filled syringe with the correct adapter is used to push the solutions through the stacked discs. A high dilution of formic acid (FA) is used to wash away contaminants. Acetonitrile (ACN) is a solvent used to elute the peptides from the STAGE tip.

Using the cutting tool, 6 discs of C_{18} Empore filters were cut out and gently packed into a 2-200 µl micropipette tip. The following was added and subsequently pushed through the filters, using the syringe to create air pressure:

i. 20 µl methanol
ii. 20 µl 0.1 % FA in 80 % ACN
iii. 20 µl 0.1 % FA

(i., ii., and iii.: Pushed through and sent in waste)

iv. Sample
v. 20 µl 0.1 % FA – Pushed into waste
vi. 40 µl 0.1 % FA in 80 % ACN, and the eluate collected into LoBind tubes.

The samples were dried in a SpeedVac at 30 °C, and stored at -70 °C. Before analysis, the peptides were resuspended 0.1% formic acid and 2% ACN.

3.7 – LC-MS/MS

The LC-MS/MS analysis was done at the Proteomics Unit at the University of Bergen in the exact manner as described in the paper by Pettersen et al. [59]:

"The MS/MS analysis was carried out at the Proteomics Unit at the University of Bergen on an Ultimate 3000 RSLC system (Thermo Scientific, Waltham, MA) connected to a LTQ Orbitrap

mass spectrometer (Thermo Scientific) equipped with a nanoelectrospray ion source. Briefly, 0.5–1 µg protein was loaded onto a preconcentration column (Acclaim PepMap 100, 2 cm × 75 µm ID nanoViper column, packed with 3 µm C18 beads) at a flow rate of 5 µl/min for 5 min using an isocratic flow of 0.1% TFA [trifluoroacetic acid] (v/v) with 2% ACN (v/v). Peptides were separated during a biphasic ACN gradient from two nanoflow UPLC pumps (flow rate of 270 nl/min) on the analytical column (Acclaim PepMap 100, 50 cm × 75 µm ID nanoViper column, packed with 3 µm C18 beads). Solvent A and B was 0.1% TFA acid (v/v) in water or ACN, respectively. Separated peptides were sprayed directly into the MS instrument during a 195 min LC run with the following gradient composition: 0–5 min 5% B, 5–6 min 5–7% B, 6–135 min 7–32% B, 135–145 min 32—% B, 145–150 min 40–90% B. Elution of very hydrophobic peptides and conditioning of the column was performed by isocratic elution with 90% B (150–170 min) and 5% B (175–195 min), respectively. Desolvation and charge production were accomplished by a nanospray Flex ion source.

The mass spectrometer was operated in data-dependent-acquisition mode to automatically switch between Orbitrap-MS and LTQMS/ MS acquisition. Survey of full-scan MS spectra (from *m/z* 300 to 2000) were acquired in the Orbitrap with resolution of r = 240,000 at *m/z* 400 (after accumulation to a target of 1,000,000 charges in the LTQ). The method used allowed sequential isolation of the most intense ions (up to 12, depending on signal intensity) for fragmentation on the linear ion trap using collisionally induced dissociation at a target value of 10,000 charges. Target ions already selected for MS/MS were dynamically excluded for 40s. General MS conditions were as follows: electrospray voltage, 1.8 kV; no sheath; and auxiliary gas flow. Ion selection threshold was 3000 counts for MS/MS, and an activation Q-value of 0.25 and activation time of 10 ms was also applied for MS/MS."

3.8 – Pulsed-Field Gel Electrophoresis

Pulsed-field gel electrophoresis (PFGE) is a method for the separation of larger DNA fragments. While smaller fragments are easily separated by regular gel electrophoresis, large DNA fragments (> 20 kb) clump together and form one cohesive band. In PFGE, the direction of the current is varied, and switches between straight down, perpendicular to the axis. The constantly changing current creates a much better separation of larger DNA fragments, such as large plasmids [85]. Large plasmids are sensitive towards fragmentation, and so the bacteria are

cast inside gel plugs. Immobilized in the agarose gel plugs, lysis of the cells and digestion of the DNA using nuclease can take place without risking the DNA clumping or being torn apart.

Materials:

- TE buffer
    - 10 mM Tris, 1 mM EDTA, pH 8
- TBE (Tris, borate, EDTA) buffer (Substrate lab, MID, 4th floor of the Laboratory Building, HUH)
- SeaKem Gold agarose (Lonza)
- Proteinase K, 20 mg/ml (Thermo Fisher Scientific)
- Cell Suspension Buffer
    - 100 mM Tris, 100 mM EDTA, pH 8
- S1 endonuclease (100 U/µl) (Thermo Fisher Scientific, Catalogue # EN0321)
- Restriction buffer for S1 endonuclease (5x) (Thermo Fisher Scientific)
- Chromasolv H_2O (Sigma-Aldrich)
- GelRed (Biotium)

The equipment used for running the PFGE was a CHEF-DR III (Bio-Rad), assembled according to the manufacturer's instructions.

Method

The PFGE was carried out by following the Standard Operating Procedure for PFGE of *E. coli* O157:H7 and non-O157 [86].

The settings for the running of the gel were the following:

- The cooling device was set to 14 °C
- 2 L 0.5x TBE buffer was poured into the cooling device
- The flow rate of the pump was set to 1 L / minute
- The pump was turned on, and (volume) 0.5x TBE buffer was added
- The cooling device was run for 30 minutes before the gel was submerged in the holder
- The following settings were used:
    - Initial switch time 1 seconds
    - Final switch time 25 seconds
    - Runtime 24 hours
    - Angle 120°
    - Gradient 6V/cm
    - Temperature 14 °C
    - Ramping factor linear

The gel was stained using intercalating dye (GelRed) and visualized in a UV transilluminator.

3.9 – Investigating the presence of smaller plasmids in H622

In order to determine whether the bacterial strain harbours any smaller plasmids, a plasmid extraction was carried out. An *E. coli* strain with a known presence of a small plasmid was used as positive control. Size and presence of DNA was verified by running an agarose gel electrophoresis.

Materials:

- Kit: QIAGEN Plasmid Midi Kit (Cat No./ID: 12143)
- Bacterial strains:
    - H622
    - BL21(DE3)pLysS
        - An *E. coli strain* containing the plasmid pLysS (approximately 4.8 kb).
- S1 endonuclease (100 U/µl) (Thermo Fisher Scientific, # EN0321)
- Restriction buffer for S1 endonuclease (5x) (Thermo Fisher Scientific)
- Agarose (SeaKem)
- TAE buffer (Substrate lab, MID, 4th floor of the Laboratory Building, HUH)
- DNA marker: EZ load 1 – 15 kb (BioRad)
- GelRed (Biotium)

Method:

The procedure to extract plasmids was carried out according to the instructions included in the kit. Post extraction, the DNA was cut using an S1 endonuclease. A negative control containing TE buffer in place of DNA was run alongside the samples. A Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific) was used to determine DNA concentration in the samples.

Recipe for enzymatic digestion:

S1 nuclease: 1 µl (5U/µl)
S1 buffer (5x): 4 µl
DNA: 200 ng
dH_2O ad 20 µl

The reaction was run for 1.5 hours on 37 °C.

Agarose gel electrophoresis of samples and staining in GelRed

A 0.8 % (w/v) agarose gel in TAE buffer was used, and 5 µl of EZ load 1 – 15 kb was loaded next to the samples. The gel was run at 120 V for 1 hr 10 min. It was stained in a GelRed solution (120 µl in 500 ml) for 2 hours and visualized in a UV transilluminator.

3.10 – Computational analysis of MS results

Software:

- Chromas Lite, v2.1, Developed by Technelysium
- MaxQuant, v1.5.5.1, J. Cox et al. [62]

- Perseus, v1.5.6.0, Tyanova et al. [65]
- Excel, various versions (Microsoft Office)

Procedure:

Protein detection using MaxQuant

The raw files from the LC-MS/MS were analysed using MaxQuant version 1.5.5.1 [62] and the integrated Andromeda search engine [63] with the following databases: Reviewed *E. coli* proteins were downloaded from the Swiss-Prot section of UniProtKB (28,709 entries, downloaded on the 15th September 2016). MaxQuant was used to create a six-frame translation of the following genomes: NMEC strain IHE3034 (NCBI Reference Sequence: NC_017628.1, 28,998 entries for chromosome), and plasmid pRS218 (NCBI Reference Sequence: NZ_CP007150.1; 3,329 entries for plasmid) from the NMEC strain RS218. Along with the experimental raw files of H622, IHE3034 and J53, raw files from a completed experiment in the same research group were run, where IHE3034 and J53 were grown on blood agar and the proteins digested as described here. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD005779, and will be made accessible upon publication [87]. Parameters used in the MaxQuant search were the standard settings, save the following:

- Lysine acetylation and glutamate/glutamine conversion to pyro-glutamate were defined as variable modifications, along with N-terminal acetylation and methionine oxidation.
- "Match between runs" was enabled, and group-specific parameters included defined enzyme specificity of trypsin or LysC, with N-terminal cleavage to proline.

Using two digestive enzymes gives a deeper coverage of protein fragments. In order to utilize this, the spectra were combined for the replicas digested with different enzymes. Allowed fragment mass tolerance value was set to a maximum of 0.50 Da, and the initial mass deviation of the precursor ion was not allowed to exceed 20 ppm. For both proteins and peptides, maximum false discovery rate (FDR), was set to 0.01. Label-free quantification (LFQ) algorithms within MaxQuant, MaxLFQ, were used to create the normalized protein intensities. They were normalized in relation to the levels of common proteins in a sample [64]. The LFQ values – "LFQ intensity" – have been used for statistical analysis, while the unnormalized spectra – "Intensity" – have been solely used to detect the presence of proteins within a sample.

Completing the identification of proteins

The resulting proteins and peptides were exported to Microsoft Excel in separate documents. The mapping of proteins lacking gene names and/or UniProt IDs was completed using protein

BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins). The amino acid sequence as defined by MaxQuant was exported to BLAST and a search was run. Only proteins that showed a 100% similarity with the amino acid sequence were considered, and the protein was chosen based on BLAST's own ranking. The information was exported and included in the established Excel document for proteins.

Using a Pearson correlation to determine correlation of intensities and LFQ intensities between replicas

For each strain, two blood cultures were inoculated creating biological replicas A and B. Each biological replica was treated in duplicate, creating replicas A1, A2, B1 and B2. To quantify the similarity between replicas, and establishing a base for examining how close they are, the Pearson correlation coefficient R was found for each strain and growth conditions. E.g., all replicas from H622 grown in blood culture were compared to each other, all replicas from IHE3034 grown on blood agar were compared to each other, but not to the replicas of IHE3034 grown in blood culture, they were analysed separately. The Pearson correlation coefficient was established for the intensities and the LFQ intensities by using Perseus: All values were transformed to $\log_2$, and a multi scatter plot of all replicas from one strain and conditions was created. In Perseus, one has the option of "Show Pearson correlation" for each scatter plot.

Filtering of proteins: Removing duplicates and proteins with very low scores

Filtering of protein duplicates was done by first identifying all proteins that shared the same gene name. This was done via Excel's innate "Find duplicates" function. As all peptides identified by MaxQuant receives a unique peptide ID, the peptide IDs by which they were identified could be compared. If double identification could be ascertained, the protein with the combination of lowest score, MS/MS count, and lowest number of unique and "razor" was deleted from the results. If the proteins shared gene names, but could not be ascertained as being duplicate identifications, the amino acid sequences were compared using protein BLAST. If the BLAST comparison showed that they did not share amino acid sequence, none were removed. Proteins only identified by site, or containing peptides that were flagged by MaxQuant as either contaminants or reverse sequence, were removed. Peptides belonging to removed proteins were deleted from the results. All proteins that were identified by MaxQuant to have a species of origin other than *Escherichia coli*, *Shigella sp.* or other closely related organisms (*Enterobacteria phage, Salmonella typhimurium*, or *Klebsiella pneumoniae*), were deleted.

Predicting subcellular localisation

Subcellular localisation was predicted using PSORTb version 3.0.2 (http://www.psort.org/psortb/index.html). This was done by exporting all the UniProt IDs from

the proteins document, and using UniProt (http://www.uniprot.org/) to search for all the proteins. UniProt has a "Download" function, and this was used to download all amino acid sequence FASTA files by choosing "Download > Download all > Format: FASTA (canonical) > Uncompressed". The resulting FASTA files were imported to PSORTb.

Identifying and removing falsely detected proteins, "false positives"

In order to identify potential false positive hits, proteins from each bacterial strain were considered isolated from the data from the other strains. Data from one strain did not interfere with data from another. Proteins from different conditions, *i.e.*, grown in blood agar and in blood culture, were considered together if they were identified in the same strain.

Visual Evaluation and Statistics to Promote Annotation (VESPA) software from Pacific Northwest National Laboratory (http://cbb.pnnl.gov/portal/software/vespa.html) was used. This software maps potential peptides, based on sequenced genomes. The aforementioned genome for IHE3034 was used to predict proteins for the IHE3034 strain, and the K12 genome MG1655 (GenBank reference: U00096.3) was used to predict proteins for the J53 strain. All peptides identified by MaxQuant that were not identified by VESPA in the strain in question, were removed from the strain's results. In order to identify false positives in H622 - for which there is yet no sequenced genome - the following, more stringent, procedure was followed: Identification type is defined as either "By MS/MS" or "By matching". "By MS/MS" is considered the strongest positive verdict for the presence of the peptide. "By matching" implies that there is hard MS evidence for the presence of the peptide only in one of two MS runs. The proteins belonging to H622 were sorted to identify the proteins where all the peptides were identified solely "By matching", or which had a number of identifications "By MS/MS" of 2 or lower. These proteins were manually curated, and the proteins with at least two replicas containing identification "By MS/MS" in identification type were kept, regardless if this was of the same peptide in different replicas. All proteins identified only by one or none "By MS/MS" were removed from the H622 strain.

Work in Perseus

Perseus is a powerful statistical software used to quantify and analyse MS data [65]. In this thesis, Perseus (http://www.perseus-framework.org, versions 1.5.6.0 and 1.5.4.1) and Microsoft Excel has been utilized in statistical analysis of the proteomic data. When the data was handled in Perseus, all intensities were transformed to $\log_2$, and all replicas from one strain or condition – e.g, all replicas from IHE3034 in blood culture, were categorised as one group using the function "categorical annotation" in Perseus. For both qualitative and quantitative analysis, the finding of a protein was considered valid only if it were found in at least two replicas per group.

An analysis of variance was carried out on proteins of strains grown in clinical blood cultures

In order to compare the differing expression of proteins detected in H622, IHE3034 and J53 when grown in blood culture, an analysis of variance (ANOVA) was carried out. An ANOVA was chosen, as it can be used when more than two means are to be compared. The analysis was done in Perseus on log_2-transformed LFQ intensities of 1 733 proteins shared by the three strains when grown in blood culture. The data from blood agar was not considered. Mode of ANOVA was permutation-based FDR, the number of randomizations 250, the FDR set to 0.01, and S0 = 0. Proteins determined valid by this test were exported to Excel, and further analysed: Those that had a maximum or minimum value in a replica that differed from the mean by two or more logarithmic units were marked as proteins of interest.

A two-tailed t test was carried out on the strains grown on blood agar and in clinical blood culture

Comparison of proteomic expression of bacteria grown in blood agar to the same strain grown in clinical blood culture was carried out in Perseus on proteins found in at least two replicas per strain. The data used was the log_2-transformed LFQ intensity levels of the proteins identified in the strains grown on blood agar, and the log_2-transformed data from the same strain grown in clinical blood culture. By changing only one variable – the growth conditions – and comparing only two components, a t-test is a valid statistical choice. A volcano plot was created using a two-tailed (non-paired) t test with 250 randomizations, FDR set to 0.01 and S0 = 0.1. Only proteins deemed significant by this test, and with levels that differed by two or more logarithmic units in one of the conditions were considered for inclusion in the results.

EcoCyc [88], DAVID [89], and UniprotKB [90] were used in the analysis work to assign the proteins biochemical functions and pathways, while STRING [91] was used as a tool to consider relations between proteins.

4. Results

In this section, the results from previously described experiments will be presented.

4.1 – Phenotypic characteristics

The phenotypic characteristics assay performed by the MID at Haukeland University Hospital according to their standardized protocols, yielded the following results:

All three strains, H622, IHE3034 and J53 demonstrated the exact same pattern, typical of *E. coli*.

Table 6 – A phenotypic characteristics assay was carried out on three strains of E. coli, *H622, IHE3034 and J53. All strains demonstrated the same results. The results are given as positive (+) and negative (-).*

Trait	**Result**	**Trait**	**Result**
ONPG, test for β-galactosidase activity	+	Movement	+
Acid	+	Urease	-
Gas	+	Indole	+
H_2S	-	Lysine	+
Mannitol fermentation	+	Lactose fermentation	+

4.2 – Multi-locus sequence typing

E. coli strains IHE3034 and H622, both isolated from the spinal fluid of neonates, were cultured on blood agar plates, lysed, and a set of fragments of specific household genes were sequenced using predetermined primers.

The resulting sequences indicate that H622 belongs to ST 421, and IHE3034 belongs to ST 95. Both strains belong to ST Complex 95 (Table 7). The sequences in their entirety, can be found in supplementary materials.

Table 7 - MLST results for H622 and IHE3034.

Strain	***Adk***	***FumC***	***GyrB***	***Icd***	***Mdh***	***PurA***	***RecA***	**ST**	**ST Complex**
H622	37	38	19	37	17	8	26	ST 421	ST 95 Complex
IHE3034	37	38	19	37	17	11	26	ST 95	ST 95 Complex

4.3 – Phylogenetic grouping

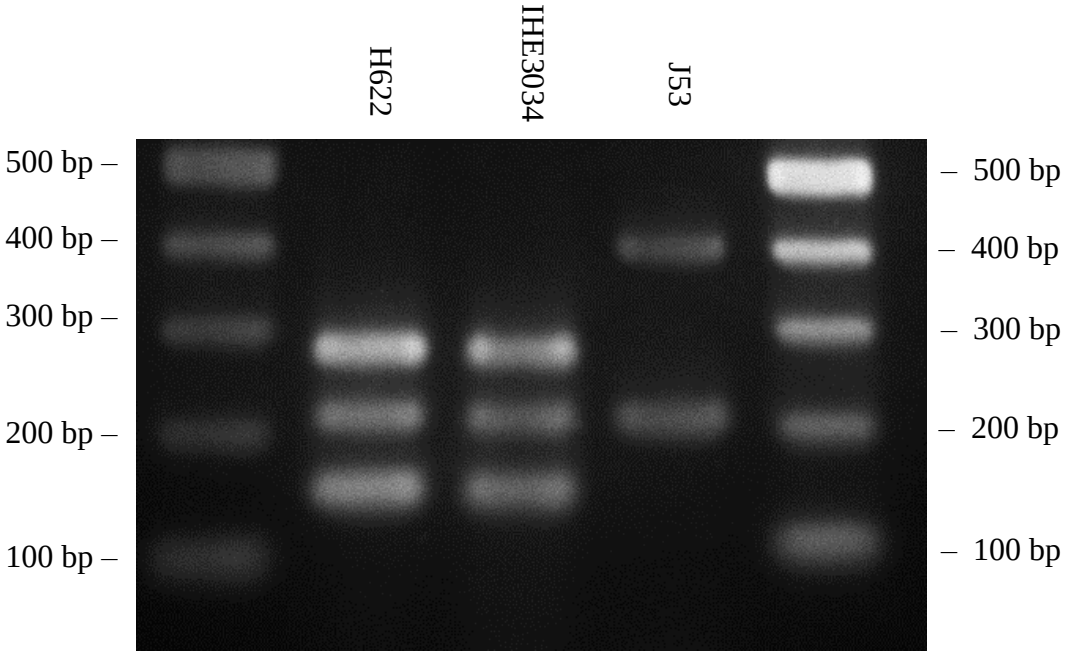

Figure 1 - Phylogenetic grouping results of H622(- + + +), IHE3034(- + + +) and J53(+ - + -).

The bacterial strains H622, IHE3034 and J53 were tested for the presence of four genes (Figure 1). H622 and IHE3034 show an identical pattern: *arpA*(-), *chuA*(+), *yja.A*(+), TspE4.C2(+). (- + + +). The K12 derivative J53 demonstrates the following pattern: *arpA*(+), *chuA*(-), *yja.A*(+), TspE4.C2(-). (+ - + -). According to Clermont et.al. [21], this indicates that the NMEC strains belong to group B2, and J53 to group A.

4.4 – Antibiotic susceptibility

The bacterial strains H622, IHE3034 and J53 were tested with a number of antibiotics (Table 8). The quality control ATCC 25922 was within the ranges established by EUCAST [92].

Only H622 demonstrated resistance towards any of the tested antimicrobials. According to clinical breakpoint tables for bacteria v.7.1 developed by EUCAST, H622 was resistant towards ampicillin [93].

Table 8 – Results of antibiotic susceptibility testing. Strip results are given in MIC, Zone indicates mm diameter of zone of inhibition around the antibiotics disc. Only H622 exhibits any resistance. "Strip" indicates MIC number, as read on the test strip, given in µg/ml. "Zone" indicates the diameter of the zone of inhibition, in millimetres. Clinical breakpoints are given. S = sensitive, R = resistant. N/A = not applicable here.

Antibiotic	**ATCC 25922**	**H622**	**IHE3034**	**J53**	**MIC breakpoint (µg/ml)**		**Zone diameter breakpoint (mm)**	
					S ≤	**R >**	**S ≥**	**R <**
Ampicillin	Strip: 8 Zone: 18	> 256 6	2 20	8 19	8	8	14	14
Cefotaxime	Strip: 0.125 Zone: 27	0.064 30	0.064 30	0.125 30	1	2	20	17
Ceftriaxone	Strip: 0.125	0.032	0.032	0.032	1	2	N/A	N/A
Ciprofloxacin	Strip: 0.016 Zone: 31	0.016 35	0.016 34	0.125 28	0.25	0.5	26	24
Gentamicin	Strip: 1 Zone: 20	1 19	1 20	0.5 23	2	4	17	24
Amoxicillin and clavulanate	Zone: 21	19	23	22	N/A	N/A	19	19
Ceftazidime	Zone: 25	27	27	28	N/A	N/A	22	19
Cefuroxime	Zone: 22	24	24	23	N/A	N/A	19	19
Ertapenem	Zone: 33	34	36	38	N/A	N/A	19	19
Imipenem	Zone: 31	29	28	31	N/A	N/A	20	17
Meropenem	Zone: 33	33	35	34	N/A	N/A	22	16
Piperacillin-tazobactam	Zone: 24	24	25	27	N/A	N/A	20	17
Trimethoprim-sulfamethoxazole	Zone: 27	29	29	32	N/A	N/A	14	11

4.5 – Pulsed-field gel electrophoresis

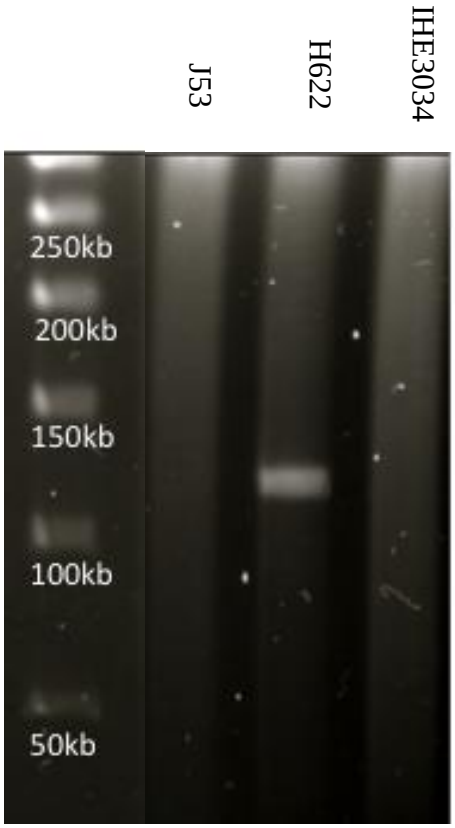

Figure 2 – PFGE results of J53, H622 and IHE3034. H622 demonstrates the presence of a plasmid between 100 – 150 kb.

Bacterial strains H622, IHE3034 and J53 were lysed, their DNA digested using endonucleases, and run on a PFGE. The resulting gel (Figure 2) shows that H622 contains a large plasmid with a size between 100 and 150 kilobase (kb). IHE3034 and J53 do not show the presence of any large plasmids.

4.6 – Presence of small plasmids

DNA was extracted from H622 and BL21(DE3)pLysS and digested using a S1 endonuclease. The agarose gel (Figure 3) shows that H622 has no genetic elements smaller than 15 kb. BL21(DE3)pLysS has a large amount of DNA at ~5 kb, and smaller amounts at 6 kb, 9 kb and 15 kb. The negative control shows no presence of DNA.

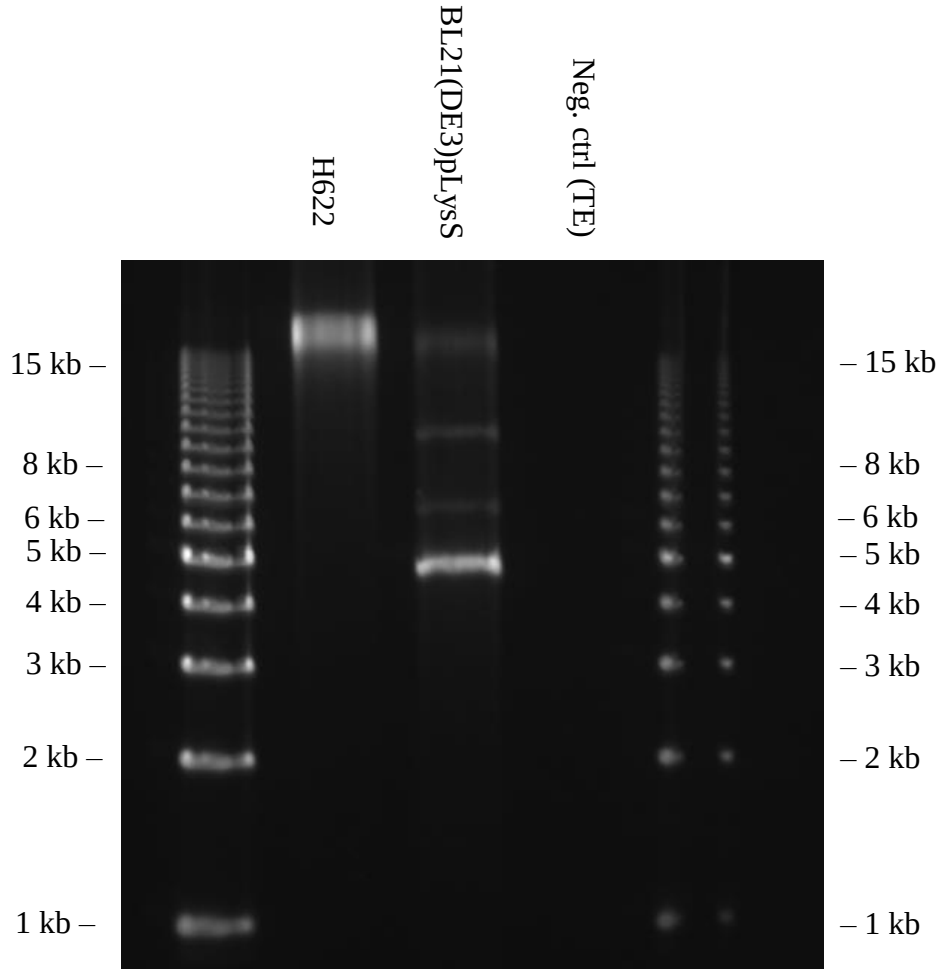

Figure 3 – DNA from H622 and BL21(DE3)pLysS digested with nuclease and run on an agarose gel electrophoresis. BL21(DE3)pLysS shows DNA at ~5kb, 6 kb, 9 kb and 15 kb. H622 has no genetic material below 15 kb. The negative control (TE) shows no DNA.

4.7 – Mass spectrometry results of bacteria grown in blood cultures and blood agar

Quantitative protein profiles of *E. coli* strains H622, IHE3034 and J53 were created using MS/MS. The bacteria were grown in clinical blood cultures, lysed and digested with proteinases, before being analysed using LC-MS/MS. Data from an earlier experiment in the same lab using the same method and the strains IHE3034 and J53 grown on blood agar were included in the analysis of MaxQuant raw spectra. The quantification technology MaxLFQ, integrated in the MaxQuant analysis platform, was used to normalize and quantify all proteins. The quantitated data is used for all statistical and quantitative analysis, such as comparison of protein levels between strains and conditions. In order to aid analysis, all LFQ values were transformed to $\log_2$ before analysed statistically or included in figures.

In order to assess the correlation of all replicas within one strain, several multi scatter plots were created in Perseus, comparing replicas within one strain to each other. This was carried out on the intensity values, and the LFQ intensities. For the normalized LFQ spectra, the Pearson correlation coefficient R varied from smallest 0.988 to largest 0.995 (Table S7 in supplementary materials). This establishes the similarity between replicas. A box plot was created to compare the distribution of LFQ levels of all proteins across strains (Supplementary materials, Figure S2). Both the scatter plots and the box plot shows a uniform distribution of LFQ intensities across strains.

There were 103 proteins identified by only one peptide, they can be found in supplementary materials Table S9, along with the peptide sequence and detailed information about the MS/MS spectrum. Proteins that shared peptides, potentially making the LFQ intensities too high or too low per protein, can be found in supplementary materials Table S8. Potential detected isoforms of proteins can be found in supplementary materials Table S10, along with alignments of the amino acid sequences.

A total of 31 785 peptides representing 2 707 proteins were detected when analysing all strains and growth conditions, with an average per strain of 2260 identified proteins (Table S2 and Table S1 in supplementary materials). The dynamic range covers proteins of LFQ intensity from smallest $2^{15.3}$ to largest $2^{34.5}$. This signifies a fold difference where the most abundant proteins are observed more than 600 000 times as often as the least abundant. Four of the proteins that were detected did not possess high enough intensity levels to be included in the quantitated LFQ intensities.

As an example of a detected peptide with a good score, Figure 4 shows a tandem mass spectrum of a peptide with the sequence HIAMHVAAK as seen in the viewer window of the MS software

MaxQuant. The fragments of the peptide can be seen as peaks along the x-axis, which represents *m/z* ratio. For the entire peptide, *m/z*=360.86, score=89.89 The y-axis represents relative abundance, with the highest peak at *m/z*= 376.22 set at 100 %, and the other peaks are given in percentages relative of this.

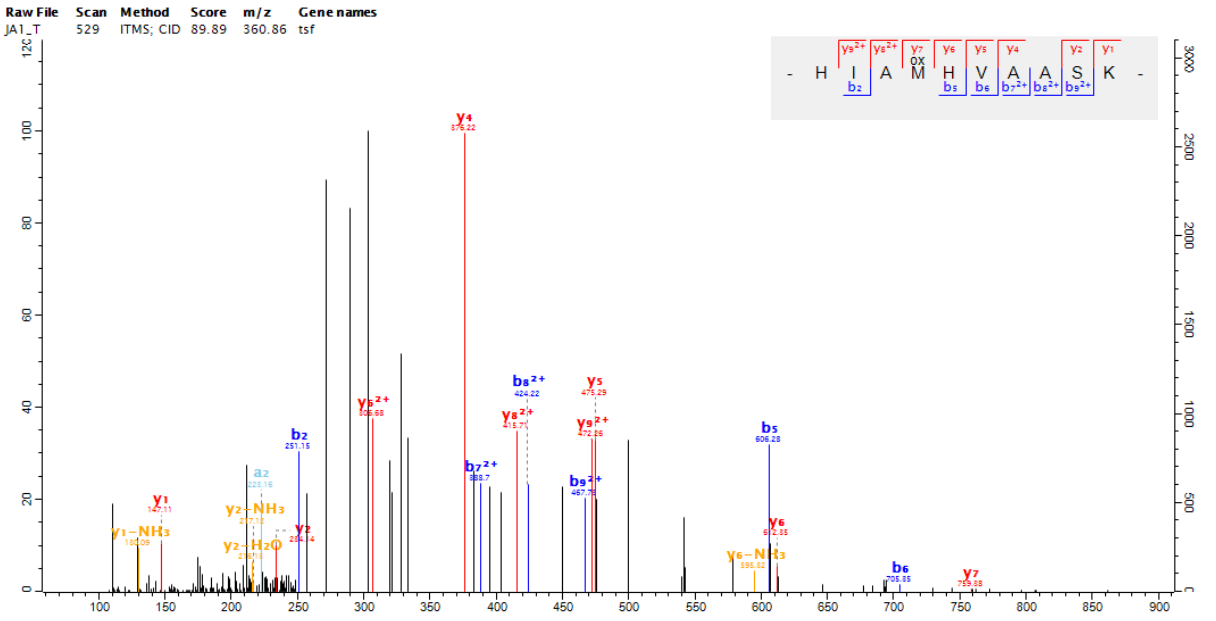

Figure 4 – Tandem mass spectrum of a peptide as seen in the viewer of MaxQuant. The y-axis represents relative abundance, and the highest peak has been assigned 100%. Mass to charge ratio, m/z *is represented by the x-axis. This peptide has been identified as having the sequence HIAMHVAAK. Source: This work.*

4.7.1 – Detection of proteins using LC-MS/MS analysis

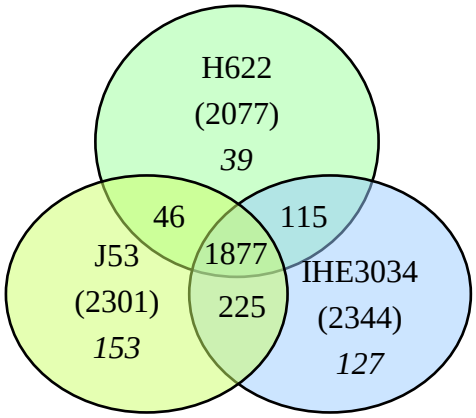

Figure 5 – Venn diagram illustrating the number of shared proteins detected between E. coli *strains H622, IHE3034 and J53 when grown in clinical blood cultures. Numbers in parentheses signify the total number of proteins detected within the strain. Numbers in italic describe the unique proteins found only in the specific strain under these conditions. This is purely the detected proteins, and do not reflect the expressed levels of proteins.*

A total of 1 877 proteins were detected in all strains, H622, IHE3034 and J53, when grown in clinical blood culture (Figure 5). When considering all strains and growth conditions, clinical blood culture and blood agar, 1730 proteins were found to be in common.

The shared proteins detected consist of the larger part of the bacteria's cellular machinery, needed for metabolism and replication. Many are found at similar levels, and do not vary significantly between strains. Proteins worth mentioning are AtpABCDEFGH (Figure 6), subunits of ATP synthase, DnaABEGJKNX, proteins considered necessary for binding to *oriC* and initiating and sustaining DNA replication, and RpsABCDEFGHIJKLMNOPQRSTU, ribosomal proteins S1-21 of the 30S ribosomal subunit [94-104]. The proteins of interest that will be discussed in this thesis are the ones that varied significantly between, or were detected solely in, strains, pathotypes and conditions.

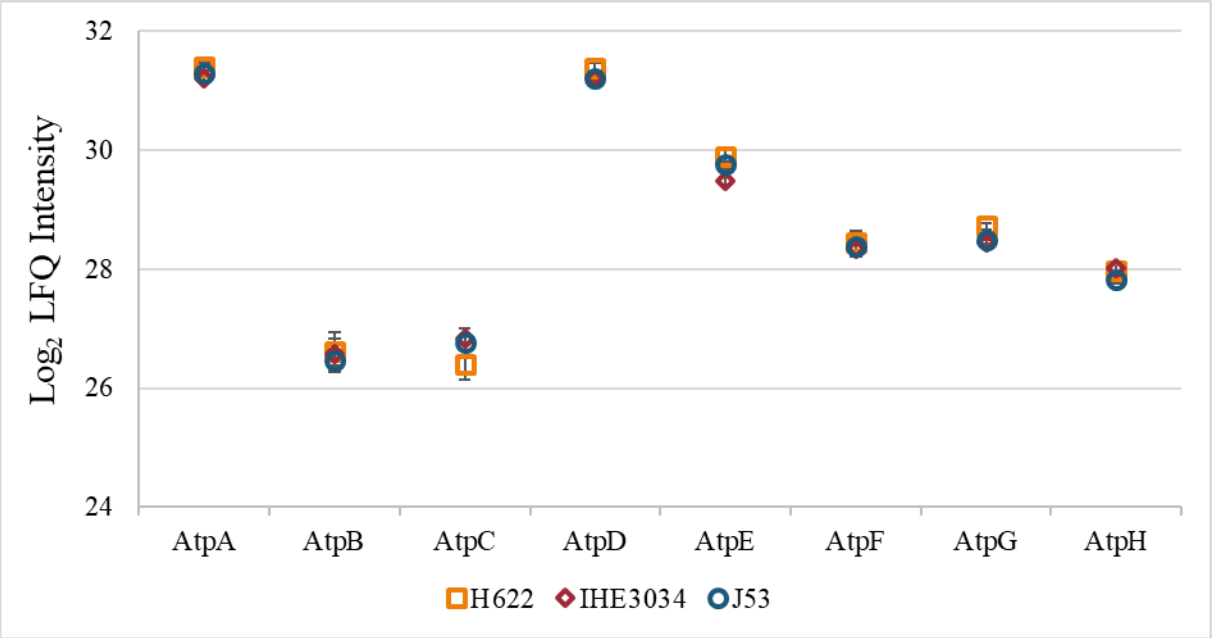

Figure 6 – Quantitative profiles of differentially expressed proteins belonging to E. coli *strains H622, IHE3034 and J53 grown in clinical blood cultures and analysed using LC-MS/MS. All error bars are ± one standard deviation of the mean. Subunits of ATP synthase are expressed similarly across strains.*

Prediction of cellular localization of proteins

A prediction of subcellular localization was carried out using PSORTb to determine the coverage and detection of proteins from different subcellular locations. PSORTb uses the FASTA sequence of proteins to give an estimation of where in the cell the proteins were likely to have been found. Strains IHE3034 and J53 have been sequenced and annotated, and their predicted proteomes were found in UniProt Proteome and downloaded from UniParc. The protein FASTA sequences were analysed in PSORTb to find the predicted subcellular location [20, 69, 105, 106]. All proteins with the predicted subcellular localization can be found in supplementary materials, Table S3.

Using PSORTb, the cellular localization of identified proteins was predicted bioinformatically (Table 9). The approximate coverage of proteins from different compartments was very similar across strains. The annotated genomes were analysed using PSORTb in the same way, and the resulting numbers used to determine how much of a certain group of proteins had been detected in the analysed strains. Approximately 50 % of annotated proteins of *E. coli* strains IHE3034 and J53 grown in clinical blood culture were detected.

Table 9 – PSORTb was used to predict the subcellular localization of detected proteins in strains grown in clinical blood cultures. The detected proteins were compared to sequenced

*and annotated genomes of IHE3034 and J53. "Number observed" indicates the number of detected proteins, sorted by PSORTb prediction. "Percent of annotated" describes how large percentage the detected proteins have covered of the annotated proteins. * The genome of H622 has not been sequenced.*

	H622		**IHE3034**		**J53**	
Localization	**Number observed**	**Percent**	**Number observed**	**Percent of annotated**	**Number observed**	**Percent of annotated**
Cytoplasmic	1246	- *	1358	71.32%	1346	65.24%
Cytoplasmic membrane	403	-	474	43.41%	471	42.32%
Extracellular	11	-	14	30.43%	9	15.79%
Outer membrane	49	-	49	51.04%	51	50.50%
Periplasmic	104	-	113	68.90%	110	67.07%
Unknown	264	-	336	26.88%	314	25.12%
Total	**2077**	**-**	**2344**	**53%**	**2301**	**48%**

4.7.2 – Quantitative analysis of proteins detected in LC-MS/MS analysis using normalized LFQ intensities

All statistical analysis of the proteins and comparison of protein levels between strains was carried out using the LFQ intensity levels. All LFQ levels shown have been log_2-transformed. Here, the strains will be generally compared and considered, before the analysis and results of differing protein levels are described. Statistical methods have been used to detect statistically significant variation between detected levels of proteins.

Correlation between biological replicas

Each strain was grown in two clinical blood culture bottles designated "A" and "B". The average between these biological replicas have been used when analysing and comparing strains, therefore a comparison of the biological replicas themselves was warranted. Figure 7 shows multi scatter plots of, and the correlation between, the LFQ intensities of the proteins in the biological replicas. *E. coli* H622 has the highest correlation between blood cultures of the three strains with a Pearson correlation coefficient R of 0.998 (Figure 7A). For the strains IHE3034 and J53, the Pearson correlation coefficients are 0.993 and 0.996, respectively (Figure 7 B and C). The variation between replicas increases with lower log_2 LFQ intensity.

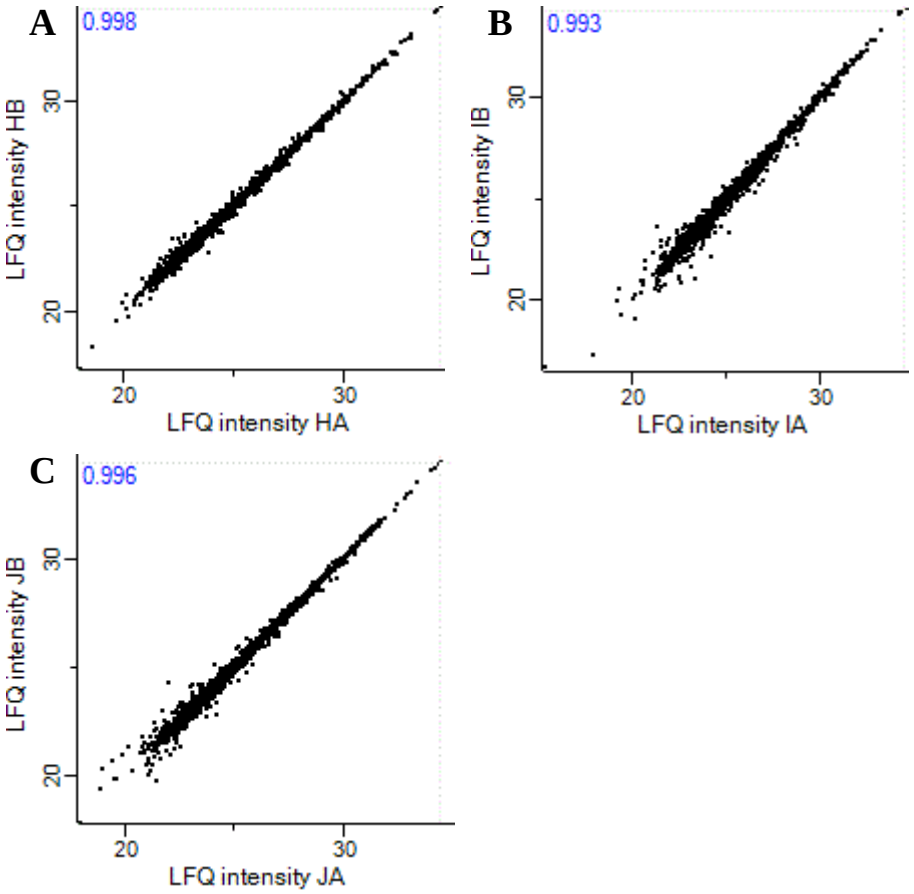

Figure 7 – E. coli *strains H622 (A), IHE3034 (B), and J53 (C) were grown in clinical blood cultures, two blood culture bottles per strains. These biological replicas were compared in scatter plots using their log_2-transformed LFQ intensities. Numbers in blue describe the Pearson correlation coefficient R.*

Variation between biological replicas

The correlation coefficient R between the biological replicas is at, or higher than, 0.993 for all three strains, and at the same time, the variation between the values of proteins found in both of the biological replicas is less than 0.3 % on average for the three strains grown in blood culture. The variation was found by determining the distance between the values per protein, and divide that value by the average value of the protein (Equation (1)). "A" and "B" refers to the values of each protein in the respective clinical blood cultures. All values used were the log_2-transformed LFQ intensities.

$$\frac{(|A-B|)/2}{(A+B)/2} \times 100\ \%$$

Equation (1) – Determining the variation between LFQ values of the biological replicas. A and B refers to the two clinical blood cultures.

J53 is used as an example, values and tendencies are representative for all three strains (Figure 8, Figure 9).

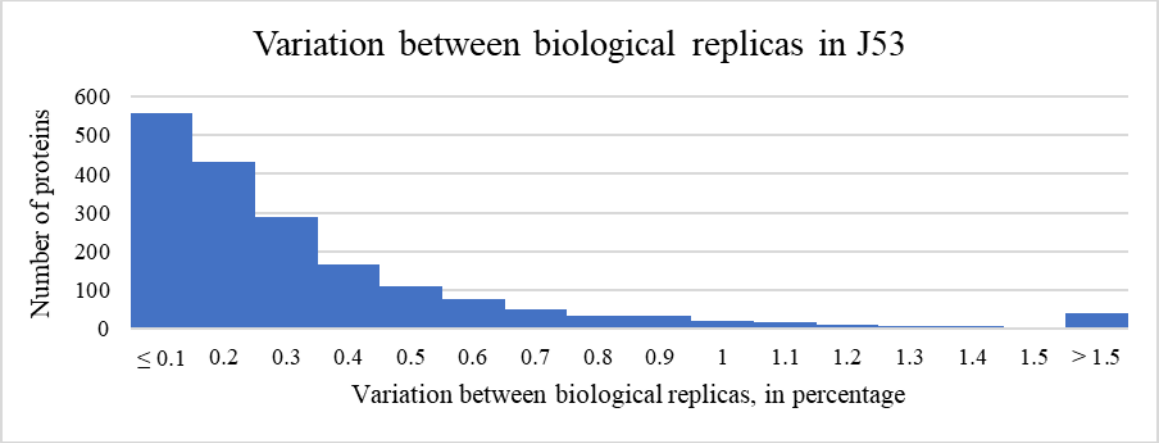

Figure 8 – Histogram showing the frequency distribution of variation between protein values in the biological replicas (blood cultures) of J53, given in percent. The majority of the proteins exhibit a variation of less than 0.2 % from the average value of the protein.

Figure 8 shows a frequency distribution of the variation between LFQ values of proteins found in both of the biological replicas. The distribution shows how accurate the quantitative data are. It demonstrates that the great majority of the proteins exhibited LFQ values that varied less than 0.2 % between blood cultures.

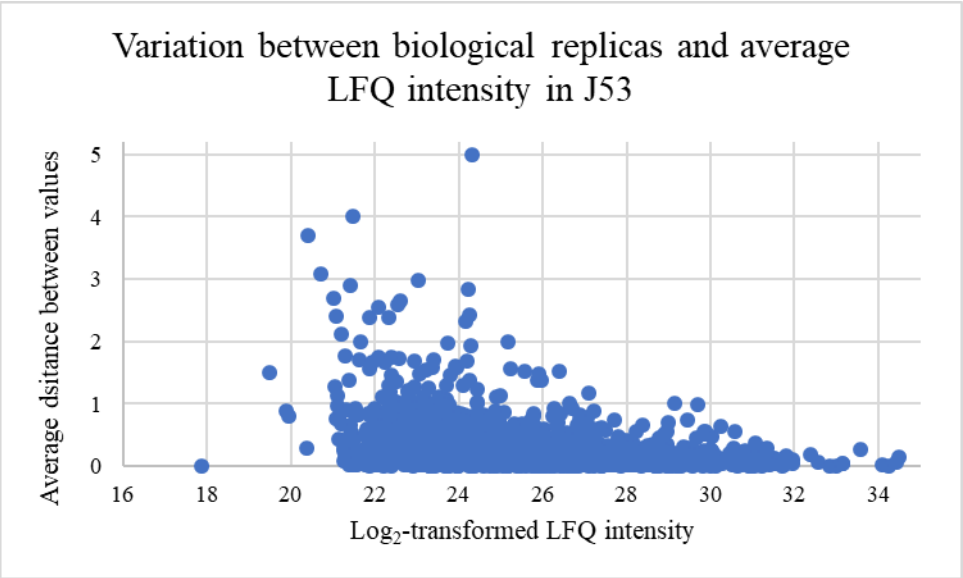

Figure 9 – The variation between blood cultures was determined and plotted against the average value per protein.

The same numbers used for Figure 8 were employed in Figure 9, where a scatter plot was created to visualize the correlation between $\log_2$ LFQ intensities and variation between the biological replicas. In general, most of the data points exhibit very low variation between the replicas, and the variation becomes lower with increasing LFQ. The largest difference between data points in the replicas is seen at lower LFQ intensity. The example shown uses values from *E. coli* strain J53, but is representative for all strains grown in blood culture. The most abundant proteins are more accurately quantified than the least abundant proteins.

Frequency distribution of protein levels and correlation between strains

Before the detailed comparison of protein levels across strains will be presented, the distribution of LFQ intensities per strain is shown. These figures describe the distribution of LFQ intensity levels found in the strains. Figure 10 shows the distribution of $\log_2$-transformed LFQ levels by frequency. All strains grown in clinical blood are relatively similar, with the majority of proteins detected having a $\log_2$ LFQ intensity between 22 and 28 (Figure 10). The curve slopes evenly towards the higher numbers on the right, and somewhat more sharply towards the lower frequencies on the left.

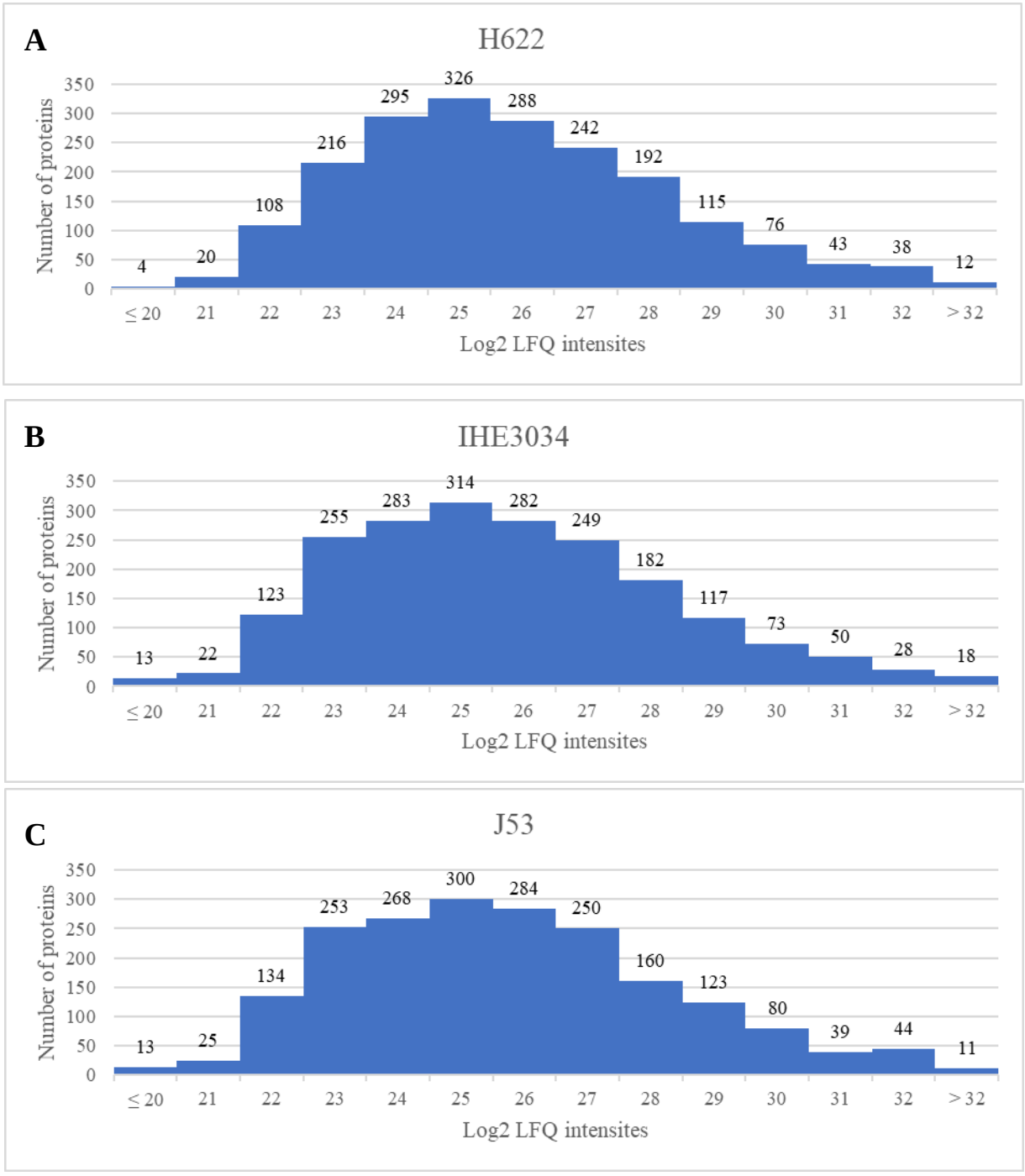

Figure 10 – Three strains of E. coli, *A) H622, B) IHE3034 and C) J53, were grown in clinical blood culture, and their proteins analysed via LC-MS/MS. The histograms show the distribution of log_2-transformed LFQ levels by frequency of proteins.*

As the distribution of LFQ levels was established within strains, further comparison of the three strains was carried out, by creating a multi scatter plot and finding the Pearson correlation between all strains. Figure 11 shows multi scatter plots comparing all strains grown in clinical blood culture using log_2-transformed LFQ levels of proteins. The Pearson correlation coefficient per scatter plot is shown in blue. The correlation between H622 and IHE3034 was

the highest, with a correlation of 0.944 (Figure 11 A). The lowest correlation could be seen between J53 and IHE3034, with a Pearson correlation coefficient of 0.868 (Figure 11 C). Slightly higher was the correlation between J53 and H622, of 0.892 (Figure 11 B).

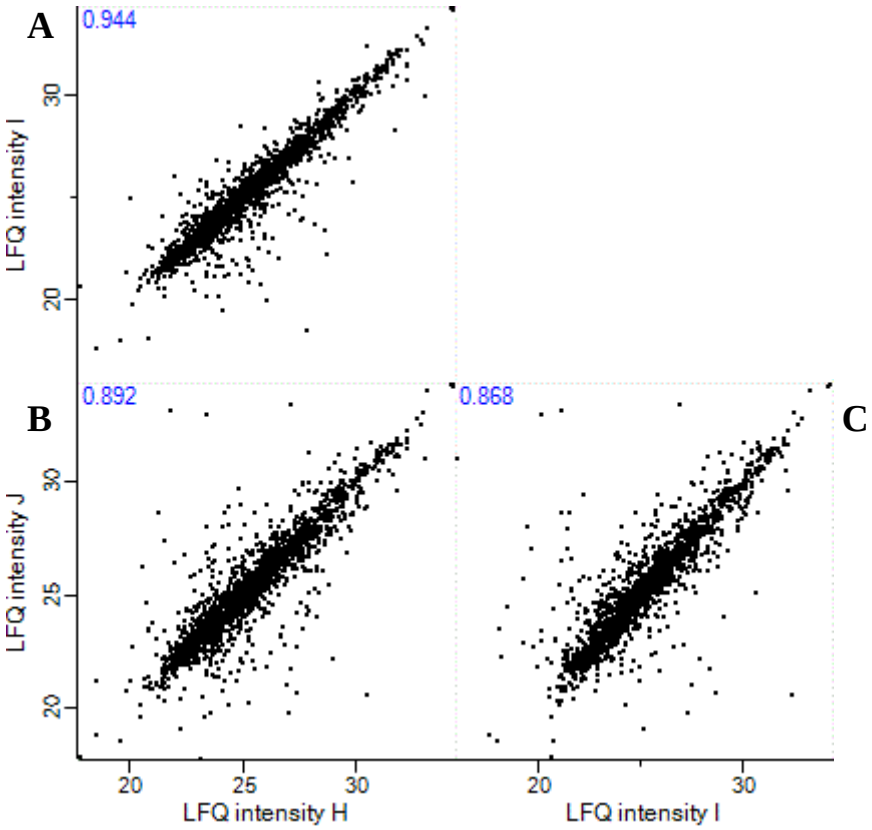

Figure 11 – A multi scatter plot comparing log_2-transformed LFQ intensities for strains H622 (H), IHE3034 (I), and J53 (J). Numbers in blue describe the Pearson correlation between strains. LFQ intensity is log_2-transformed.

Results from ANOVA analysis of strains grown in clinical blood culture

In order to compare the strains H622, IHE3034 and J53 when grown in clinical blood cultures, an ANOVA was carried out using Perseus and the results analysed using Excel. The ANOVA of the log_2-transformed LFQ spectra resulted in the finding of 79 proteins. These were the proteins that were found in all three strains and determined as significant by the Perseus ANOVA, while also having LFQ intensities in at least one replica that varied from the mean by more than two logarithmic units. I.e., LFQ intensities in one or more replicas that were four times as high as, or a quarter of, the mean (Table S4 in supplementary materials). In the following section, proteins found to differ by more than this, or not detected in all strains, will be presented. All figures are equipped with error bars representing ± one standard deviation. In many figures, the error bars are so short that they become obscured by the marker.

4.7.2.1 – Proteins upregulated in the NMEC strains when grown in clinical blood cultures

The aims of this work include comparing the protein expression levels between the *E. coli* strains H622, IHE3034 and J53 when grown in clinical blood culture. The resulting MS data were analysed statistically, and some of these results are presented here.

In this section, proteins that were deemed significant by the ANOVA analysis of shared proteins between H622, IHE3034 and J53, and/or were detected only in the NMEC strains are presented. The proteins are categorized by their functions as deemed by current research, along with a brief description. This entails the following groups of proteins:

1. Proteins detected in all strains when grown in clinical blood cultures and found to have LFQ intensities more than two logarithmic units above the mean in either H622, IHE3034, or both.
2. Proteins that were detected in only one or both of the NMEC strains when grown in clinical blood cultures.

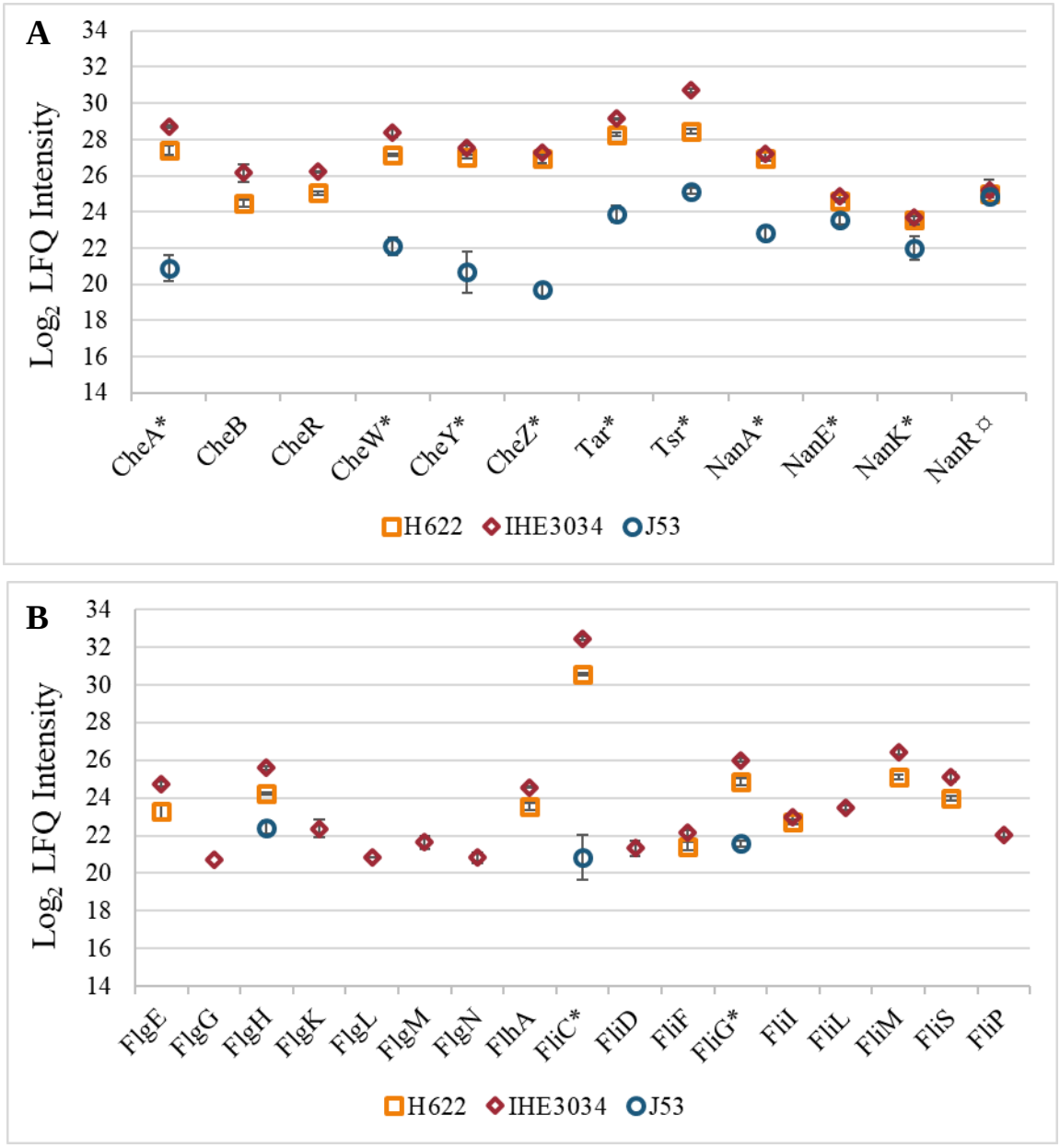

Figure 12 – Quantitative profiles of differentially expressed proteins belonging to E. coli *strains H622, IHE3034 and J53 grown in clinical blood cultures. All error bars are ± one standard deviation of the mean. Names with (*) designate proteins present in all three strains that varied by more than two logarithmic units from the mean. If unmarked and detected in all three strains, the proteins were deemed significant by the ANOVA analysis in Perseus. A) Proteins Che, Tar and Tsr, thought to be involved in chemotaxis, and proteins NanAEK, responsible for degradation of sialic acid were detected at higher level in strains H622 and IHE3034, when compared to J53. NanR, marked with ¤ , was not deemed significant by the ANOVA analysis. B) The same pattern could be observed in detected Flg and Fli proteins, important for flagellar movement.*

Carbohydrate metabolism

Degradation of sialic acid has been proposed to be mediated by the proteins NanAEK, and repressed by NanR [107, 108]. In the two NMEC strains, NanAK was found at similar levels, markedly elevated compared to the data from J53 (Figure 12A). NanE, while more similar across strains, was found at slightly higher levels, and the variation was considered significant by the ANOVA analysis done in Perseus. The repressor NanR was not considered statistically significant in Perseus, and is detected in almost identical abundance in all three strains grown in blood culture.

Motility and chemotaxis

The protein machinery involved in motility, such as flagellum- and chemotaxis-associated proteins were found expressed to a larger degree in the pathogenic strains (Figure 12B). Some of these proteins were below the detection limit in J53. According to van Albada et al. and Cardozo et al., the complex CheABRWYZ work together to transmit signals from methyl-accepting chemoreceptors Tar and Tsr to the flagellar machinery [109, 110]. Tar and Tsr both sense amino acids, and have especially high affinity for aspartate and serine, respectively [111]. Notably, these proteins follow a similar pattern; H622 and IHE3034 are close together, and J53 slightly lower. The proportional distance between the protein levels across strains is similar for several of these proteins.

As for flagellar proteins, these data depicted a clear tendency, although the exact levels of quantitated protein varied (Figure 12B). Proteins that are purported to be involved in creating a functional flagellar motor were expressed at overwhelmingly higher levels in H622 and IHE3034. FlgE, FliFIMS, and FlhA could only be detected only in the NMECs. While present in J53, FlgH and FliCG were found in higher levels in H622 and IHE3034. FlgKLMN and FliDLP were not found in H622 or J53 when considering LFQ values. These proteins, apart from FlgM, are considered parts of the flagellar motor, the filament, or movement and assembly of these [112]. Experimental evidence suggests that FlgM participates in the regulation of gene expression of the aforementioned proteins in *Salmonella typhimurium* [113]. In general, the NMEC strains exhibit similar levels, with IHE3034 slightly above H622. All proteins of the NMEC strains in Figure 12 AB demonstrate a shorter standard deviation than their counterpart in J53.

Nucleotide metabolism

Turnbough et al. suggest that PyrE, orotate phosphoribosyltransferase, catalyzes *de novo* biosynthesis of pyrimidines. This protein was found at pronouncedly higher levels in both

NMEC strains than in the K12 (Figure 13) [114]. The protein was found in similar abundance in both IHE3034 and H622.

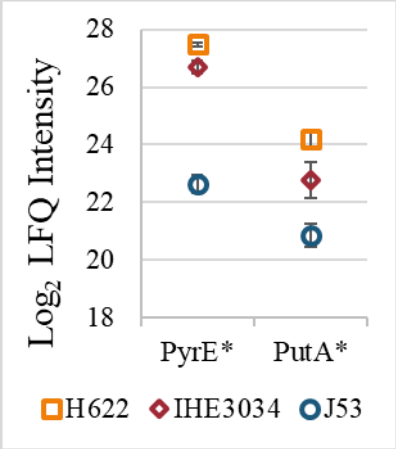

Figure 13 – Quantitative profiles of differentially expressed proteins belonging to E. coli *strains H622, IHE3034 and J53 grown in clinical blood cultures. All error bars are ± one standard deviation of the mean. Names with (*) designate proteins present in all three strains that varied by more than two logarithmic units from the mean. Proteins PyrE and PutA are considered important in nucleotide and amino acid metabolism, respectively.*

Amino acid metabolism

PutA is thought to be a bifunctional membrane-associated protein that catalyses the oxidation of L-proline to glutamate. It may also function as a transcriptional repressor of other *put* genes when proline is sufficiently available [115]. This protein is detected at lower levels in J53, with IHE3034 expressing the most (Figure 13). Compared to PyrE, the NMEC strains are further apart, and not as similar in their expression of the protein.

Capsular polysialic acid synthesis

Capsule synthesis proteins showed very different expression in the three strains. Synthesis of capsular polysialic acid NeuNAc(α2-8), is orchestrated by gene clusters *neu* and *kps* [116, 117]. For the pathogenic strains, proteins NeuABCDES and KpsCDEFMU were detected on similar levels, while not detected at all in J53 (Figure 14). KpsT was only seen in H622. These genes are all essential in creating the polysialic acid capsule seen in *E. coli* exhibiting K1 capsular antigen [47, 118]. Both NMEC strains express the proteins at very similar levels, almost identical in some proteins, except KpsF. Of this protein, H622 expresses a lower level than IHE3034.

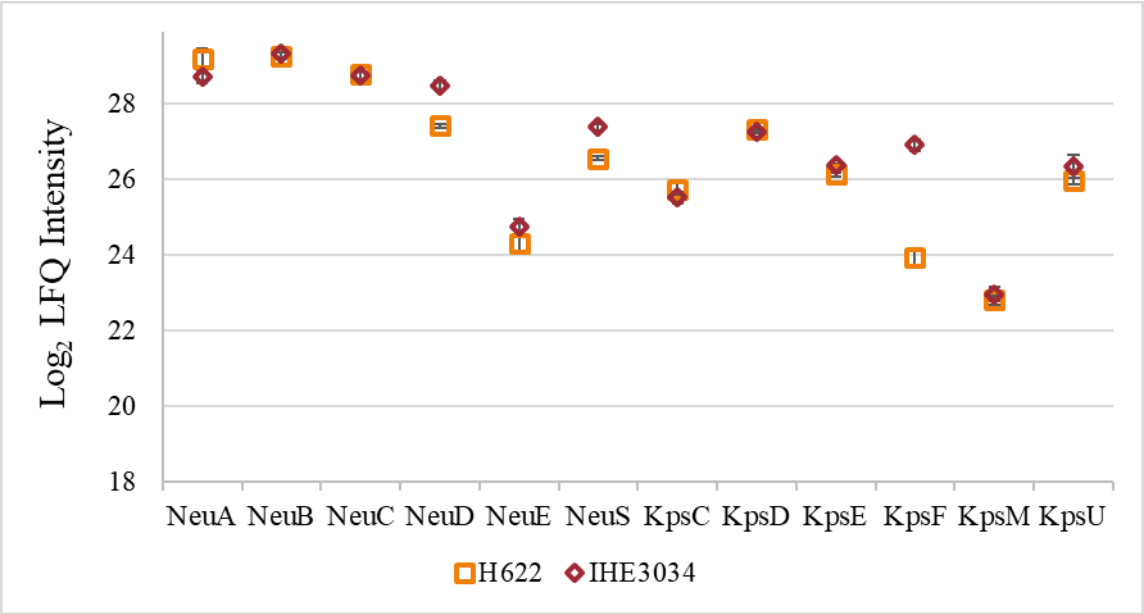

Figure 14 – Quantitative profiles of differentially expressed proteins belonging to E. coli *strains H622 and IHE3034 grown in clinical blood cultures show high levels of* kps *and* neu *proteins. All error bars are ± one standard deviation of the mean. The figure portrays relative LFQ levels of proteins important for the synthesis of capsular polysialic acid.*

Cellulose biosynthesis

A similar pattern could be seen for proteins of the *bcs* operon, which are thought to be involved in the synthesis of bacterial cellulose, and important for the formation of biofilms [119, 120]. Proteins BcsBCEQZ were significantly elevated, or detected solely in H622 and IHE3034 (Figure 15). BcsAG could only be detected in IHE3034. The majority of these proteins, especially in the NMEC strains, varied very little, and possess low standard deviations.

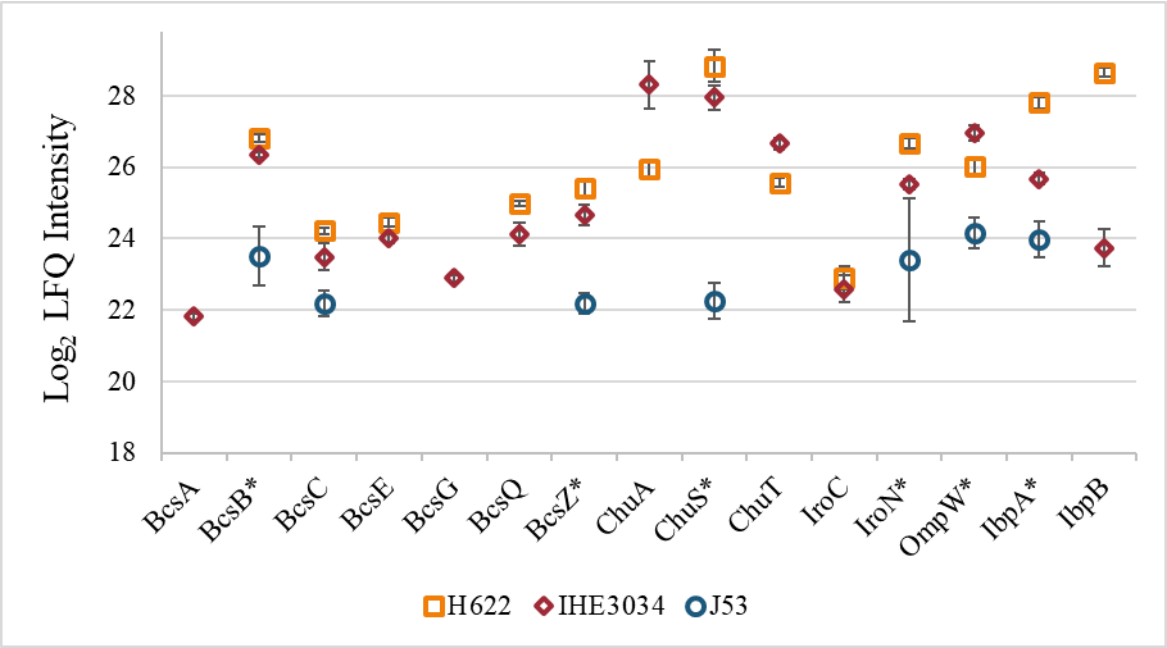

Figure 15 – Quantitative profiles of differentially expressed proteins belonging to E. coli *strains H622, IHE3034 and J53 grown in clinical blood cultures. All error bars are ± one standard deviation of the mean. Names with (*) designate proteins present in all three strains that varied by more than two logarithmic units from the mean. If unmarked and detected in all three strains, the proteins were deemed significant by the ANOVA analysis in Perseus. The three* E. coli *strains demonstrated very different levels of protein expression. Proteins BcsABCEGQZ, functional in cellulose biosynthesis, were detected at higher levels in H622 and IHE3034 than J53. Some proteins (BcsAEG) were not above the detection limit in J53. Proteins ChuAST, IroCN and OmpW, deemed important in iron acquisition, were also seen at higher levels in the NMEC strains, with three proteins (ChuAT and IroC) not above the detection limit in J53. Small heat shock proteins IbpAB were also seen differing in levels; IbpA exhibited a lower LFQ intensity in J53 than in the NMEC strains, while IbpB was not seen above the detection limit in J53.*

Iron acquisition and iron-related virulence

ChuAST, alleged receptor and oxygenase for the utilization of heme, were present in high levels in H622 and IHE3034, while not being detected in J53 [49, 121] (Figure 15). IroCN, considered important in uptake and excretion of iron chelating siderophores, were detected at higher levels in the NMEC strains, with iroC not detected in J53 [122, 123]. The ChuA protein in IHE3034 and IroN protein in J53 possessed slightly higher standard deviations due to a larger difference between the biological replicas. Outer membrane protein OmpW is a β-barrel membrane protein with a hydrophobic channel whose expression is regulated by iron. It has been tentatively suggested to be involved in iron uptake, and postulated to be involved in iron-related virulence [124, 125].

Stress response

Small heat shock proteins IbpAB were more expressed in the NMEC strains (Figure 15). These proteins are thought to bind to and stabilize unfolded proteins, readying them for refolding [126]. Despite IbpB not being found in J53, there is little similarity in the expression levels between IHE3034 and H622.

4.7.2.2 – Proteins expressed at a higher level in J53 than in H622 and IHE3034 when grown in clinical blood cultures

In this section, proteins that were detected at higher levels in the LFQ intensities of J53, compared with those of NMEC strains H622 and IHE3034, will be presented. The proteins are categorised by their functions as deemed by current research, along with a brief description. While J53 displayed a decidedly lower level of expression of virulence factors, it differed from IHE3034 and H622 in its level of expression of several metabolically related proteins.

This entails the following groups of proteins:

1. Proteins detected in all strains when grown in clinical blood cultures and found to have LFQ intensities of more than two logarithmic units above the mean in J53.
2. Proteins that were detected only in J53 when grown in clinical blood cultures.

Figure 16 – Quantitative profiles of differentially expressed proteins belonging to E. coli *strains H622, IHE3034 and J53 grown in clinical blood cultures. All error bars are ± one standard deviation of the mean. Names with (*) designate proteins present in all three strains that varied by more than two logarithmic units from the mean. If unmarked and detected in all three strains, the proteins were deemed significant by the ANOVA analysis in Perseus. J53 exhibited a higher expression of A) proteins involved in peptide and amino acid transport, and B) members of the* trp *regulon, responsible for biosynthesis of L-tryptophan, along with proteins important for amino acid metabolism, DadAX, TnaA, GlsA1, and GdhA.*

Peptide and amino acid transport

Proteins OppABCDF are regarded as vital for creating the Opp transporter for oligopeptides, and DppABCDF as a mediator in dipeptide transport [127, 128]. Apart from OppD, these were

all decidedly higher in, or solely detected in J53 (Figure 16B). The biological replicas expressed very similar levels, as seen by the very small standard deviations. Experimental evidence suggests that DppABCDF can function as a heme permease [129].

Other amino acid transport proteins could be observed at a higher expression level in J53 as well: Mtr, and SdaC, responsible for tryptophan and L-serine transport, respectively [130, 131]. SdaC differs from the other proteins presented in this paragraph, as IHE3034 expresses a similar level of the protein, and H622 is seen with an expression level lower than the other two.

Amino acid metabolism

Members of the *trp* regulon TrpABCE are considered important components of the biosynthesis of L-tryptophan [132, 133]. Proteins of this regulon all followed the same pattern, but only TrpCE differed from the mean by more than two logarithmic units. These are also the only Trp-proteins where J53 is markedly higher in its expression than the NMEC strains. TrpAB, while considered significant by the Perseus analysis, are expressed more similarly across strains than TrpCD. TrpGD was not detected in the LFQ intensities of IHE3034, while TrpR, the transcriptional repressor of the *trp* regulon, and the tryptophan-tRNA ligase TrpS were found at very similar levels across strains (Figure 16B). DadA, important for alanine degradation, and DadX, an alanine racemase thought to catalyse the conversion between L-alanine and D-alanine, were also detected in higher quantities in J53 [134]. Along with TnaA, which has been suggested to degrade tryptophan to ammonia, pyruvate and indole [135]. Proteins considered important in glutamine and glutamate metabolism, GlsA1, a glutaminase, and GdhA, a glutamate dehydrogenase, could also be observed at higher levels of expression in J53 [136, 137]. GlsA1 differs from GdhA by being similarly expressed in J53 and H622, while GdhA exhibits a very similar level of expression in the two NMEC strains, lower than in J53.

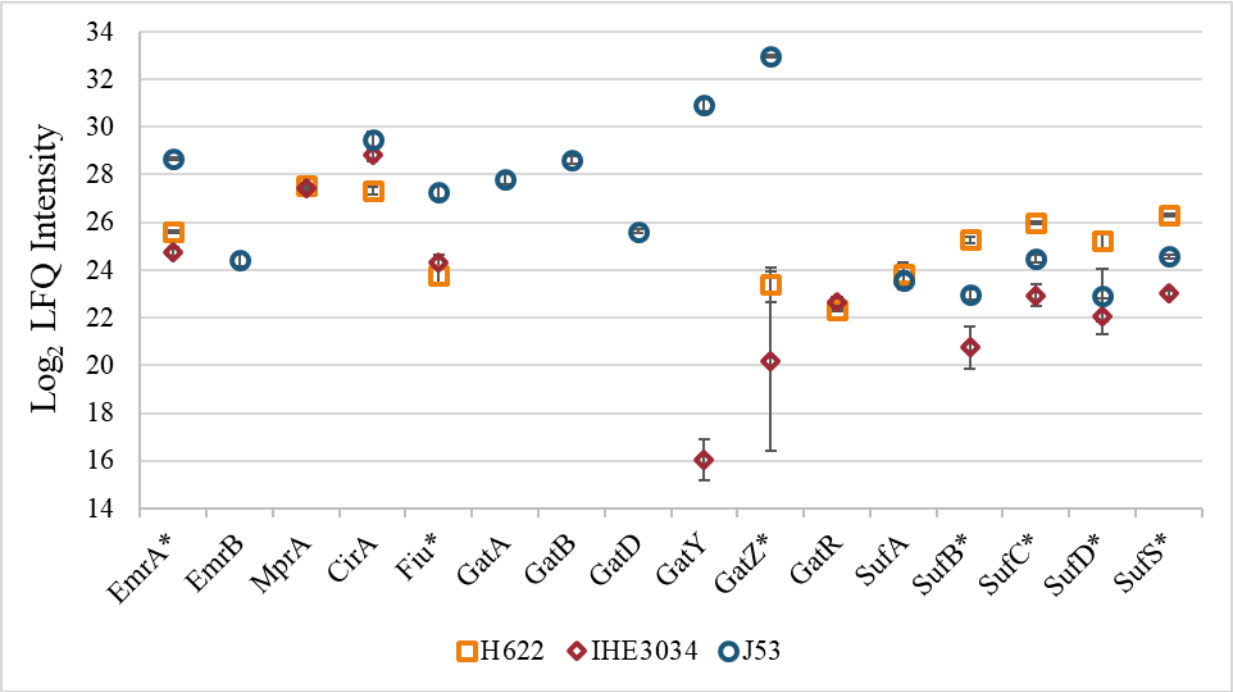

Figure 17 – Quantitative profiles of differentially expressed proteins belonging to E. coli *strains H622, IHE3034 and J53 grown in clinical blood cultures. All error bars are ± one standard deviation of the mean. Names with (*) designate proteins present in all three strains that varied by more than two logarithmic units from the mean. If unmarked and detected in all three strains, the proteins were deemed significant by the ANOVA analysis in Perseus. Proteins EmrAB, responsible for drug efflux and its repressor MprA are both expressed higher in J53. Proteins CirA and Fiu, important in iron acquisition are expressed higher in J53 than in the NMEC strains. Galactitol metabolism proteins of the* gat *operon are detected almost solely in J53, apart from the repressor GatR, which is not seen in J53. Suf proteins, important for iron-sulfur cluster assembly differ by being more highly expressed in H622.*

Drug resistance

The purported multidrug efflux pump EmrAB demonstrated significant elevation, while its repressor, MprA, was below the limit for inclusion within LFQ data in J53 (Figure 17) [138, 139].

Iron acquisition

While OmpW was seen at much higher levels in H622 and IHE3034 (Figure 15), J53 demonstrated a significant upregulation of siderophore uptake receptors CirA and Fiu (Figure 17). These proteins are thought to be able to compensate for a lower expression of OmpW [125, 140].

Carbohydrate metabolism

The Gat proteins were either exclusively expressed in J53, or at a significantly higher level. Experimental evidence suggests that proteins GatABDYZ all participate in the transport and

degradation of the sugar alcohol galactitol. GatR stood out from the rest by being undetected in J53, and at a similar level in the other two strains (Figure 17). GatR is considered a negative regulator of the *gat* operon [141] .

Iron-sulfur cluster assembly

The SufABCDS proteins involved in iron-sulfur cluster assembly were very differently expressed between the NMEC strains. They were elevated in H622, while being lower in IHE3034 than both other strains (Figure 17). SufA was only detected in H622 and J53, when considering LFQ values. This differs from the other patterns of expression levels presented here, by the NMEC strains not sharing similar expression levels, and serves as an example of proteins where all three strains differ.

4.7.2.3 – Analysis of IHE3034 and J53 cultured under different conditions

Two *E. coli* strains, the NMEC strain IHE3034 and the K12 commensal strain derivative J53 were grown on blood agar plates and in clinical blood cultures, before being enzymatically digested and analysed using LC-MS/MS, and subsequently analysed computationally. The culturing of bacteria on blood agar plates was done in an earlier experiment in the Gade Research Group for Infection and Immunity, and the raw data included in this thesis.

Effect of different culture conditions on proteomic expression in IHE3034 and J53

In order to assess the correlation between the strains when grown under different conditions, two scatter plots were created using the proteomic data in Perseus (Figure 18). When comparing IHE3034 grown in clinical blood culture to itself when grown on blood agar, the Pearson correlation coefficient was 0.739. The K12 strain J53 had a similar correlation coefficient, with 0.725. Both strains showed an even distribution of differences in proteomic expression across LFQ intensities.

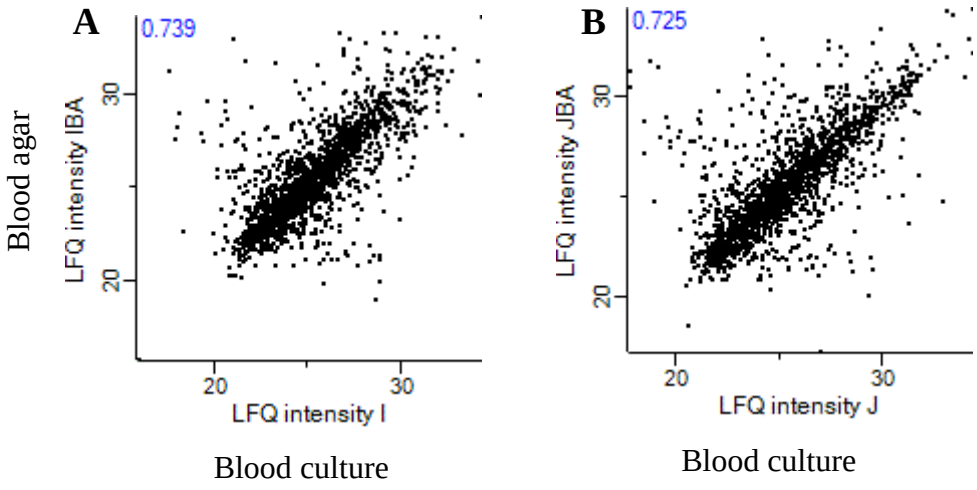

Figure 18 – E. *coli strains IHE3034 (A) and J53 (B) were grown in clinical blood culture and on blood agar plates, before being analysed using LC-MS/MS. Scatter plots were created in order to compare quantitative protein profiles across culture conditions. LFQ intensities are* log_2*-transformed.*

In order to detect the proteins that differed the most between the two conditions, the data were analysed statistically using the program Perseus. The different conditions were compared by subjecting the two datasets from the same strain to a two-tailed t test with 250 permutations and

a FDR of 0.01 using the program Perseus (Figure 19). Between IHE3034 grown in blood culture and on blood agar, 1 388 proteins were found to have a significant difference in LFQ intensity levels. Of these, 270 proteins differed in LFQ intensities between conditions by more than two logarithmic units, i.e., more than four times as much in one condition than in the other. The same numbers for J53 are 1 326 and 268, respectively (Table S5 and Table S6 in supplementary materials show all proteins that differed by more than two logarithmic units between growth conditions).

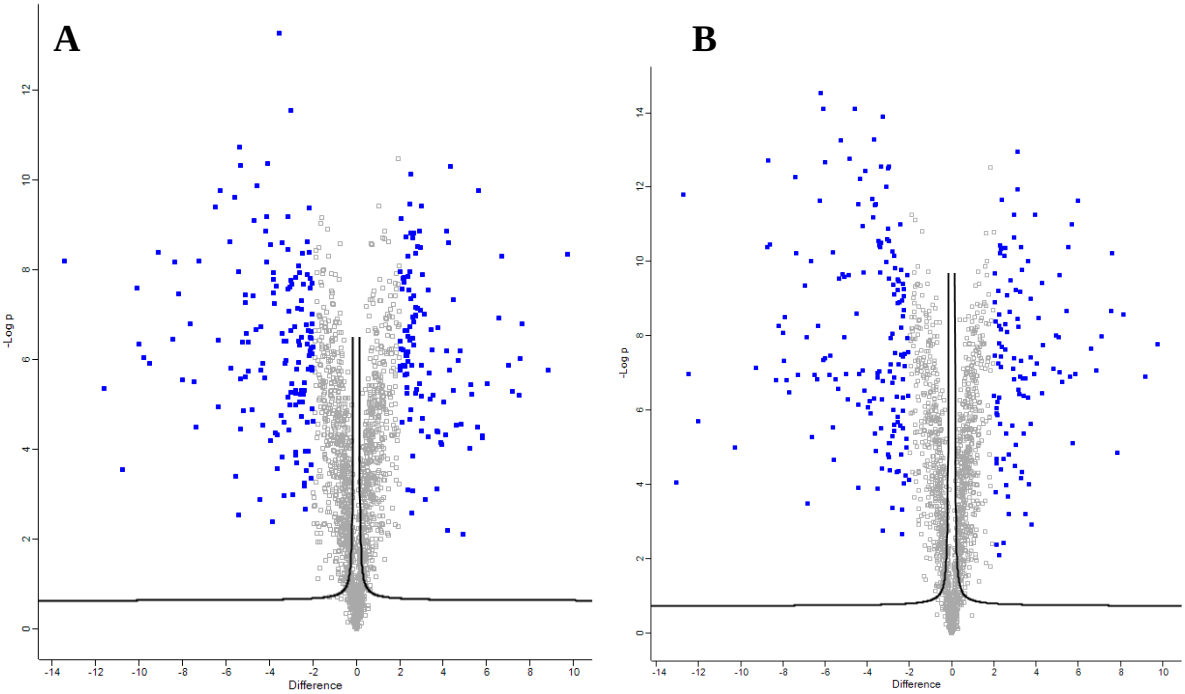

Figure 19 – E. coli *strains IHE3034 and J53 were grown in clinical blood culture, and on blood agar, before being enzymatically digested and analysed using LC-MS/MS. A two-tailed t test with 250 permutations an FDR of 0.01, and s0 of 0.1 was used to detect proteins that differed in expression between growth conditions within the same strain. The volcano plots illustrate the logarithmic fold change (x-axis) and the negative logarithm of the p-value (y-axis) of plotted proteins. Blue squares illustrate proteins that differed between conditions by more than two logarithmic units.*
A: In IHE3034, 270 proteins were detected that differed in expression levels between growth conditions by more than two logarithmic units. B: In J53, 268 proteins that differed in expression levels between growth conditions by more than two logarithmic units were detected.

The two strains shared several common traits in up- and downregulation of certain pathways and operons when grown in clinical blood culture, compared to being grown on blood agar.

Yet, a number of proteins could only be detected as to differ by more than two logarithmic units between conditions in one of the strains, not both.

The two strains shared the following changes in certain proteins:

In the following paragraphs, a selection of the proteins that follow a similar expression pattern between conditions in both strains is presented.

Proteins that were higher expressed in strains grown on blood agar

Carbohydrate metabolism

When grown on blood agar, both strains saw a higher expression of proteins purported to be involved in the utilization of different carbon sources: AceAB, AcnAB, Acs, GlcBCEG, GltA, and Mdh are proteins thought to be involved in the metabolism of isocitrate, citrate and malate. These enzymes are considered important constituents of the glyoxylate cycle, responsible for enabling the bacterium in using acetyl-CoA as a carbon source [142]. The glyoxylate cycle is closely tied to the tricarboxylic acid (TCA) cycle, and share enzymes such as isocitrate dehydrogenase, Icd. When analysed, this catabolic pathway followed the same pattern as the enzymes of the glyoxylate cycle, and saw significantly higher intensities in the bacteria grown on blood agar compared to the ones harvested from blood cultures. Important proteins of the prokaryotic TCA cycle I that were found to be significantly elevated on blood agar were SdhABCD, subunits of succinate dehydrogenase enzyme, SucAB, subunits of 2-oxoglutarate dehydrogenase complex, and SucCD, succinyl-CoA synthetase [143, 144]. FbaB, active in gluconeogenesis, was more expressed in the samples from blood agar than the ones grown in liquid blood culture [145]. However, not all proteins of carbohydrate metabolism were varied in this way. FruABK, proteins regarded as important for fructose catabolism and transport, could be detected at much lower intensities in the strains when grown on blood agar plates than in blood culture [146].

Fatty acid metabolism

FadADL, proteins that are widely accepted as catalysing β-oxidation and transporting fatty acids, were detected at lower levels in the strains grown in clinical blood culture [147-149].

Amino acid metabolism and transport

As with carbohydrate metabolism, amino acid transport proteins were detected at higher levels in blood agar-grown bacteria. Argt, a probable transporter of L-ornithine, L-arginine, and L-lysine, and Glnh, a L-glutamine transporter, could all be seen at higher expression levels when grown on blood agar [150]. Other proteins that were seen at a higher intensity in blood agar-grown bacteria include PutAP, catalysing proline degradation and transport, and Csta, a peptide transport protein induced by carbon starvation in experiments [115, 151, 152].

Bacterial electron chain transport
CyoAB, subunits of bacterial ubiquinol oxidase were seen at elevated expression levels in bacteria grown on blood agar [153].

Osmotic stress
Experimental evidence has demonstrated increased expression of OsmBCEY when bacteria have been exposed to osmotic stress [154-157]. These proteins were seen in higher levels when the bacteria were grown on blood agar. As were ProVWX, constituents of an osmosensitive transport system for glycine betaine and proline betaine [158].

Proteins that were higher expressed in strains grown in clinical blood culture

Chemotaxis
Both strains demonstrated an upregulation of chemotaxis protein CheA in the blood culture grown bacteria. The strains differed in expression of CheY: in J53, it was expressed higher in the bacteria grown on blood agar, whereas IHE3034 showed high levels of both chemotaxis proteins in the clinical blood culture.

Nucleotide biosynthesis
The proteins PurCEHLMT, important enzymes in purine biosynthesis, were more expressed in the bacteria grown in blood culture [159].

Ribosomal proteins
J53 only had one 50S ribosomal protein at elevated levels in the blood culture-grown bacteria, RpmH, while IHE3034 showed significantly higher levels of RplBDMNOPQSTVX and RpmABCDFGH in the clinical blood culture.

4.7.2.4 – Other MS results

When the MS results were examined for presence of β-lactamase, two enzymes were found: β-lactamase, gene name *ampC*, and β-lactamase TEM, gene name *bla*. Bla is found only in H622 grown in clinical blood culture (Figure 20). This enzyme is a class A β-lactamase, which can often be inhibited by small molecule inhibitors like clavulanic acid and tazobactam [160, 161]. AmpC was seen in IHE3034 and J53, in blood culture and when grown on blood agar (Figure 20). It was detected in IHE3034 when grown in blood culture, but the levels were too low to be quantified.

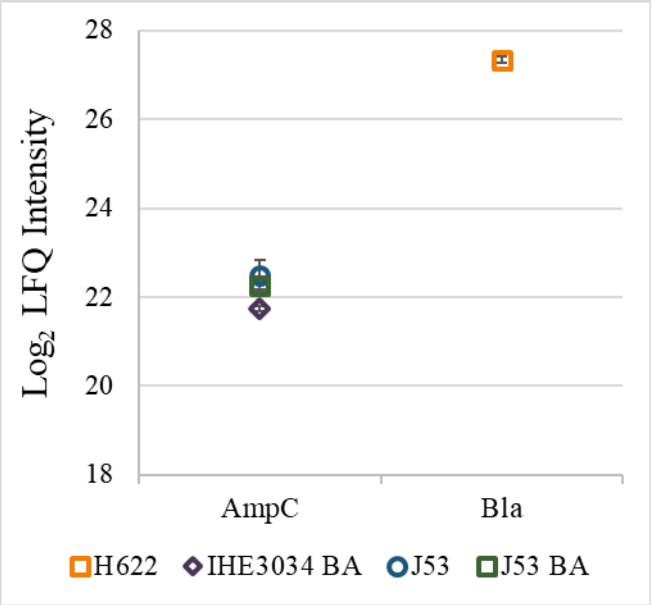

Figure 20 – Quantitative profiles of differentially expressed proteins belonging to E. coli *strains H622, IHE3034 and J53 grown in clinical blood culture and on blood agar. "BA" designates grown on blood agar. All error bars are ± one standard deviation of the mean.* AmpC *is only quantitated in J53 in clinical blood culture and on blood agar, as well as IHE3034 on blood agar.* Bla *is quantified solely in H622 in clinical blood culture.*

5. Discussion

Bacterial meningitis in infants is a potentially debilitating disease with a 15 % mortality rate and the potential of serious neurological sequelae. *E. coli* is a common bacterial cause of this infection, and the incidence of *E. coli* meningitis in an industrialized country has been estimated to 0.04 cases out of 1000 live births. In this work, three *E. coli* strains, H622, IHE3034 and J53, two of which were isolated from the spinal fluid of a neonate, were cultivated in clinical blood culture before being lysed, enzymatically digested and analysed using LC-MS/MS. Quantitative protein profiles were created for all three, and the quantitated protein expression compared for differences in relative protein levels between the strains. Before the LC-MS/MS analysis, several tests were carried out in order to characterise the *E. coli* strains genetically and biochemically.

5.1 – Phenotypic characterisation and genetic analysis

An investigation of phenotypic expression was carried out at Haukeland University Hospital following their standardized protocols. All strains shared the same results, results which are typical for *E. coli*. This establishes that the strains are very similar in expressing proteins involved in the pathways that were tested for here.

The MLST was carried out on the NMEC strains. They are both members of the same ST complex; ST complex 95. IHE3034 has been classified as ST 95 earlier, so this test confirms the earlier results [20]. This indicates that H622 and IHE3034 are closely related. There exist 23 ST complexes, which are groups of closely related STs. Each ST complex "includes at least three STs that differ from their nearest neighbour by no more than two of the seven loci" [71]. Wirth et.al. investigated 462 *E. coli* isolates from around the world, and found that ST complex 95 includes ST types from K1 strains and enteropathogenic *E. coli* (EPEC). In fact, several different pathotypes shared ST complex groups, and similar genetic traits. It has been postulated that closely related pathogenic *E. coli* such as EPEC and enterohaemorrhagic *E. coli* (EHEC) are able to change pathotype through the acquisition or loss of very few genes [71].

The strains were assigned to phylogroups through the use of the quadruplex phylo-typing method. Moriel et al. classified IHE3034 as a B2, and our results reaffirms that [20]. These results are in tune with research performed earlier, as NMEC strains usually belong to the phylogroup B2, and non-pathogenic strains to group A [37]. It also further establishes the close relations between IHE3034 and H622, and their evolutionary distance to J53. During the investigation of 462 strains, Wirth et.al. found that pathogenic *E. coli,* represented by members of groups B1, B2, and D, had a greater rate of homologous recombination in their genomes than

group A. They concluded that "pathogenic strains have accelerated rates of mutation and recombination" [71]. Pathogenic strains are subject to other evolutionary pressures than non-pathogenic ones, including evasion of the host's immune system. These evolutionary distances are reflected in the proteomic results; The multi scatter plot shows that IHE3034 and H622 are very similar in protein expression (Pearson correlation coefficient = 0.944). They are close to equal in distance to J53, with a Pearson correlation coefficient of 0.868 and 0.892, respectively (Figure 11).

5.2 – Apart from H622 exhibiting resistance towards ampicillin, the strains were susceptible to all tested antimicrobials

An antibiotic susceptibility assay was carried out on all three strains following standard protocol as established by EUCAST. The assay determined that the only resistance shown toward the antimicrobials tested here, was H622. This bacterium is resistant towards ampicillin, with a MIC reading of > 256 µg/ml and a zone of inhibition of 6 mm (Table 8). The MIC breakpoint and zone of inhibition necessary to be categorised as resistant is, respectively, > 8 µg/ml and < 14 mm. Figure S1 in supplementary materials shows that H622 was not inhibited at all by the ampicillin test strip. Some of the other strains are on the limit of being categorised as resistant, such as J53. The MIC value towards ampicillin is just within the range to be classified as sensitive. The zone of inhibition, however, is 19 mm, well above the 14 mm required. In the LC-MS/MS results, two β-lactamases were found (Figure 20). β-lactamase TEM, found in H622, is expressed in high enough abundancies to confer ampicillin resistance. Interestingly, the bacterium is sensitive towards amoxicillin and clavulanate (clavulanate is a β-lactamase inhibitor). As we did not test it against amoxicillin alone, it cannot be concluded that H622 is resistant towards amoxicillin without clavulanate. Yet, clavulanate probably inhibits the enzyme expressed here, as is common for this type of β-lactamase. The β-lactamase AmpC found in IHE3034 and J53 has not conferred resistance to these strains. Most likely, the protein is expressed in insufficient amounts. J53 expresses the enzyme higher than IHE3034 does, by about 40 percent (Figure 20, LFQ intensities of all proteins can be found in supplementary materials, Table S1). The difference in protein expression is enough to cause a MIC three times as high. Despite too low levels to successfully confer resistance, the potential for resistance is there, should the bacteria be exposed to β-lactam antibiotics and given the chance to express higher levels of the enzyme.

Antibiotic resistance is an increasing problem in health care, and resistant pathogenic *E. coli* strains are on the rise [162]. Ampicillin resistance is not uncommon in *E. coli* isolates. In fact, in a study published in 2002, 85 % of *E. coli* isolated from sepsis in very low birth-weight

infants were resistant towards ampicillin [163]. It has been speculated that the increased use of ampicillin during birth to reduce transmission of group B *Streptococcus* has contributed to this increase, as "the mothers of infants with ampicillin-resistant strains of *E. coli* were significantly more likely to have received intrapartum ampicillin than were those with ampicillin-sensitive strains" [162, 163]. However, the rate of antimicrobial resistance in NMEC strains varies, as a study from 2012 found the rate of ampicillin resistance in NMECs just above 20 % [164]. It is unknown to this thesis whether the mother of the infant from whose spinal fluid H622 was isolated had received ampicillin during birth.

5.3 – H622 carries a large plasmid

A PFGE was performed in order to examine whether H622 carries large plasmids.

The PFGE shows that H622 carries a relatively large plasmid between 100 kb and 150 kb (Figure 2). Earlier research has established that IHE3034 currently does not carry any plasmids, which is supported here [20]. Large plasmids have been shown to increase the virulence of NMEC strains, and often carry virulence genes such as iron uptake systems and antibiotic resistance [42, 164]. It is yet unknown whether this plasmid is a carrier of the resistance genes contributing to the ampicillin resistance of H622.

While it does demonstrate the presence of a large plasmid, H622 shows no signs of smaller plasmids. NMEC strains have not been shown to carry particularly small plasmids. Small plasmids may be uncommon in these bacteria, as Nicholson et al. investigated plasmids from isolated from NMEC strains, where none were smaller than 97 kb [42].

5.4 – Mass spectrometry results of bacteria grown in blood cultures and blood agar

The *E. coli* strains H622, IHE3034 and J53 were grown in clinical blood culture, lysed and digested by proteases before being analysed using LC-MS/MS. The resulting data was analysed in MaxQuant, and the detected peptides were used to identify proteins. The resulting protein intensity data underwent several statistical analyses and comparisons between strains.

This part of the discussion will focus solely on the strains grown in clinical blood culture. The strains grown on blood agar will be discussed in section 5.4.4.

5.4.1 – Protein detection and prediction of cellular localization of proteins

The detection of proteins depends upon the peptides being present in high enough abundancies to be detected, and on correct identification by MaxQuant. More than 80 percent of the detected proteins are found in all three strains (Figure 5). When considering only Figure 5, it may look as if the NMEC strains share fewer proteins than IHE3034 and J53 does. However, if one considers Figure 11, it becomes clear that the NMEC strains have the highest Pearson correlation coefficient. This can be explained by pointing out that Figure 5 depicts only the detected proteins, and does not include their respective amounts. As protein levels are included in Figure 11, this figure clearly shows that the NMEC strains have more in common with each other than either of them have in common with J53. The NMEC strains share a Pearson correlation coefficient of 0.944.

The strains express proteins necessary for regular cell maintenance and replication on similar levels (Figure 6, all protein LFQ intensities are available in supplementary materials, Table S1). This demonstrates that the core genomes of these bacteria are very similar.

PSORTb was used to predict the localization of proteins based on their FASTA sequences (Table 9). As the genomes of IHE3034 and J53 have been genome sequenced, it was possible to find their predicted proteomes in the UniProt database. The numbers for all three strains are relatively similar, with H622 displaying slightly fewer detected proteins overall. Fewer detected proteins in H622 are probably caused by the very stringent demands for filtering out false positives employed in this strain. The fact that the cytoplasmic proteins are the best covered in general is not a surprise, as they can be expected to be amongst the most abundant proteins in the lysate, and thus take up a large portion of the detected proteins. The proteins predicted to originate in the outer and cytoplasmic membranes have also been relatively well covered. Membrane proteins is one of the protein groups where MS has a lot higher coverage than 2DE, due to its ability to analyse non-soluble membrane proteins. The high coverage of proteins predicted to be of periplasmic origin is an interesting point, and the number is equally high

across strains. It is possible that these proteins have been "trapped" between the inner and outer membranes. This physical "trapping" gives an opposite effect to what may have happened to give such a low number – 25 percent coverage – for the detected extracellular proteins. Before the LC-MS/MS analysis, the bacteria are rinsed in several stages to reduce human blood proteins in the sample. This process has probably rinsed away a certain amount of the extracellular proteins. A coverage of 50 percent of predicted proteins per strain is the same level of coverage we have seen in earlier research, which suggests that the sensitivity of protein detection portrayed here is on par with earlier publications [59, 60].

5.4.2 – Validity and reliability of biological replicas

The quantitated proteins were first compared on the basis of correlation between biological replicas, the blood cultures. For all three strains grown in clinical blood cultures, the Pearson correlation coefficient was 0.993 or higher. Figure 7, Figure 8 and Figure 9 speak for the robustness of the data, showing that protein levels vary little between biological replicas. It can be concluded that the biological replicas are very similar. This gives credibility to the results.

A reason for the variation at low LFQ levels can be how the MS analysis functions. The most abundant peptide is set as 100 % intensity, and the lower abundance peptides are set relative to that. Because of this, peptides from high-abundancy proteins can "drown out" peptides from low-abundancy proteins. This makes the exact level of low-abundancy proteins less precise, as well as creating a cut-off level below which the proteins are no longer detected. By looking at Figure 9, this limit can be established as being between 20 and 21 for the data in this thesis, given in $\log_2$-transformed LFQ intensities. This cut-off level and the sensitivity of the LC-MS/MS becomes even more important when considering the frequency distribution of protein LFQ intensities (Figure 10). The frequency distribution appears to follow the curve of a normal distribution on the right side of the figure, but tapers sharply on the left. It should be pointed out that the sharp taper starts around 23 on the $\log_2$-transformed LFQ intensities. This is just above the sensitivity cut-off for of the analyses conducted here. It is very probable that the frequency distribution follows a normal distribution, but the proteins of lower intensities have not been detected due to being below the detection limit. The statistical analyses performed in this work are based on the assumption that the LFQ intensities follow a normal distribution. This assumption is very common in work pertaining statistical analyses of protein quantities [165].

5.4.3 – Virulence factors are found at higher levels in the NMEC strains

The most striking differences between the NMEC strains and J53 when grown in clinical blood culture, is that H622 and IHE3034 have a consistently higher expression of virulence factors.

The phenotypic characteristics assay shows that all the strains exhibit movement (Table 6). Despite this, the NMEC strains express higher levels of proteins considered important for motility, flagella, and chemotaxis than found in J53 (Figure 12). This indicates that these strains are better swimmers, able to move towards sources of nutrients or other chemoattractants. Movement plays a vital role in the virulence of NMEC strains. By being strong swimmers in blood, they are less the victims of random Brownian motion. A bacterium moving purposefully will be more efficient in its spread and its hunt for nutrients. Here, they are grown in blood, but it is probable that these characteristics also help the bacteria when they are suspended in spinal fluid.

An important factor connected with movement is the formation of biofilm. Wood et al. showed that *E. coli* motility was strongly correlated with its ability to form biofilms [166]. The more motile bacteria formed thicker biofilms with a larger coverage. Formation of biofilm is dependent upon the bacteria excreting extracellular polysaccharides in order to embed themselves in an extracellular matrix along with other bacterial cells. Not uncommonly, these polysaccharides include cellulose. Proteins that are members of the *bcs* operon are involved in cellulose synthesis, and experimental evidence has strongly suggested that these genes are necessary for *E. coli* biofilm formation [167]. Cellulose expression is also a factor in creation of the red, dry and rough (rdar) morphotype, along with the synthesis of curli fimbriae, controlled by genes of the *csg* operon [167, 168]. No Csg proteins were detected in this work. That does not imply they are not present, as they may be present below the detection limit. All this implies that the NMEC strains IHE3034 and H622 are able to move purposefully, and that they are proficient at forming a biofilm. Movement also plays a vital role in chemotaxis, something the NMEC strains strongly express proteins for (Figure 12). The ability to sense a chemical gradient, while being able to effectively move towards it, give these strains a strong advantage when colonizing a host. Both NMEC strains are especially rich in the chemoreceptors Tar and Tsr, which sense amino acids. These two are among the most abundant chemoreceptors on the surface of *E. coli* [169]. Tar primarily detects aspartate, asparagine, and glutamate, while Tsr is mainly attracted by serine, cysteine, alanine and glycine [111]. Yang et al. found that the degree of chemotaxis towards amino acid attractants shown by an *E. coli*, correlates with its consumption and utilization of the same amino acids [111]. The K12 strain J53 differed from

the NMEC strains in the expression of proteins related to peptide and amino acid transport. Transporters for oligo- and dipeptides, as well as single amino acids, are upregulated compared to H622 and IHE3034. In addition to transport, several proteins involved in amino acid metabolism see higher levels in this strain. It could be speculated that J53 has devoted more resources towards importing and catabolizing amino acids than the NMEC strains, while H622 and IHE3034 have devoted more resources towards chemotactic detection of amino acids.

A mammalian host environment is limited in available iron, an essential nutrient for bacteria. Pathogenic bacteria are often able to sense that they are inside a host, by noticing a period of iron starvation [170]. If one considers Figure 15, it becomes clear that iron acquisition proteins have strong presences in H622 and IHE3034. Noteworthy, ChuA is not detected in J53, nor is the gene fragment from the *chuA* gene (Figure 1).

Effectiveness in acquiring iron is considered a virulence factor, and one often seen in NMEC strains [37, 42]. Iron levels are in fact so vital for microorganisms, that one effect of the human inflammatory response is to secrete molecules that sequester iron from the plasma, lowering the levels available for the pathogen [171]. Interestingly, the IroCN proteins show a strong presence, and IroN has been pointed out earlier by Russo et al. as a potential vaccine antigen [172]. Several proteins described in this work may have multiple roles, many of which are not yet fully elucidated. One of these is the Outer membrane protein OmpW. In addition to being postulated as involved in iron uptake and iron-related virulence, experimental evidence suggests this protein also confers resistance to phagocytosis [124, 125]. Wu et al. found that high levels of OmpW protected *E. coli* strains from being phagocytosed by mouse macrophages [125]. The fact that H622 and IHE3034 express high levels of iron acquisition proteins reflect their adaptation as efficient pathogens. Or “efficient pathogens when the opportunity arises”. These are very adaptable bacteria, and can be carried in the gut without causing illness.

While J53 expresses low levels of these proteins, it expresses higher levels of peptide transport proteins OppABCF and DppABCDF. It has been postulated that DppABCDF are able to function as a heme permease, enabling the bacterium to internalize and use heme iron [129]. One protein does differ from the pattern, however; OppD. This protein was only detected in one replica in J53 when grown in blood culture, at a $\log_2$-normalized LFQ level of 23.6 (Supplementary materials Table S1). In order to be regarded as detected in a strain in this work, a protein has had to be detected in at least two replicas in a strain. OppD and OppF are the two ATP-binding and -hydrolysis proteins in the *opp* operon [128]. The cause for OppD not being detected in J53 cannot fully be concluded here.

All three strains express proteins for iron acquisition, though they differ in which proteins they favour. It is also worth noting that the iron acquisition proteins favoured by J53 are fewer in number, but are expressed at higher levels than those favoured by the NMEC strains. It would seem that the three strains employ different tactics in order to acquire iron in an iron-limited environment.

Sialic acid is usually not synthesized by bacteria, but can be used by them as a source for carbon, nitrogen and energy [173]. The NMEC strains show a high expression of proteins involved in both catabolism of sialic acid and synthesis of polysialic acid (Figure 12A, Figure 14). Interestingly, the repressor of *nan*-proteins, NanR, is detected at the same levels across strains, despite the general higher level of the proteins it represses in H622 and IHE3034. Synthesis of polysialic acid by *neu*- and *kps*-genes is responsible for synthesis of the K1 capsule, and these proteins are not seen in J53 at all. It is very unlikely it carries the genes for these proteins, as they are not found in any databases pertaining derivatives of *E. coli* K12. The *E. coli* H622 has not been serotyped, but the presence of these proteins makes it highly likely that it has the capsular antigen K1.

The K1 capsule, as mentioned earlier, is a virulence factor that enables the bacterium to "camouflage" itself using polysialic acid NeuNAc(α2-8) [45]. This makes it less likely to be spotted by the host's innate immune system. Not only does it help the bacterium avoid detection, it is anti-phagocytic, protects the bacterium from surface accumulation of C3b, and assists the bacterium in crossing the BBB [45, 46]. A protein that may also play an important role in avoiding complement activation and serum killing, is Outer membrane protein OmpA. This protein is well-studied, and may have many roles. It has been suggested as functioning as an outer membrane porin, and/or as an anchor between the outer membrane and the peptidoglycan cell wall [174, 175]. This protein has been demonstrated to protect *E. coli* K1 from complement activation, possibly by binding C4bp and stopping the rest of the complement cascade [38]. OmpA is found at very similar levels in all strains grown in clinical blood culture, with an average $\log_2$-normalized LFQ level of 31.4, ± 0.4 (Not explicitly shown, LFQ intensities of all proteins can be found in supplementary materials, Table S1). The similar levels of detected OmpA indicates that although non-pathogenic and pathogenic *E. coli* differ in protein expression, they also share many traits. As OmpA may have several functions, no conclusions can be drawn as to the role it fills in the respective bacteria presented here.

The drastically higher LFQ intensities of galactitol-degrading *gat*-proteins in J53 can be seen in direct relation to the lack of detected GatR, the repressor of these proteins.

The multidrug efflux pump proteins EmrAB are unrelated to the resistance towards sodium azide exhibited by J53. That resistance is thought to be conferred by a mutation of the *SecA* gene [176]. These proteins did not grant J53 resistance towards any of the tested antimicrobials (Table 8).

The stress response proteins IbpAB are not similarly expressed between H622 and IHE3034. IbpB is expressed very differently between them, with the highest levels seen in H622. While IbpB is not detected in J53 at all, the important factor here, is that the expression is higher in both NMEC strains than in J53. Kuczyńska-Wiśnik et al. demonstrated that a lack of IbpAB lead to higher levels of extracellular indole, and an inhibition of biofilm formation [135]. This is interesting when considering the high levels of TnaA found in J53 (Figure 16B). Perhaps it should be seen the other way; TnaA is not only detected at high levels in J53, but at low levels in the NMEC strains. As the tryptophanase TnaA increases extracellular indole concentrations, it may also inhibit biofilm formation [135]. TnaA could be considered in relation to the generally higher level of proteins involved in amino acid metabolism found in J53. None of these explanations contradict the other. TnaA may be downregulated in H622 and IHE3034, while upregulated in J53.

Some of the strongly expressed proteins differ between the NMEC strains. This is not surprising, and serves as a reminder that despite the close relations and similar protein expression, they are two different strains of *E. coli*.

The challenges of maintaining a bacterial strain *in vitro*

It is clear that IHE3034 has changed during its laboratory passage, as it originally had one small and one large plasmid, and now it has none [20, 177]. Bacteria losing plasmids during *in vitro* passage is a well-known phenomenon, and can happen due to cell division (segregational loss), or when the cost of keeping the plasmid outweighs the benefits conferred to the bacterium [178]. Moriel et.al. claims in the materials and methods section of their paper that IHE3034 "is more virulent than the ancestor strain in the sepsis mouse model" [20]. It is uncommon to hear of a bacterial strain that becomes *more* virulent during laboratory passage, unless it is cultured through an animal model. Usually, the opposite is true; maintaining strong virulence does not benefit the bacterium if there is no pressure to do so. Moriel et al. do not describe the basis for their comparison – there are no references to an experiment having been done with the parent strain. Their assessment of it being more virulent therefore seems unfounded, and cannot be used to draw any conclusions about the data presented here.

While the NMEC strains clearly are closely related and share several genetic and proteomic traits, it is important to note the large temporal distance between their isolations. While IHE3034 was isolated in Finland in 1976, H622 was isolated at Haukeland University Hospital in 2012 [177] (Nothing is yet published on H622). This entails that IHE3034 has been maintained in laboratory conditions for around 40 years, while H622 has been kept locally since its isolation. Keeping a bacterial strain unchanged for 40 years is – to put it lightly – a demanding task. Commercial vendors and organisations such as the ATCC follow strict rules and guidelines to keep their cultures close to the original. They use a seed lot system where several cultures of the same strain are frozen at passage zero, and can be thawed when new cultures are needed [179, 180]. Each seed can give rise to several "working" cultures, thus keeping the original strain from mutating and changing, as can often happen through long-term passage *in vitro.*

5.4.4 – Culturing conditions have an impact on protein expression

Included in the MaxQuant data analysis were raw files from an earlier experiment in the research group. Here, IHE3034 and J53 had been cultured on blood agar before being subjected to the same digestion and LC-MS/MS analysis as described.

The scatter plot comparing the strains under different growth conditions indicate that the changes in protein expression between cultivation in blood agar and blood culture is very similar between the strains (Figure 18). They have a Pearson correlation coefficient of 0.739 and 0.725, respectively. The number of proteins that vary is very similar, supporting the notion that the strains respond to the changes between clinical blood culture and blood agar similarly (Table S6 and Table S7 in supplementary materials). The proteins that change in expression between conditions are directly linked to the differences in media. These include proteins involved in carbohydrate metabolism, degradation of fatty acids, and proteins triggered by osmotic stress. This reflects the changes in available nutrients, as well as the osmolarity of the medium. The response is the same as was seen when Pettersen et.al. compared UPEC strain 536 grown in clinical blood culture and on blood agar [59]. Higher expression levels of systems responsible for the uptake of nutrients can be seen in connection with blood agar being a nutrient-rich medium. Reflecting the findings in this work, Pettersen et al. also detected a distinct upregulation of AceAB, AcnAB, Acs, Icd, SdhABCD, and SucABCD, some of which they contributed to glyoxylate degradation systems being induced by growth on acetate [59]. Interestingly, both strains have a higher expression of the chemotaxis proteins CheAY in

clinical blood culture. This suggests that they are both capable of sensing when they are grown in liquid media, and adapting to this. The results confirm an important aspect of *E. coli*, that it is a very versatile bacterium, able to adapt to changing environments.

Concluding remarks

The main differences between *E. coli* strains H622, IHE3034 and J53 reflect adaptations of pathogenic and non-pathogenic bacteria. While the commensal K12 derivative J53 directs energy into nutrient scavenging, the pathogenic NMEC strains H622 and IHE3034 express higher levels of virulence factors such as proteins responsible for motility, biofilm formation, iron acquisition, and capsule synthesis. The NMEC strains show a greater similarity in genetic makeup and evolutionary relations with each other than either does with J53.

When grown on blood agar, the strains demonstrated many of the changes that have been detected in earlier works, most of which pertain to the availability of nutrients in blood agar compared to clinical blood culture.

We have compared the proteomic expressions of H622, IHE3034 and J53 successfully using mass spectrometry. We have also determined the evolutionary relations between the strains, and characterised H622 genetically and phenotypically.

Future perspectives

This work has focused mainly on comparing the proteomic expression of H622 and IHE3034 to that of J53. The results presented here shed light on some of the processes that differentiate pathogenic *E. coli* strains from commensal ones.

Mass spectrometric analyses yield massive amounts of data. Future and more detailed analyses of the raw data acquired here can be used to expand our knowledge about the protein expression of the strains presented. The raw files have been deposited to the ProteomeXchange Consortium with the dataset identifier PXD005779, and will be made accessible upon publication. The data may also be employed to further the search for vaccine antigens found in pathogenic *E. coli*. Valuable for this endeavour, mass spectrometry has the strength of being able to detect outer membrane proteins. The two NMEC strains H622 and IHE3034 should be more thoroughly compared to detect differences between the strains, and to create a detailed and inclusive pathotype. The mass spectrometry data can be used in general studies of *E. coli* and comparison of these strains to other pathotypes such as ETEC, UPEC, and SEPEC. The sequenced genomes should be utilized to conduct more extensive proteogenomic research.

An interesting experiment that could create a basis for comparison would be to grow the same strains in a liquid medium such as Luria Bertani broth, in order to compare *E. coli* grown in liquid non-human medium to those grown in clinical blood culture. This way, one can assess whether or not the observed changes are caused by liquid culture conditions, or other factors such as iron availability.

References

1. Blount, Z.D., *The unexhausted potential of E. coli.* eLife, 2015. **4**.
2. Darnton, N.C., et al., *On Torque and Tumbling in Swimming Escherichia coli.* J Bacteriol, 2007. **189**(5): p. 1756-64.
3. Escherich, T., *The intestinal bacteria of the neonate and breast-fed infant. 1884.* Rev Infect Dis, 1988. **10**(6): p. 1220-5.
4. Shulman, S.T., H.C. Friedmann, and R.H. Sims, *Theodor Escherich: the first pediatric infectious diseases physician?* Clin Infect Dis, 2007. **45**(8): p. 1025-9.
5. Garrity, G., et al., *Volume 2: The Proteobacteria, Part B: The Gammaproteobacteria*, ed. G. Garrity. 2007: New York: Springer.
6. Bentley, R. and R. Meganathan, *Biosynthesis of vitamin K (menaquinone) in bacteria.* Microbiol Rev, 1982. **46**(3): p. 241-80.
7. Lawrence, J.G. and J.R. Roth, *Evolution of Coenzyme B(12) Synthesis among Enteric Bacteria: Evidence for Loss and Reacquisition of a Multigene Complex.* Genetics, 1996. **142**(1): p. 11-24.
8. Turner, L., W.S. Ryu, and H.C. Berg, *Real-Time Imaging of Fluorescent Flagellar Filaments.* J Bacteriol, 2000. **182**(10): p. 2793-801.
9. Wadhams, G.H. and J.P. Armitage, *Making sense of it all: bacterial chemotaxis.* Nat Rev Mol Cell Biol, 2004. **5**(12): p. 1024-37.
10. Beloin, C., A. Roux, and J.M. Ghigo, *Escherichia coli biofilms.* Curr Top Microbiol Immunol, 2008. **322**: p. 249-89.
11. Leimbach, A., J. Hacker, and U. Dobrindt, *E. coli as an all-rounder: the thin line between commensalism and pathogenicity.* Curr Top Microbiol Immunol, 2013. **358**: p. 3-32.
12. Croxen, M.A. and B.B. Finlay, *Molecular mechanisms of Escherichia coli pathogenicity.* Nat Rev Micro, 2010. **8**(1): p. 26-38.
13. Kaper, J.B., J.P. Nataro, and H.L. Mobley, *Pathogenic Escherichia coli.* Nat Rev Microbiol, 2004. **2**(2): p. 123-40.
14. Medini, D., et al., *The microbial pan-genome.* Curr Opin Genet Dev, 2005. **15**(6): p. 589-94.
15. Segerman, B., *The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories.* Front Cell Infect Microbiol, 2012. **2**.
16. Land, M., et al., *Insights from 20 years of bacterial genome sequencing.* Funct Integr Genomics, 2015. **15**(2): p. 141-61.
17. Kohler, C.D. and U. Dobrindt, *What defines extraintestinal pathogenic Escherichia coli?* Int J Med Microbiol, 2011. **301**(8): p. 642-7.
18. Dobrindt, U., *(Patho-)Genomics of Escherichia coli.* Int J Med Microbiol, 2005. **295**(6-7): p. 357-71.
19. Johnson, T.J., et al., *The genome sequence of avian pathogenic Escherichia coli strain O1:K1:H7 shares strong similarities with human extraintestinal pathogenic E. coli genomes.* J Bacteriol, 2007. **189**(8): p. 3228-36.
20. Moriel, D.G., et al., *Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic Escherichia coli.* Proc Natl Acad Sci U S A, 2010. **107**(20): p. 9072-7.
21. Clermont, O., et al., *The Clermont Escherichia coli phylo-typing method revisited: improvement of specificity and detection of new phylo-groups.* Environ Microbiol Rep, 2013. **5**(1): p. 58-65.
22. Maiden, M.C., *Multilocus sequence typing of bacteria.* Annu Rev Microbiol, 2006. **60**: p. 561-88.

23. Johnson, J.R. and A.L. Stell, *Extended virulence genotypes of Escherichia coli strains from patients with urosepsis in relation to phylogeny and host compromise.* J Infect Dis, 2000. **181**(1): p. 261-72.
24. Picard, B., et al., *The link between phylogeny and virulence in Escherichia coli extraintestinal infection.* Infect Immun, 1999. **67**(2): p. 546-53.
25. Tenaillon, O., et al., *The population genetics of commensal Escherichia coli.* Nat Rev Microbiol, 2010. **8**(3): p. 207-17.
26. Doumith, M., et al., *Rapid Identification of Major Escherichia coli Sequence Types Causing Urinary Tract and Bloodstream Infections.* J Clin Microbiol, 2015. **53**(1): p. 160-6.
27. Bonacorsi, S. and E. Bingen, *Molecular epidemiology of Escherichia coli causing neonatal meningitis.* Int J Med Microbiol, 2005. **295**(6-7): p. 373-81.
28. Baud, O. and Y. Aujard, *Neonatal bacterial meningitis.* Handb Clin Neurol, 2013. **112**: p. 1109-13.
29. Heath, P.T., N.K. Nik Yusoff, and C.J. Baker, *Neonatal meningitis.* Arch Dis Child Fetal Neonatal Ed, 2003. **88**(3): p. F173-8.
30. Gaschignard, J., et al., *Neonatal Bacterial Meningitis: 444 Cases in 7 Years.* Pediatr Infect Dis J, 2011. **30**(3): p. 212-7.
31. Houdouin, V., et al., *Association between mortality of Escherichia coli meningitis in young infants and non-virulent clonal groups of strains.* Clin Microbiol Infect, 2008. **14**(7): p. 685-90.
32. Basmaci, R., et al., *Escherichia Coli Meningitis Features in 325 Children From 2001 to 2013 in France.* Clin Infect Dis, 2015. **61**(5): p. 779-86.
33. Stevens, J.P., et al., *Long term outcome of neonatal meningitis.* Arch Dis Child Fetal Neonatal Ed, 2003. **88**(3): p. F179-84.
34. Okike, I.O., et al., *Incidence, etiology, and outcome of bacterial meningitis in infants aged <90 days in the United kingdom and Republic of Ireland: prospective, enhanced, national population-based surveillance.* Clin Infect Dis, 2014. **59**(10): p. e150-7.
35. Grimwood, K., et al., *Twelve year outcomes following bacterial meningitis: further.* Arch Dis Child, 2000. **83**(2): p. 111-6.
36. Kim, K.S., *Pathogenesis of bacterial meningitis: from bacteraemia to neuronal injury.* Nat Rev Neurosci, 2003. **4**(5): p. 376-85.
37. Wijetunge, D.S.S., et al., *Characterizing the pathotype of neonatal meningitis causing Escherichia coli (NMEC).* BMC Microbiol, 2015. **15**.
38. Wooster, D.G., et al., *Logarithmic phase Escherichia coli K1 efficiently avoids serum killing by promoting C4bp-mediated C3b and C4b degradation.* Immunology, 2006. **117**(4): p. 482-93.
39. Prasadarao, N.V., et al., *Outer Membrane Protein A-Promoted Actin Condensation of Brain Microvascular Endothelial Cells Is Required for Escherichia coli Invasion.* Infect Immun, 1999. **67**(11): p. 5775-83.
40. Johnson, J.R., et al., *Phylogenetic distribution of virulence-associated genes among Escherichia coli isolates associated with neonatal bacterial meningitis in the Netherlands.* J Infect Dis, 2002. **185**(6): p. 774-84.
41. Picard, B., et al., *Genetic structures of the B2 and B1 Escherichia coli strains responsible for extra-intestinal infections.* J Gen Microbiol, 1993. **139**(12): p. 3079-88.
42. Nicholson, B.A., et al., *Genetic Characterization of ExPEC-Like Virulence Plasmids among a Subset of NMEC.* PLoS One, 2016. **11**(1).

43. Bonacorsi, S.P., et al., *Identification of regions of the Escherichia coli chromosome specific for neonatal meningitis-associated strains.* Infect Immun, 2000. **68**(4): p. 2096-101.
44. Severi, E., D.W. Hood, and G.H. Thomas, *Sialic acid utilization by bacterial pathogens.* Microbiology, 2007. **153**(Pt 9): p. 2817-22.
45. Kaczmarek, A., A. Budzyńska, and E. Gospodarek, *Detection of K1 antigen of Escherichia coli rods isolated from pregnant women and neonates.* Folia Microbiol (Praha), 2014. **59**(5): p. 419-22.
46. Willis, L.M. and C. Whitfield, *Structure, biosynthesis, and function of bacterial capsular polysaccharides synthesized by ABC transporter-dependent pathways.* Carbohydrate Research, 2013. **378**: p. 35-44.
47. Daines, D.A., et al., *NeuD plays a role in the synthesis of sialic acid in Escherichia coli K1.* FEMS Microbiol Lett, 2000. **189**(2): p. 281-4.
48. Moissenet, D., et al., *Meningitis Caused by Escherichia coli Producing TEM-52 Extended-Spectrum β-Lactamase within an Extensive Outbreak in a Neonatal Ward: Epidemiological Investigation and Characterization of the Strain.* J Clin Microbiol, 2010. **48**(7): p. 2459-63.
49. Nagy, G., et al., *Expression of Hemin Receptor Molecule ChuA Is Influenced by RfaH in Uropathogenic Escherichia coli Strain 536.* Infect Immun, 2001. **69**(3): p. 1924-8.
50. Graves, P.R. and T.A.J. Haystead, *Molecular Biologist's Guide to Proteomics.* Microbiol Mol Biol Rev, 2002. **66**(1): p. 39-63.
51. Armengaud, J., *Microbiology and proteomics, getting the best of both worlds!* Environ Microbiol, 2013. **15**(1): p. 12-23.
52. Witzmann, F.A. and J. Li, *Cutting-edge technology. II. Proteomics: core technologies and applications in physiology.* Am J Physiol Gastrointest Liver Physiol, 2002. **282**(5): p. G735-41.
53. Seng, P., et al., *MALDI-TOF-mass spectrometry applications in clinical microbiology.* Future Microbiol, 2010. **5**(11): p. 1733-54.
54. Pitt, J.J., *Principles and Applications of Liquid Chromatography-Mass Spectrometry in Clinical Biochemistry.* Clin Biochem Rev, 2009. **30**(1): p. 19-34.
55. Ho, C., et al., *Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications.* Clin Biochem Rev, 2003. **24**(1): p. 3-12.
56. Wells, J.M. and S.A. McLuckey, *Collision-induced dissociation (CID) of peptides and proteins.* Methods Enzymol, 2005. **402**: p. 148-85.
57. Hodge, K., et al., *Cleaning up the masses: Exclusion lists to reduce contamination with HPLC-MS/MS().* J Proteomics, 2013. **88**: p. 92-103.
58. Han, M.J. and S.Y. Lee, *The Escherichia coli Proteome: Past, Present, and Future Prospects.* Microbiol Mol Biol Rev, 2006. **70**(2): p. 362-439.
59. Pettersen, V.K., et al., *Coordination of Metabolism and Virulence Factors Expression of Extraintestinal Pathogenic Escherichia coli Purified from Blood Cultures of Patients with Sepsis.* Mol Cell Proteomics, 2016. **15**(9): p. 2890-907.
60. Krug, K., et al., *Deep Coverage of the Escherichia coli Proteome Enables the Assessment of False Discovery Rates in Simple Proteogenomic Experiments.* Mol Cell Proteomics, 2013. **12**(11): p. 3420-30.
61. Iwasaki, M., et al., *One-dimensional capillary liquid chromatographic separation coupled with tandem mass spectrometry unveils the Escherichia coli proteome on a microarray scale.* Anal Chem, 2010. **82**(7): p. 2616-20.
62. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.* Nat Biotechnol, 2008. **26**(12): p. 1367-72.

63. Cox, J., et al., *Andromeda: a peptide search engine integrated into the MaxQuant environment.* J Proteome Res, 2011. **10**(4): p. 1794-805.
64. Cox, J., et al., *Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ.* Mol Cell Proteomics, 2014. **13**(9): p. 2513-26.
65. Tyanova, S., et al., *The Perseus computational platform for comprehensive analysis of (prote)omics data.* Nat Methods, 2016. **13**(9): p. 731-40.
66. *UniProt: the universal protein knowledgebase.* Nucleic Acids Res, 2017. **45**(Database issue): p. D158-69.
67. Keseler, I.M., et al., *The EcoCyc database: reflecting new knowledge about Escherichia coli K-12.* Nucleic Acids Res, 2017. **45**(Database issue): p. D543-50.
68. Caspi, R., et al., *The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases.* Nucleic Acids Res, 2012. **40**(Database issue): p. D742-53.
69. Yi, H., et al., *Genome Sequence of Escherichia coli J53, a Reference Strain for Genetic Studies.* J Bacteriol, 2012. **194**(14): p. 3742-3.
70. Clermont, O., D. Gordon, and E. Denamur, *Guide to the various phylogenetic classification schemes for Escherichia coli and the correspondence among schemes.* Microbiology, 2015. **161**(Pt 5): p. 980-8.
71. Wirth, T., et al., *Sex and virulence in Escherichia coli: an evolutionary perspective.* Mol Microbiol, 2006. **60**(5): p. 1136-51.
72. Achtman, M. *E. coli primers in MLST Databases at University of Warwick*. 2017 [cited 2017 02.03.2017]; Available from: http://mlst.warwick.ac.uk/mlst/dbs/Ecoli/documents/primersColi_html.
73. University of Bergen. *Protocol BigDye v.3.1*. [cited 2016 7th december]; Available from: http://www.uib.no/en/seqlab/55363/protocol-bigdye-v31.
74. Achtman, M. *MLST Database at University of Warwick (Now retired)*. 2017 [cited 2017 02.03.2017]; Available from: http://mlst.warwick.ac.uk/mlst/dbs/Ecoli.
75. Clermont, O., S. Bonacorsi, and E. Bingen, *Rapid and simple determination of the Escherichia coli phylogenetic group.* Appl Environ Microbiol, 2000. **66**(10): p. 4555-8.
76. VanMeter, K.C. and R.J. Hubert, *Microbiology for the Healthcare Professional*. 2016.
77. Madigan, M.T., et al., *Brock biology of microorganisms*. 13. edition / ed. 2012, San Francisco, Calif.: Pearson/Benjamin Cummings. xxviii, 1061 sider.
78. EUCAST. *The European Committee on Antimicrobial Susceptibility Testing*. 2016; Available from: http://www.eucast.org/.
79. Kahlmeter, G., *The 2014 Garrod Lecture: EUCAST – are we heading towards international agreement?* Journal of Antimicrobial Chemotherapy, 2015.
80. Rodloff, A., et al., *Susceptible, Intermediate, and Resistant – The Intensity of Antibiotic Action.* Dtsch Arztebl Int, 2008. **105**(39): p. 657-62.
81. Matuschek, E., D.F. Brown, and G. Kahlmeter, *Development of the EUCAST disk diffusion antimicrobial susceptibility testing method and its implementation in routine microbiology laboratories.* Clin Microbiol Infect, 2014. **20**(4): p. O255-66.
82. Wisniewski, J.R., et al., *Universal sample preparation method for proteome analysis.* Nat Methods, 2009. **6**(5): p. 359-62.
83. Wisniewski, J.R. and M. Mann, *Consecutive proteolytic digestion in an enzyme reactor increases depth of proteomic and phosphoproteomic analysis.* Anal Chem, 2012. **84**(6): p. 2631-7.

84. Rappsilber, J., Y. Ishihama, and M. Mann, *Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics.* Anal Chem, 2003. **75**(3): p. 663-70.
85. Schwartz, D.C. and C.R. Cantor, *Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis.* Cell, 1984. **37**(1): p. 67-75.
86. National Center for Emerging and Zoonotic Infectious Diseases (NCEZID), D.o.F., Waterborne, and Environmental Diseases (DFWED). *Standard Operating Procedure for PulseNet PFGE of Escherichia coli O157:H7, Escherichia coli non-O157 (STEC), Salmonella serotypes, Shigella sonnei and Shigella flexneri*. 2013 [cited 2016 15.12.2016]; Available from: https://www.cdc.gov/pulsenet/pdf/ecoli-shigella-salmonella-pfge-protocol-508c.pdf.
87. Vizcaíno, J.A., et al., *2016 update of the PRIDE database and its related tools.* Nucleic Acids Res, 2016. **44**(Database issue): p. D447-56.
88. Keseler, I.M., et al., *EcoCyc: a comprehensive database of Escherichia coli biology.* Nucleic Acids Res, 2011. **39**(Database issue): p. D583-90.
89. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.
90. Magrane, M., *UniProt Knowledgebase: a hub of integrated protein data.* Database (Oxford), 2011. **2011**: p. bar009.
91. Szklarczyk, D., et al., *STRING v10: protein-protein interaction networks, integrated over the tree of life.* Nucleic Acids Res, 2015. **43**(Database issue): p. D447-52.
92. EUCAST. *Routine and extended internal quality control for MIC determination and disk diffusion as recommended by EUCAST*. 2017; Available from: http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/QC/v_7.0_EUCAST_QC_tables_routine_and_extended_QC.pdf.
93. EUCAST. *EUCAST Clinical Breakpoint tables*. 2016; Available from: http://www.eucast.org/clinical_breakpoints/.
94. Donczew, R., J. Zakrzewska-Czerwinska, and A. Zawilak-Pawlik, *Beyond DnaA: the role of DNA topology and DNA methylation in bacterial replication initiation.* J Mol Biol, 2014. **426**(12): p. 2269-82.
95. Fillingame, R.H., C.M. Angevine, and O.Y. Dmitriev, *Mechanics of coupling proton movements to c-ring rotation in ATP synthase.* FEBS Lett, 2003. **555**(1): p. 29-34.
96. Funnell, B.E., T.A. Baker, and A. Kornberg, *In vitro assembly of a prepriming complex at the origin of the Escherichia coli chromosome.* J Biol Chem, 1987. **262**(21): p. 10327-34.
97. Nakanishi-Matsui, M., et al., *Inhibition of F1-ATPase rotational catalysis by the carboxyl-terminal domain of the subunit.* J Biol Chem, 2014. **289**(44): p. 30822-31.
98. Rao, R., J. Pagan, and A.E. Senior, *Directed mutagenesis of the strongly conserved lysine 175 in the proposed nucleotide-binding domain of alpha-subunit from Escherichia coli F1-ATPase.* J Biol Chem, 1988. **263**(31): p. 15957-63.
99. Rowen, L. and A. Kornberg, *Primase, the dnaG protein of Escherichia coli. An enzyme which starts DNA chains.* J Biol Chem, 1978. **253**(3): p. 758-64.
100. Sekiya, M., et al., *Single molecule behavior of inhibited and active states of Escherichia coli ATP synthase F1 rotation.* J Biol Chem, 2010. **285**(53): p. 42058-67.
101. Sengupta, J., R.K. Agrawal, and J. Frank, *Visualization of protein S1 within the 30S ribosomal subunit and its interaction with messenger RNA.* Proc Natl Acad Sci U S A, 2001. **98**(21): p. 11991-6.

102. Siegenthaler, R.K. and P. Christen, *Tuning of DnaK chaperone action by nonnative protein sensor DnaJ and thermosensor GrpE.* J Biol Chem, 2006. **281**(45): p. 34448-56.
103. Studwell-Vaughan, P.S. and M. O'Donnell, *Constitution of the twin polymerase of DNA polymerase III holoenzyme.* J Biol Chem, 1991. **266**(29): p. 19833-41.
104. Yao, N., et al., *DNA structure requirements for the Escherichia coli gamma complex clamp loader and DNA polymerase III holoenzyme.* J Biol Chem, 2000. **275**(15): p. 11440-50.
105. UniProt. *Proteomes - Escherichia coli O18:K1:H7 (strain IHE3034 / ExPEC)*. 2017 [cited 2017; Available from: http://www.uniprot.org/proteomes/UP000002364.
106. UniProt. *Proteomes - Escherichia coli J53*. 2017 February 4, 2017 [cited 2017; Available from: http://www.uniprot.org/proteomes/UP000004552.
107. Vimr, E.R. and F.A. Troy, *Identification of an inducible catabolic system for sialic acids (nan) in Escherichia coli.* J Bacteriol, 1985. **164**(2): p. 845-53.
108. Kalivoda, K.A., S.M. Steenbergen, and E.R. Vimr, *Control of the Escherichia coli sialoregulon by transcriptional repressor NanR.* J Bacteriol, 2013. **195**(20): p. 4689-701.
109. van Albada, S.B. and P.R. Ten Wolde, *Differential affinity and catalytic activity of CheZ in E. coli chemotaxis.* PLoS Comput Biol, 2009. **5**(5): p. e1000378.
110. Cardozo, M.J., et al., *Disruption of chemoreceptor signalling arrays by high levels of CheW, the receptor-kinase coupling protein.* Mol Microbiol, 2010. **75**(5): p. 1171-81.
111. Yang, Y., et al., *Relation between chemotaxis and consumption of amino acids in bacteria.* Mol Microbiol, 2015. **96**(6): p. 1272-82.
112. Berg, H.C., *The rotary motor of bacterial flagella.* Annu Rev Biochem, 2003. **72**: p. 19-54.
113. Aldridge, P., et al., *Flk prevents premature secretion of the anti-σ factor FlgM into the periplasm.* Mol Microbiol, 2006. **60**(3): p. 630-43.
114. Turnbough, C.L. and R.L. Switzer, *Regulation of Pyrimidine Biosynthetic Gene Expression in Bacteria: Repression without Repressors.* Microbiol Mol Biol Rev, 2008. **72**(2): p. 266-300.
115. Zhou, Y., et al., *Direct linking of metabolism and gene expression in the proline utilization A protein from Escherichia coli.* Amino Acids, 2008. **35**(4): p. 711-8.
116. Andreishcheva, E.N. and W.F. Vann, *Gene Products Required for De Novo Synthesis of Polysialic Acid in Escherichia coli K1.* J Bacteriol, 2006. **188**(5): p. 1786-97.
117. Steenbergen, S.M. and E.R. Vimr, *Biosynthesis of the Escherichia coli K1 group 2 polysialic acid capsule occurs within a protected cytoplasmic compartment.* Mol Microbiol, 2008. **68**(5): p. 1252-67.
118. Whitfield, C. and I.S. Roberts, *Structure, assembly and regulation of expression of capsules in Escherichia coli.* Mol Microbiol, 1999. **31**(5): p. 1307-19.
119. Omadjela, O., et al., *BcsA and BcsB form the catalytically active core of bacterial cellulose synthase sufficient for in vitro cellulose synthesis.* Proc Natl Acad Sci U S A, 2013. **110**(44): p. 17856-61.
120. Römling, U. and M.Y. Galperin, *Bacterial cellulose biosynthesis: diversity of operons, subunits, products and functions.* Trends Microbiol, 2015. **23**(9): p. 545-57.
121. Suits, M.D., et al., *Identification of an Escherichia coli O157:H7 heme oxygenase with tandem functional repeats.* Proc Natl Acad Sci U S A, 2005. **102**(47): p. 16955-60.
122. Zhu, M., et al., *Functions of the siderophore esterases IroD and IroE in iron-salmochelin utilization.* Microbiology, 2005. **151**(Pt 7): p. 2363-72.
123. Muller, S.I., M. Valdebenito, and K. Hantke, *Salmochelin, the long-overlooked catecholate siderophore of Salmonella.* Biometals, 2009. **22**(4): p. 691-5.

124. Hong, H., et al., *The outer membrane protein OmpW forms an eight-stranded beta-barrel with a hydrophobic channel.* J Biol Chem, 2006. **281**(11): p. 7568-77.
125. Wu, X.B., et al., *Outer membrane protein OmpW of Escherichia coli is required for resistance to phagocytosis.* Res Microbiol, 2013. **164**(8): p. 848-55.
126. Kumar, M. and V. Sourjik, *Physical map and dynamics of the chaperone network in Escherichia coli.* Mol Microbiol, 2012. **84**(4): p. 736-47.
127. Abouhamad, W.N., et al., *Peptide transport and chemotaxis in Escherichia coli and Salmonella typhimurium: characterization of the dipeptide permease (Dpp) and the dipeptide-binding protein.* Mol Microbiol, 1991. **5**(5): p. 1035-47.
128. Klepsch, M.M., et al., *Escherichia coli peptide binding protein OppA has a preference for positively charged peptides.* J Mol Biol, 2011. **414**(1): p. 75-85.
129. Létoffé, S., P. Delepelaire, and C. Wandersman, *The housekeeping dipeptide permease is the Escherichia coli heme transporter and functions with two optional peptide binding proteins.* Proc Natl Acad Sci U S A, 2006. **103**(34): p. 12891-6.
130. Heatwole, V.M. and R.L. Somerville, *The tryptophan-specific permease gene, mtr, is differentially regulated by the tryptophan and tyrosine repressors in Escherichia coli K-12.* J Bacteriol, 1991. **173**(11): p. 3601-4.
131. Shao, Z., R.T. Lin, and E.B. Newman, *Sequencing and characterization of the sdaC gene and identification of the sdaCB operon in Escherichia coli K12.* Eur J Biochem, 1994. **222**(3): p. 901-7.
132. Yee, M.C., V. Horn, and C. Yanofsky, *On the role of helix 0 of the tryptophan synthetase alpha chain of Escherichia coli.* J Biol Chem, 1996. **271**(25): p. 14754-63.
133. Yakhnin, H., et al., *The trp RNA-binding attenuation protein of Bacillus subtilis regulates translation of the tryptophan transport gene trpP (yhaG) by blocking ribosome binding.* J Bacteriol, 2004. **186**(2): p. 278-86.
134. Strych, U. and M.J. Benedik, *Mutant Analysis Shows that Alanine Racemases from Pseudomonas aeruginosa and Escherichia coli Are Dimeric.* J Bacteriol, 2002. **184**(15): p. 4321-5.
135. Kuczynska-Wisnik, D., E. Matuszewska, and E. Laskowska, *Escherichia coli heat-shock proteins IbpA and IbpB affect biofilm formation by influencing the level of extracellular indole.* Microbiology, 2010. **156**(Pt 1): p. 148-57.
136. Brown, G., et al., *Functional and Structural Characterization of Four Glutaminases from Escherichia coli and Bacillus subtilis.* Biochemistry, 2008. **47**(21): p. 5724-35.
137. van Heeswijk, W.C., H.V. Westerhoff, and F.C. Boogerd, *Nitrogen Assimilation in Escherichia coli: Putting Molecular Data into a Systems Perspective.* Microbiol Mol Biol Rev, 2013. **77**(4): p. 628-95.
138. Borges-Walmsley, M.I., et al., *Identification of oligomerization and drug-binding domains of the membrane fusion protein EmrA.* J Biol Chem, 2003. **278**(15): p. 12903-12.
139. Lomovskaya, O., K. Lewis, and A. Matin, *EmrR is a negative regulator of the Escherichia coli multidrug resistance pump EmrAB.* J Bacteriol, 1995. **177**(9): p. 2328-34.
140. Patzer, S.I., et al., *The colicin G, H and X determinants encode microcins M and H47, which might utilize the catecholate siderophore receptors FepA, Cir, Fiu and IroN.* Microbiology, 2003. **149**(Pt 9): p. 2557-70.
141. Nobelmann, B. and J.W. Lengeler, *Molecular analysis of the gat genes from Escherichia coli and of their roles in galactitol transport and metabolism.* J Bacteriol, 1996. **178**(23): p. 6790-5.
142. Mainguet, S.E., et al., *A reverse glyoxylate shunt to build a non-native route from C4 to C2 in Escherichia coli.* Metab Eng, 2013. **19**: p. 116-27.

143. Yankovskaya, V., et al., *Architecture of succinate dehydrogenase and reactive oxygen species generation.* Science, 2003. **299**(5607): p. 700-4.
144. Yu, B.J., et al., *sucAB and sucCD are mutually essential genes in Escherichia coli.* FEMS Microbiol Lett, 2006. **254**(2): p. 245-50.
145. Martínez-Gómez, K., et al., *New insights into Escherichia coli metabolism: carbon scavenging, acetate metabolism and carbon recycling responses during growth on glycerol.* Microb Cell Fact, 2012. **11**: p. 46.
146. Kornberg, H.L., *Routes for fructose utilization by Escherichia coli.* J Mol Microbiol Biotechnol, 2001. **3**(3): p. 355-9.
147. Lepore, B.W., et al., *Ligand-gated diffusion across the bacterial outer membrane.* Proc Natl Acad Sci U S A, 2011. **108**(25): p. 10121-6.
148. Black, P.N., et al., *Cloning, sequencing, and expression of the fadD gene of Escherichia coli encoding acyl coenzyme A synthetase.* J Biol Chem, 1992. **267**(35): p. 25513-20.
149. Yang, S.Y., *Location of the fadBA operon on the physical map of Escherichia coli.* J Bacteriol, 1991. **173**(23): p. 7405-6.
150. Bermejo, G.A., et al., *Ligand-Free Open–Closed Transitions of Periplasmic Binding Proteins: the Case of Glutamine-Binding Protein.* Biochemistry, 2010. **49**(9): p. 1893-902.
151. Dubey, A.K., et al., *CsrA Regulates Translation of the Escherichia coli Carbon Starvation Gene, cstA, by Blocking Ribosome Access to the cstA Transcript.* J Bacteriol, 2003. **185**(15): p. 4450-60.
152. Ganea, C. and K. Fendler, *Bacterial transporters: charge translocation and mechanism.* Biochim Biophys Acta, 2009. **1787**(6): p. 706-13.
153. Abramson, J., et al., *The structure of the ubiquinol oxidase from Escherichia coli and its ubiquinone binding site.* Nat Struct Biol, 2000. **7**(10): p. 910-7.
154. Gutierrez, C. and J.C. Devedjian, *Osmotic induction of gene osmC expression in Escherichia coli K12.* J Mol Biol, 1991. **220**(4): p. 959-73.
155. Boulanger, A., et al., *Multistress Regulation in Escherichia coli: Expression of osmB Involves Two Independent Promoters Responding either to σ(S) or to the RcsCDB His-Asp Phosphorelay.* J Bacteriol, 2005. **187**(9): p. 3282-6.
156. Gutierrez, C., S. Gordia, and S. Bonnassie, *Characterization of the osmotically inducible gene osmE of Escherichia coli K-12.* Mol Microbiol, 1995. **16**(3): p. 553-63.
157. Lennon, C.W., et al., *Folding optimization in vivo uncovers new chaperones.* J Mol Biol, 2015. **427**(18): p. 2983-94.
158. Schiefner, A., et al., *Cation-pi interactions as determinants for binding of the compatible solutes glycine betaine and proline betaine by the periplasmic ligand-binding protein ProX from Escherichia coli.* J Biol Chem, 2004. **279**(7): p. 5588-96.
159. Zhang, Y., M. Morar, and S.E. Ealick, *Structural Biology of the Purine Biosynthetic Pathway.* Cell Mol Life Sci, 2008. **65**(23): p. 3699-724.
160. Georgopapadakou, N.H., *Beta-lactamase inhibitors: evolving compounds for evolving resistance targets.* Expert Opin Investig Drugs, 2004. **13**(10): p. 1307-18.
161. Brown, N.G., et al., *Structural and Biochemical Evidence That a TEM-1 β-Lactamase N170G Active Site Mutant Acts via Substrate-assisted Catalysis.* J Biol Chem, 2009. **284**(48): p. 33703-12.
162. Russo, T.A. and J.R. Johnson, *Medical and economic impact of extraintestinal infections due to Escherichia coli: focus on an increasingly important endemic problem.* Microbes Infect, 2003. **5**(5): p. 449-56.
163. Stoll , B.J., et al., *Changes in Pathogens Causing Early-Onset Sepsis in Very-Low-Birth-Weight Infants.* New England Journal of Medicine, 2002. **347**(4): p. 240-247.

164. Logue, C.M., et al., *Genotypic and Phenotypic Traits That Distinguish Neonatal Meningitis-Associated Escherichia coli from Fecal E. coli Isolates of Healthy Human Hosts.* Appl Environ Microbiol, 2012. **78**(16): p. 5824-30.
165. Jorge, I., et al., *Statistical Model to Analyze Quantitative Proteomics Data Obtained by (18)O/(16)O Labeling and Linear Ion Trap Mass Spectrometry: Application to the Study of Vascular Endothelial Growth Factor-induced Angiogenesis in Endothelial Cells.* Mol Cell Proteomics, 2009. **8**(5): p. 1130-49.
166. Wood, T.K., et al., *Motility influences biofilm architecture in Escherichia coli.* Appl Microbiol Biotechnol, 2006. **72**(2): p. 361-7.
167. Da Re, S. and J.M. Ghigo, *A CsgD-Independent Pathway for Cellulose Production and Biofilm Formation in Escherichia coli.* J Bacteriol, 2006. **188**(8): p. 3073-87.
168. Bokranz, W., et al., *Expression of cellulose and curli fimbriae by Escherichia coli isolated from the gastrointestinal tract.* J Med Microbiol, 2005. **54**(Pt 12): p. 1171-82.
169. Li, M. and G.L. Hazelbauer, *Cellular Stoichiometry of the Components of the Chemotaxis Signaling Complex.* J Bacteriol, 2004. **186**(12): p. 3687-94.
170. Skaar, E.P., *The Battle for Iron between Bacterial Pathogens and Their Vertebrate Hosts.* PLoS Pathog, 2010. **6**(8).
171. Cherayil, B.J., *The role of iron in the immune response to bacterial infection.* Immunol Res, 2011. **50**(1): p. 1-9.
172. Russo, T.A., et al., *The Siderophore Receptor IroN of Extraintestinal Pathogenic Escherichia coli Is a Potential Vaccine Candidate.* Infect Immun, 2003. **71**(12): p. 7164-9.
173. Kalivoda, K.A., et al., *Regulation of Sialic Acid Catabolism by the DNA Binding Protein NanR in Escherichia coli.* J Bacteriol, 2003. **185**(16): p. 4806-15.
174. Arora, A., et al., *Refolded outer membrane protein A of Escherichia coli forms ion channels with two conductance states in planar lipid bilayers.* J Biol Chem, 2000. **275**(3): p. 1594-600.
175. Samsudin, F., et al., *OmpA: A Flexible Clamp for Bacterial Cell Wall Attachment.* Structure, 2016. **24**(12): p. 2227-2235.
176. Oliver, D.B., et al., *Azide-resistant mutants of Escherichia coli alter the SecA protein, an azide-sensitive component of the protein export machinery.* Proc Natl Acad Sci U S A, 1990. **87**(21): p. 8227-31.
177. Achtman, M., et al., *Six widespread bacterial clones among Escherichia coli K1 isolates.* Infect Immun, 1983. **39**(1): p. 315-35.
178. San Millan, A., et al., *Positive selection and compensatory adaptation interact to stabilize non-transmissible plasmids.* Nat Commun, 2014. **5**: p. 5208.
179. Simione, F.P. *Cryopreservation Manual, Nalge Nunc International*. 1998 [cited 2017 26.09.2017]; Available from: https://www.vanderbilt.edu/viibre/CryoPreservationManual.pdf.
180. American Type Culture Collection. *Reference Strains: How many passages are too many?* 2003 [cited 2017 26.09.2017]; Available from: https://www.atcc.org/~/media/PDFs/Technical%20Bulletins/tb06.ashx.

Appendix

Two documents containing supplementary materials have been attached, one PDF file and one Excel document.

The PDF file Supplementary materials.pdf contains Figure S1, Figure S2, and the resulting sequences from the Sanger sequencing of the gene fragments used in MLST.

The supporting proteomic file Supplementary tables.xlsx contains Tables S1 – S10.