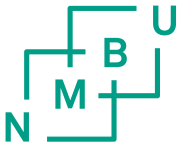

U
B
M
N

Acknowledgements

This master thesis was performed at the Norwegian University of Life Science, Department of Chemistry, Biotechnology and Food Science, with Professor Knut Rudi as supervisor and Ph.D. student Anuradha Ravi as secondary supervisor.

First, I want to thank Professor Knut Rudi for giving me the opportunity to study antibiotic resistance through this thesis, which is my great interest in science. Thank you for helping me planning my laboratory work and for performing the PLS analyses. Your enthusiasm and positive attitude encouraged me throughout the thesis. I want to thank Anuradha for helping me executing my laboratory work and for performing the metagenome assembly. I could not have asked for a better supervisor. In addition, I want to thank Jane for always being so kind and helpful. And not at least, thanks to all in the Microbial Diversity group for making me feel welcome. You made me look forward to each day in the lab.

Furthermore, I want to thank the staff at the hospitals in Boston, Chicago and Evanston for providing us with this unique sample material, and the staff at Genetic Analysis for performing the DNA extraction.

Finally, a special thanks to my boyfriend Dennis. Your patience and support have meant a lot to me - I could not have managed this without you. Thanks to my parents, for always believe in me and for always being supportive.

Ås, May 2015

Eva Lena F. Estensmo

Sammendrag

Premature barn er mer utsatt for kolonisering av opportunistiske patogener som kan resultere i nekrotiserende enterokolitt (NEC). Denne tilstanden er forbundet med høy morbiditet og mortalitet og krever ofte mye antibiotikabehandling. Antibiotikabehandlinger kan forstyrre bakteriefloraen, og selektere for antibiotikaresistente bakterier. Kunnskapen om utbredelsen av disse bakteriene og antibiotikaresistensgener i den premature tarmfloraen og deres bidrag i utviklingen av NEC er begrenset. Derfor undersøker denne studien utbredelsen og persistensen av integroner - genetiske elementer som inneholder antibiotikaresistensgener - i den fekale bakteriefloraen fra en kohort av premature barn med og uten NEC. Gjennom kvantitativ PCR og metagenom shotgun sekvensering, oppdaget vi en høyere frekvens av integroner, persistens av integroner hos flere pasienter og en rekke antibiotikaresistensgener hos premature barn med NEC. Det er derfor grunn til å tro at integroner kan være assosiert med NEC. I tillegg, taksonomisk klassifisering gjennom 16S rRNA sekvensering viste en signifikant høyere forekomst av *Escherichia coli* i premature barn med NEC. Denne bakterien har tidligere vært assosiert med NEC i andre studier. Imidlertid, denne studien er den første vi vet om til å assosiere integroner med NEC. Den danner derfor et grunnlag for videre forståelse om den premature tarmfloraen som et reservoar for antibiotikaresistensgener og integroner, da dette kan spille en viktig rolle i patogenesen av NEC.

Abstract

Preterm infants are more susceptible to colonization by opportunistic pathogens that may result in necrotizing enterocolitis (NEC). This condition is associated with high morbidity and mortality and often requires excessive antibiotic treatment. Antibiotic treatments can disturb the microbiota and select for antibiotic resistant bacteria. The knowledge about the prevalence of these bacteria and antibiotic resistance genes in the preterm gut microbiota and their contribution in the development of NEC is limited. Therefore, this study investigates the prevalence and persistence of integrons - genetic elements harboring antibiotic resistance genes - in the fecal microbiota from a cohort of preterm infants with and without NEC. Through quantitative PCR and shotgun metagenome sequencing, we detected a higher abundance of integrons, persistence of integrons in several patients and a variety of antibiotic resistance genes in the preterm infants with NEC. Therefore, it is reason to believe that integrons can be associated with NEC. In addition, taxonomic classification through 16S rRNA sequencing revealed a significantly higher abundance of *Escherichia coli* in the preterm infants with NEC. This bacterium has previously been associated with NEC in other studies. However, this study is of our knowledge the first to associate integrons with NEC. It therefore provides a foundation for further understanding about the preterm gut microbiota as a reservoir for antibiotic resistance genes and integrons, as this may play an important role in the pathogenesis of NEC.

Abbreviations

BLAST	-	Basic Local Alignment Search Tool
bp	-	base pair
ddNTP	-	dideoxy nucleotide triphosphate
DNA	-	Deoxyribonucleic acid
dNTP	-	deoxy nucleotide triphosphate
dsDNA	-	double stranded DNA
GI	-	Gastrointestinal
HRM	-	High resolution melting
*int*1	-	Class 1 integron integrase gene
MG-RAST	-	Metagenomics-Rapid Annotation using Subsystem Technology
NEC	-	Necrotizing enterocolitis
ORF	-	Open reading frame
OTU	-	Operational taxonomic unit
PCoA	-	Principal coordinates analysis
PCR	-	Polymerase chain reaction
QIIME	-	Quantitative Insights Into Microbial Ecology
qPCR	-	quantitative polymerase chain reaction
RNA	-	Ribonucleic acid
rRNA	-	ribosomal ribonucleic acid
SEM	-	Standard error of mean
ssDNA	-	single stranded DNA

Table of Content

1. Introduction

Necrotizing enterocolitis (NEC) is a gastrointestinal disease that occurs in 10 % of preterm infants and has a mortality rate close to 30 % (Morrow et al., 2013; Neu & Walker, 2011). The high mortality rate is due to the complex pathogenicity of NEC. There is not revealed a common mechanism that causes NEC, only several factors that contributes to it. This lack of knowledge makes treatment of NEC complicated.

Preterm infants are infants born before 37 completed weeks of gestation (Goldenberg, Culhane, Iams, & Romero, 2008). They are more vulnerable to infections due to their immaturity; both their organs and their immune system may be underdeveloped. Therefore, they often require excessive antibiotic treatments in order to survive. Antibiotic treatments may cause dysbiosis of the microbiota and select for antibiotic resistant bacteria. How these antibiotic resistant bacteria affect the infants and how they spread their antibiotic resistance genes is not fully understood. In this thesis, I will therefore address the connection between antibiotic resistance genes and integrons with NEC in preterm infants.

1.1 The Human Gut Microbiota

The microbiota is described as a community of microbes that resides in a specific habitat. The human gut microbiota therefore includes all microorganisms present in the gastrointestinal (GI) tract. The human GI tract consists of the upper GI tract and the lower GI tract. The lower GI tract consists of the small intestine and the large intestine. The vast majority of the human gut microbiota is located in the large intestine. The population of these bacteria is a debated topic, but is thought to be ten times higher than the number of human cells(Palmer, Bik, DiGiulio, Relman, & Brown, 2007; Qin et al., 2010). These bacteria usually lives in a symbiotic relationship with their hosts - a mutualistic or commensalistic relationship that is either beneficial or leaves the host unaffected (Collins, 2014; Collins, Surette, & Bercik, 2012). However, some bacteria present are opportunistic pathogens.

Revealing the impact of the relationship between the human gut microbiota and human health is of great significance. This has led to several global projects such as the Human Microbiome Project (HMP) (HMP, 2012; Turnbaugh et al., 2007; Zankari et al., 2012) and MetaHit (Qin et al., 2010). These projects have contributed to a better understanding of the human gut microbiota, particularly regarding to its function in health and disease.

1.1.1 Adult Gut Microbiota

The development and composition of the human gut microbiota is important to human health. It prevents pathogens from colonizing our gut, provides nutrients and energy from compounds the body cannot utilize and triggers the immune system (Candela et al., 2008; Collins, 2014; Round & Mazmanian, 2009; Sommer & Backhed, 2013). The adult gut microbiota is dominated by bacteria belonging to the phyla *Bacteroidetes* and *Firmicutes*. In addition, the gut microbiota is believed to be uniquely composed for each individual, but features of this composition can be shared throughout the human population. Some scientists believe that a core, a number of identical bacteria, is shared among individuals (Collins et al., 2012; Salonen, Salojarvi, Lahti, & de Vos, 2012; Sekelja, Berget, Naes, & Rudi, 2011). Others have classified the microbiota into enterotypes based on states of bacteriological ecosystems in the gut (Arumugam et al., 2011).

1.1.2 Infant Gut Microbiota

The human gut microbiota dramatically changes from birth to adulthood. The colonization of the infant gut is a remarkable process, changing from almost sterile to entirely colonized in just a few days (Palmer et al., 2007; Sommer & Backhed, 2013). Initially, the infant gut was believed to be completely sterile (Yunwei Wang et al., 2009), but this theory has been challenged with the detection of microorganisms in the placenta (Aagaard et al., 2014) and the meconium (Jiménez et al., 2008).

Immediately after birth, the infant gut is colonized with facultative anaerobic bacteria, such as *Streptococcus* and *Escherichia coli* (Wallace et al., 2011). These bacteria deplete the oxygen, creating a more anaerobic environment in the gut. This allows anaerobe bacteria like *Bacteroides*, *Bifidobacteria*, and *Lactobacilli* to colonize (Palmer et al., 2007). From this stage, the infant gut microbiota is generally dominated by *Bifidobacteria*. During two years of life, the microbiota changes further to be similar to that of the adult state (Avershina et al., 2013; Sommer & Backhed, 2013).

The colonization and the development of the infant gut are dependent on several factors, but two of the most important is mode of delivery and type of feeding. Vaginal delivery exposes the infant to the mother's fecal and vaginal microbiota, whereas caesarian section delivery exposes the infant to the skin microbiota and the environmental microbiota (Palmer et al., 2007; Sommer & Backhed, 2013). Further, the gut microbiota in breast-fed infants are

dominated by *Bifidobacteria* (Avershina et al., 2013; Marques et al., 2010), whereas *Enterobacteria* and *Clostridia* dominates in formula-fed infants.

1.1.3 Preterm Gut Microbiota

The microbiota in preterm infants is less complex than of term infants and the colonization by beneficial bacteria is delayed (Westerbeek et al., 2006). In addition to early colonization dominated by *Firmicutes*, preterm infants also tend to be colonized with *Proteobacteria* (Schwiertz et al., 2003; Y. Wang et al., 2009). Schwiers et al. (2003) studied fecal samples from 29 preterm infants in their first weeks of life and showed that the microbiota between the preterm infants became both more stable and similar to each other compared to term infants over time. In addition, their microbiota was different from the microbiota in full-term infants. This may also be due to hospitalization as preterm infants are more exposed to the hospital environment, creating a similar microbiota between the preterm infants.

1.2 Necrotizing Enterocolitis

Necrotizing enterocolitis is a serious gastrointestinal disease primarily occurring in preterm infants (Alexander, Northrup, & Bizzarro, 2011). NEC is caused by one or a combination of factors such as immature GI tract, damage of the intestinal mucosal layer and microbial alterations. (Thompson & Bizzarro, 2008). There might be other factors contributing to or causing NEC that is still not studied, which would be important due to prevention and treatment of NEC.

A variety of symptoms is connected to the development of NEC. The most common initial symptoms includes feeding intolerance, abdominal distension and bloody stools (Neu & Walker, 2011). The severity of the disease can rapidly deteriorate, and medical or surgical treatment is necessary. NEC can be classified into three stages based on the severity of the disease, which can be described as benign, moderate and severe (Bell et al., 1978).

The microbial colonization is thought to be important in the pathogenesis of NEC. Colonization of potentially pathogenic bacteria can lead to bacterial invasion of the intestinal wall, translocation and inflammation (Panigrahi, 2006). This inflammation of the intestinal wall may lead to NEC (Westerbeek et al., 2006). Several pathogens have been associated with NEC (Hunter, Upperman, Ford, & Camerini, 2008; McMurtry et al., 2015; Neu & Walker, 2011). However, there is still no bacterial species that has been defined as definitively causative of NEC.

It has been proposed that preterm infants with NEC have a reduced microbial diversity than healthy preterm infants (Y. Wang et al., 2009). Wang et al. (2009) studied twenty preterm infants with and without NEC. The study revealed that preterm infants are mainly colonized with *Proteobacteria*, *Firmicutes*, *Bacteroidetes* and *Fusobacteria*, whereas preterm infants with NEC are only colonized with *Proteobacteria* and *Firmicutes*. This study also included pairs of preterm twins - one with NEC and the other without - revealing several differences in colonization patterns between them. The twin with NEC had a more similar microbiota to the other infants with NEC than of the other twin, even if they were monozygotic twins. Other studies could not detect a reduction in microbial diversity between healthy preterm infants and infants with NEC (Mai et al., 2011). The study of Mai et al. (2011) compared the microbiota between 9 preterm infants with NEC to 9 control infants, and did not detect a reduction in microbial diversity between the two groups.

Today, the most common medical treatment of NEC is the use of broad-spectrum antibiotics, trying to inhibit pathogenic bacteria from causing infection. Alexander et al. (2011) studied 124 cases of NEC matched with 248 controls from 2000 to 2008. They found that increased duration of antimicrobial exposure in infants without sepsis increased the probability of developing NEC (Alexander, Northrup, & Bizzarro, 2011). An explanation for this is that the use of antibiotics can delay the colonization process in the gut.

A solution to be able to treat NEC more efficiently can be to combine antibiotics and probiotics. Promising studies found that the development of NEC may be prevented with supplementation of probiotics; foods or supplements containing live microorganisms promoting a healthy gut microbiota (Lozupone, Stombaugh, Gordon, Jansson, & Knight, 2012). Researchers found that probiotics reduced the incidence and severity of NEC in preterm infants, both by strengthening the immune system and by inhibit growth of pathogenic bacteria (Alfaleh, Anabrees, Bassler, & Al-Kharfi, 2011). Their study included preterm infants where enternal administration of probiotics was compared against a placebo group.

1.3 Acquired Antibiotic Resistance

The first antibiotic was discovered in 1928 by Alexander Fleming, which was penicillin isolated from the mould *Penicillium* (Fleming, 1929). Since then, other different antibiotics have been discovered such as tetracycline, streptomycin and chloramphenicol. Many of these have been further modified to increase their effect. Antibiotics are chemical substances

produced by microorganisms that in low concentrations kill or prevent growth of other microorganisms. Antibiotics that kills bacteria are called bactericides (Kohanski, Dwyer, & Collins, 2010), while those that prevent their growth are called bacteriostatic. Different antibiotics attack the bacteria in various ways. The most common antibiotics inhibits the synthesis of RNA, DNA, the cell wall or proteins (Kohanski et al., 2010).

The discovery of antibiotics was a revolution in the medical industry, but their effect in the future was uncertain. In order for the bacteria to survive, they developed mechanisms making them resistant to the different antibiotics. Initially, some bacteria developed resistance to a single antibiotic, but later they developed resistance to multiple antibiotics. These multidrug resistant bacteria are hard to combat with antibiotics and normal infections can become life-threatening diseases. This issue has become an increasing threat to human health and the emerge of antibiotic resistant bacteria are one among the greatest challenges throughout the world (Hu et al., 2013).

Antibiotics are of great value and importance, but they can have a negative impact on the human microbiota. Antibiotics can both disturb the normal flora and contribute to antibiotic resistance by eliminating most of the non-resistant bacteria. This leads to a selection pressure leaving behind a greater proportion of resistant bacteria. This is a issue due to disease and spread of resistance genes, making the human gut microbiota a possible reservoir for antibiotic resistance genes (Salyers, Gupta, & Wang, 2004). Salyers et al. (2004) proposed the reservoir hypothesis stating that both bacteria that reside in the human colon, and those that is passing through, may transfer or acquire resistance genes among themselves and in addition have the possibility to transfer these further to new environments or hosts.

Acquired antibiotic resistance occurs when bacteria achieves resistance towards an antibiotic they previously were susceptible for. This happens through horizontal gene transfer, which is transfer of genes outside reproduction. The bacteria acquire antibiotic resistance when resistance genes are incorporated and expressed. These resistance genes can change the cell surface proteins which inhibits the antibiotic to attach, create efflux pumps that actively pumps incoming antibiotics out of the cell, produce antibiotic degrading enzymes, and/or change the permeability of the cell wall preventing the antibiotics to enter the cell (Blair, Webber, Baylay, Ogbolu, & Piddock, 2014; Cox & Wright, 2013). Genes coding for these traits can be located on mobile genetic elements.

1.3.1 Mobile Genetic Elements and Transfer Mechanisms

Mobile genetic elements are DNA segments found on the bacterial genome that mediates movement within genomes or between bacterial cells (Frost, Leplae, Summers, & Toussaint, 2005). The most common mobile genetic elements are plasmids and transposons. Plasmids are circular DNA molecules that replicates independently of the chromosome. Transposons are genetic elements that can change their position in the genome. In chromosomes or plasmids exists conjugative transposons which are self-transmittable elements that both codes for their own conjugation and contributes to transmission of antibiotic resistance genes. (Mazel, 2006). Another genetic element is integrating conjugative elements that both can be mobile and integrate into the chromosome. Mobile genetic elements can be transferred from one bacterium to another through horizontal gene transfer mediated by transduction, transformation and conjugation.

Transduction is transmission of genes via a bacteriophage, a virus that only infects bacteria, where the phage transfers resistance genes from one bacterium and injects it into another. This is thought to be a process occurring as a consequence of excision errors of bacteriophage DNA from the bacteria genome (Huddleston, 2014). Because of the bacteriophages specificity, transmission of resistance genes are restricted to closely related species and often limited to small DNA fragments due to the limited size of the phage genomes (Gaustad, 2001).

DNA fragments from the environment can be accessed through transformation, a mechanism allowing admission of free DNA from the environment. The transformation can happen within the same species or between species and genera, and allows the bacteria to absorb larger DNA fragments from the environment and incorporate it into their genome. To be able to absorb the DNA, the bacteria have to be in a state that allows DNA admission; they have to be competent (Johnsborg, Eldholm, & Håvarstein, 2007). In some species, competence may happen simultaneously in larger bacterial groups through quorum sensing, that is communication between bacteria through signal molecules. Transformation is widely distributed among bacteria, and many reasons for transformation has been suggested such as nutrition needs, chromosome repair and creating genetic diversity (Johnston, Martin, Fichant, Polard, & Claverys, 2014).

Another mechanism contributing to genetic diversity is conjugation, which allows transfer of genetic material from one bacterium to another by direct contact. When the cells have made

contact, DNA exchange usually happens via a pore or a pilus that allows transfer of DNA from the recipient cell to the donor. Transfer of these elements requires several genes that ensures DNA mobilization and mating pair formation, and these genes can be encoded by plasmids or by integrating conjugative elements in the chromosome (Cabezón, Ripoll-Rozada, Peña, de la Cruz, & Arechaga, 2014). Further, the transfer of DNA is mediated by secretion systems that also are involved in transport of virulence factors. Plasmids and transposons are the most frequently genetic elements that are transferred through conjugation.

1.3.2 Integrons

Integrons are DNA elements that functions as recombination platforms where open reading frames (ORFs) are incorporated, rearranged and expressed (Labbate, Case, & Stokes, 2009; Mazel, 2006). The ORFs are incorporated in cassettes recognized by the recombination platform that captures the gene cassette. The gene cassettes are mobile making them able to move from one integron to another, enabling the possibility for one integron to contain several cassettes. The cassettes lack a promoter, making the expression of the genes dependent of the promoter in the integron (Pc). In addition, the integron contains a gene coding for an integrase (*int*I) and a primary recombination site (attI) (Gillings, 2014; Labbate et al., 2009; Mazel, 2006). A site specific recombination (attC) between the attachment site and the recombination site of the resistance gene allows incorporation and ensures the expression of the resistance gene through the integron promoter (Figure 1.1). The integration of new cassettes happens closest to the promoter, giving them higher expression of their resistance genes.

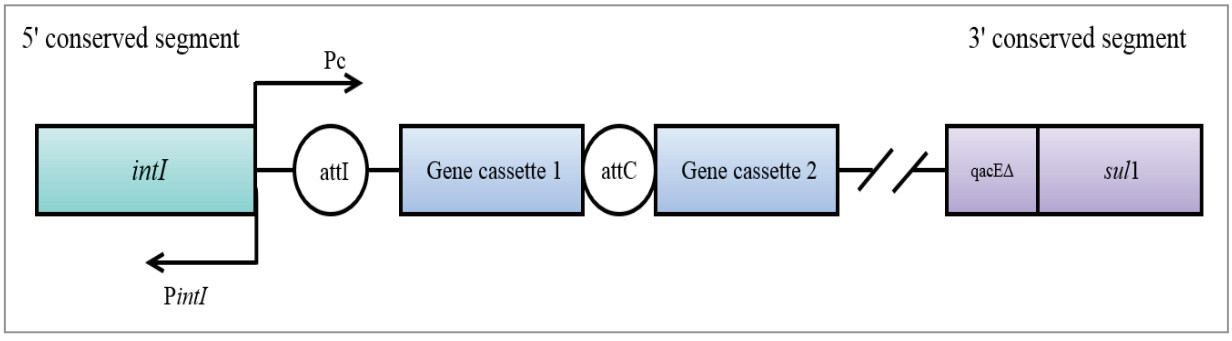


Figure 1.1 The structure of a class 1 integron and its gene cassettes. The class 1 integron also contains a 5' and a 3' conserved segment (Mazel, 2006). The 3' segment contains genes encoding resistance against quaternary ammonium compounds (*qacE*Δ) and sulfonamide (*sul*1) (Estensmo, this thesis).

Integrons are commonly divided into classes, and super integrons are distinguished from integrons. Super integrons are different from integrons, mainly because it contains a large number of gene cassettes with high identity between their recombination sites. The super integron is not mobile as it is located on the chromosome (Mazel, 2006). Mazel (2006) divides the integrons into different classes according to the sequence of the integrase, resulting in five classes of integrons: Class 1 integrons are associated with transposons derived from the transposon Tn402, whereas class 2 integrons are associated with derivatives from the transposon Tn7. Class 3 integrons are believed to be located on a plasmid, whereas class 4 and 5 has been defined according to their contribution to the development of trimethoprim resistance in species of *Vibrio* (Mazel, 2006). Gillings (2014) divides the integrons into three classes: class 1 and 2 similar to that of Mazel (2006), while class 3 is described as similar to the class 1 integron but with a less active integrase.

The activity of the integron can be enhanced through a stress response that is induced when larger amounts of single stranded DNA (ssDNA) is present in the cell, called the SOS response. This response is a widespread regulatory network that induces DNA repair events (Guerin et al., 2009), since ssDNA is associated with DNA breakage. Accumulation of ssDNA usually happens during replication of damaged DNA, but also happens during horizontal gene transmission. Guerin et al (2009) detected that the transcriptional repressor governing the SOS response overlapped with the promoter sequence of the integron. This indicates that the SOS response is related to the activity of the integrase. Further, Baharoglu et al. (2013) demonstrated that the integrase is regulated by the SOS response by measuring the activity of the integrase from multidrug resistant integrons and super integrons. The activity of the integrase was induced after mitomycin C treatment - an agent that destroys DNA and induces SOS. Therefore, due to a more active integrase, the integron can more easily integrate new resistance cassettes.

Integrons were first discovered in pathogenic bacteria associated with antibiotic resistance, but they are generally ancient elements developed through evolution (Labbate et al., 2009). The evolution of class 1 integrons from mobile genetic elements created an effective, easily transferrable element accumulating resistance genes from the environment. The class 1 integron have been found to be most active in accumulating new gene cassettes, and is now a common element in pathogens (Huddleston, 2014). Multidrug resistant integrons has been isolated from transposons involved in rapid development of antibiotic resistance in gram negative bacteria (Baharoglu, Garriss, & Mazel, 2013). This possesses a threat to the human

gut microbiota due to the large population size and high cell density. In such environments, the integron may spread, integrate new cassettes with antibiotic resistance genes and participate in new rearrangements with mobile elements.

1.4 Nucleic Acid Quantification and Sequencing

Nucleic acid based technologies give the ability to perform culture-independent analysis of a microbial community. These technologies are less time consuming and generate more data. They require DNA isolation from a sample, which include separation of the DNA from other components of the cell. The DNA yield can be quantified and specific regions can be amplified through polymerase chain reaction (PCR). This may be necessary if sequencing of the target DNA is desired.

Sequencing improved considerably about ten years ago, when next generation sequencing technologies were developed. Before then, first generation sequencing technology - referred to as Sanger sequencing - was the used technology. Next generation sequencing gave an improvement in sample preparation, the number of sequencing reactions and in detection of sequencing output (van Dijk, Auger, Jaszczyszyn, & Thermes, 2014). Through the past decade, next generation sequencing has been improved and many technologies have been developed. Recently, even third generation sequencing is developed. This technology allows detection of single molecules in real-time.

1.4.1 Quantitative Polymerase Chain Reaction

Quantitative polymerase chain reaction (qPCR) is considered one of the most powerful tools for quantitative nucleic acid analysis (Kubista et al., 2006). The method was developed by Higuchi and his colleagues in the early 90s (Higuchi, Dollinger, Walsh, & Griffith, 1992). Since then, this method has been widely used due to its fast, accurate and easily reproducible culture-independent quantification of microorganisms.

The qPCR is in many ways a refinement of the qualitative PCR. The method still exploits DNA polymerase's ability to amplify many copies of target DNA through denaturation, annealing and extension. In addition, fluorescence labeling is used to measure the number of copies generated through each PCR cycle (Kubista et al., 2006). Different fluorescent reagents used for qPCR is available, such as sequence specific probes and DNA binding dyes. TaqMan is a sequence specific probe that only binds to the target sequence. The probe has a fluorescent label in the 5' end and a quencher attached in the 3' end. The quencher absorbs the

fluorescence when the probe is intact, but the probe fluoresces when it gets cleaved by the polymerase.

The qPCR machine has a cycle threshold (Ct) for detecting the fluorescence signal. The Ct value therefore reflects the amount of template DNA in the sample and the amplification can be studied through a response curve (Kubista et al., 2006). This curve plots the fluorescence (y) against the cycle number (x). For samples amplified with a DNA binding dye it is also beneficial to perform high resolution melt (HRM) analysis to study the melting curve. This curve plots the 2nd derivative (y) against the temperature (x), and shows a rapid drop in fluorescence when the DNA denaturates due to the temperature increase. This is essential to ensure that most of the signal is due to amplification of the target sequence.

1.4.2 First Generation Sequencing

Sanger sequencing was developed by Frederick Sanger and his colleagues in 1977. The sequencing is based on a DNA polymerase that copies DNA by adding deoxyribonucleotide triphosphates (dNTPs) to the 3' end of a primer that is annealed to the template. Extension occurs in 5' to 3' direction by formation of a phosphodiester bond between the 3' hydroxyl (OH) group of the primer and the 5' phosphate group of the incoming dNTP. In addition to dNTP, dideoxyribonucleotide triphosphates (ddNTPs) are added (Sanger, Nicklen, & Coulson, 1977). In these nucleotides, the 3' OH group is replaced by a hydrogen atom. Incorporation of a ddNTP will therefore lead to termination. Both dNTP and ddNTP competes to bind. Therefore, the ratio between them will decide the frequency of termination and the size of the DNA fragment.

Sequencing can be performed with dye terminators. To detect the sequence, the ddNTPs are labeled with different dyes - one for each base (Michael L Metzker, 2005). The labeled fragments are separated by capillary gel electrophoresis. High voltage causes the negatively charged DNA to migrate through the gel towards the positive electrode. On its way, a laser excites the dyes to fluoresce. The fluorescence signals are detected and processed by the sequencing machine, which then reveals the sequence.

1.4.3 Next Generation Sequencing

Illumina sequencing

Illumina sequencing is an efficient tool to analyze the composition of a bacterial community. Illumina is a sequencing by synthesis technology, which is one of the most successful next-generation sequencing platform worldwide. This sequencing strategy is based on reversible color termination, where each base is detected as they are incorporated into a growing DNA strand (van Dijk et al., 2014). The process can be divided into three steps: sample preparation, colony formation and sequencing. During preparation, the DNA is fragmented before the ends are repaired and adenylated. Then, an adapter oligonucleotide is ligated to the ends of the DNA strand before the product is cleaned (Figure 1.2A)

Colony formation happens when these oligonucleotides binds to the chip called a flowcell. This flowcell has many adaptor oligonucleotides attached to it, allowing the adaptor oligonucleotides on the DNA strands to bind. After binding, the DNA will be copied. The end of the copied DNA will then bend to another oligonucleotide on the chip, creating a bridge formation. Amplification of these bridges results in colony generations on the flowcell (Figure 1.2B). After amplification cleaves the reverse strands off, and a sequencing primer is attached to the free end of the DNA before the sequencing starts.

Sequencing happens base by base where all four bases compete to bind, giving very high accuracy. The bases are fluorescence labeled; a fluorochrome is attached to each base creating an unique color. In this way, a laser can excite the fluorochromes and read their color before they are removed. A new base cannot bind before fluorochrome is removed because the fluorochrome blocks the 3`OH end on the growing DNA strand. Several bases binds to the growing DNA and their color signal reveals the DNA sequence (Figure 1.2C). This technology makes it possible to sequence several millions of sequences simultaneously, and gives information about all the bacteria present in the sample - not only those who dominate.

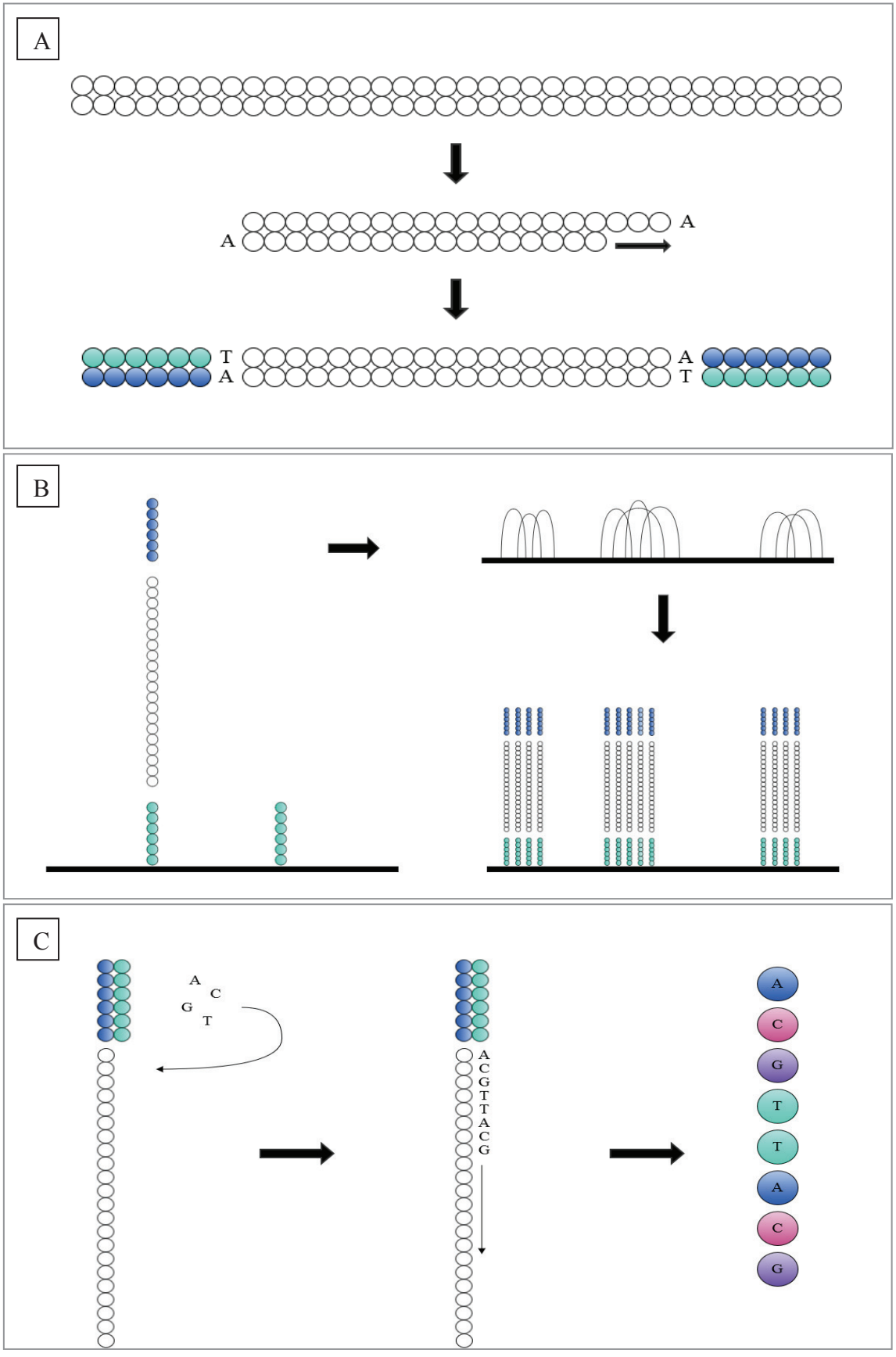


Figure 1.2 Illumina sequencing. A: Sample preparation. B: Colony formation. C: Sequencing (Estensmo, this thesis).

Roche 454

Another form of sequencing by synthesis technology is 454 pyrosequencing created by Roche. This technology is based on a PCR on beads in an emulsion of oil and water. This gives us one DNA fragment on each bead that is amplified, making each drop an independent PCR reaction (Buermans & den Dunnen, 2014). This mixture of oil, water and beads are then added to a picotiter plate, where each bead falls into one well. The sequencing starts with adding primers and nucleotides with pyrophosphate. During synthesis, the pyrophosphate cleaves off after binding of a nucleotide. This creates a light signal that is detected and that reveals the DNA sequence.

Ion Torrent Sequencing

This sequencing technology is developed by Life Technologies. The sample preparation is similar to that of 454, bead based PCR in an emulsion of water and oil. The DNA is denaturated and the beads are distributed into the wells of a fiber-optic slide (van Dijk et al., 2014). The beads contains only one amplified DNA fragment each, and only one bead is present in one well. The sequencing is based on proton release during nucleotide incorporation, which causes pH changes which are detected by ion sensors.

SOLiD

Sequencing by Oligonucleotide Ligation and Detection (SOLiD) is developed by Applied Biosystems. The sample preparation is similar to that of 454, except that the beads are distributed on a glass-slide rather that a picotiter plate (van Dijk et al., 2014). The sequencing happens by ligation: A sequencing primer hybridizes to an adapter which has a free end available for ligation to an oligonucleotide. A mixture of oligonucleotides competes for ligation to the primer, which are labeled with one of four colors. The oligonucleotide hybridizes to the adjacent sequence and the color is detected. The label is then removed and the cycle is repeated until all the bases are detected.

1.4.4 Third Generation Sequencing

Third generation technologies aim to increase the throughput and decrease both time and costs. Pacific Biosciences is the leader of this technology, which developed the Single Molecule Real Time (SMRT) technology (Eid et al., 2009). This technology is single molecule DNA sequencing by synthesis detected with zero-mode waveguide (ZMW) nanostructure arrays. The ZMW has a single illuminated φ29 polymerase attached to the bottom with a single molecule of DNA as template. The ZMW has a well-like structure that is

small enough to detect the single bases that is incorporated by the DNA polymerase. The bases are flowing freely in a solution added to the well and each of the four bases has a unique fluorescent label on the γ-phosphate. This phosphate is naturally cleaved off when the base is incorporated by the polymerase and hence the fluorescent label. The fluorescence is detected by a detector when the fluorochrome gets excited by the illuminated polymerase through incorporation (Eid et al., 2009). The detector makes the base call according to the fluorescence signal.

Several other third generation sequencing technologies are under development. One of these technologies are nanopore DNA sequencing (Buermans & den Dunnen, 2014). This technology reads the electrical signals that occur when nucleotides are passing by α-hemolysin pores. The nanopore changes its ion current when the DNA passes through. Each type of nucleotide changes the current independently, making it possible to recognize and detect the nucleotide sequence. Oxford Nanopore Technologies is one of the companies developing nanopore DNA sequencing.

1.5 Metagenome Analyses

Metagenomics is the study of organisms in a specific microbial community by analyzing the DNA directly within a sample (van Dijk et al., 2014). Next generation sequencing has revolutionized metagenomic analysis and improved our understanding of the function and diversity of microbial communities. Different sequencing methods exist for analyzing these communities, such as 16S rRNA sequencing and shotgun metagenome sequencing. These methods also require appropriate bioinformatic tools in order to analyze the huge amount of raw data generated.

1.5.1 Metagenome Sequencing

Sequencing of the 16S rRNA gene can be used for taxonomic classification of microorganisms (J. Gregory Caporaso et al., 2011; Olsen & Woese, 1993). This gene is highly conserved in all organisms due to its importance in the protein synthesis. However, the 16S rRNA gene consists of both conserved and variable regions (Olsen & Woese, 1993). This is because certain areas are more susceptible to mutations, creating conserved, variable and highly variable areas. The conserved regions can be used for primer design and the variable areas can be used for taxonomic classification. The variable regions are used to differentiate distant related organisms, whereas the highly variable areas are used to differentiate closely related species.

Shotgun metagenome sequencing investigates all the genes present in the sample. This method therefore provides information about both the organisms that is present and their metabolic processes (Segata et al., 2013). Shotgun metagenome sequencing happens by fragmentation of the DNA, sequencing of the short fragments and reconstructing them into a consensus sequence.

1.5.2 Bioinformatic Tools

QIIME

The 16S rRNA raw sequences can be analyzed through Quantitative Insights Into Microbial Ecology (QIIME) - an open-source bioinformatics pipeline. The analyses include demultiplexing and quality filtering, operational taxonomic unit (OTU) picking, taxonomic assignment, phylogenetic reconstruction and diversity analyses (J Gregory Caporaso et al., 2010). QIIME can be used to compare billions of sequences from thousands of samples from a microbial community, and is available at http://qiime.org/.

MG-RAST

Shotgun metagenome sequences can be analyzed in Metagenomics Rapid Annotation using Subsystem Technology (MG-RAST). MG-RAST was launched in 2007, and is a public available platform based on the SEED framework for comparative genomics (Meyer et al., 2008). Automated analysis provides quantitative insights into a microbial community based on metagenome sequence data. The server performs quality filtering, annotation and analysis providing both taxonomic and functional data for the uploaded samples.

ResFinder

ResFinder is a web-based database developed by Zankari and colleagues in 2012. The database uses Basic Local Alignment Search Tool (BLAST) for identification of acquired antimicrobial resistance genes in metagenome sequences (Zankari et al., 2012). Pre-assembled, complete genomes, partial genomes and short sequence reads can be used from four different sequencing platforms: 454, Illumina, Ion Torrent and SOLiD. The database is continuously updated as new resistance genes are detected, and can be accessed at www.genomicepidemiology.org.

1.6 Aim of Thesis

Antibiotic resistant bacteria possess a challenge to human health (Davies & Davies, 2010; Hu et al., 2013; Stalder, Barraud, Casellas, Dagot, & Ploy, 2012). The World Health Organization (WHO) stated in their report on global surveillance of antibiotic resistance that "*antibiotic resistance is no longer a prediction for the future; it is happening right now, across the world*" (WHO, 2014). This challenge involves bacteria resistance to single antibiotics, but also multidrug resistant bacteria. One of the mechanisms contributing to multidrug resistant bacteria are integrons, as they easily spread and acquire, exchange and express antibiotic resistant genes (Stalder et al., 2012). Moreover, factors contributing to bacterial stress, such as antibiotics, selects for integrons.

An environment that is highly exposed for antibiotics is the preterm gut. In developed countries, preterm birth is the leading cause of perinatal morbidity and mortality (Goldenberg et al., 2008). For the infants to survive, they need antibiotics to prevent infections. In addition, preterm infants are more susceptible to NEC. Even though several studies have associated several pathogens with NEC, little progress has been made in this field. In addition, the impact of antibiotic resistance and integrons on the pathogenesis and development of NEC is incompletely understood.

The aim of this thesis was therefore to investigate if integrons may be related to NEC and if the integrons can be transferred to other bacteria and hence is persistent in the preterm gut. In addition, a sub goal was to investigate if any bacteria could be related to NEC.

Approaches to achieve these goals included qPCR, Sanger sequencing, Illumina 16S rRNA metagenome sequencing and shotgun metagenome sequencing. For detection of integrons, qPCR, gel electrophoresis and Sanger sequencing was used. Further, 16S rRNA metagenome sequencing was used in order to investigate the microbial community. Finally, shotgun metagenome sequencing was used to provide taxonomic and functional information.

2. Materials and Methods

2.1 Cohort Description and Sample Information

Fecal samples were collected from patients at three different hospitals located in Boston, Chicago and Evanston. This study included preterm infants with and without NEC borne before gestational week 33. Sample characteristics are described in Table 2.1, and patient information is available in Appendix A.

Table 2.1 Sample characteristics.

	Patients	Fecal samples	Patients with NEC	Fecal samples from patients with NEC
Boston	26	107	9	39
Chicago	31	66	17	52
Evanston	9	9	3	3
Total	**66**	**182**	**29**	**94**

The samples were sent to Genetic Analysis (GA) AS Norway, which performed DNA extraction using different methods: GA method, magnetic beads and/or Qiagen. The DNA extracted samples along with 101 reference fecal samples were delivered to the Norwegian University of Life Sciences (NMBU) and stored at -40˚C. A flowchart of the experimental work is shown in Figure 2.1.

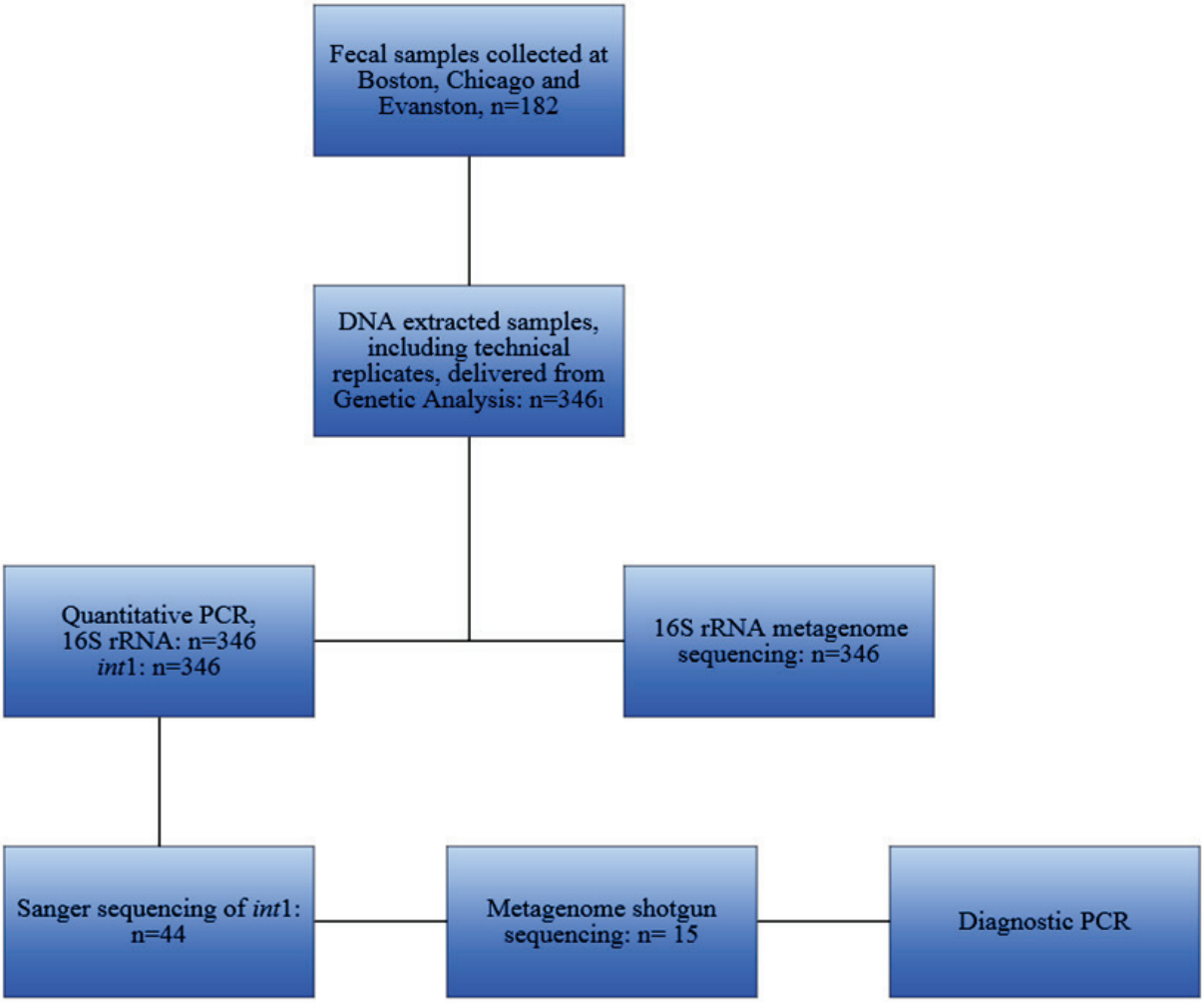


Figure 2.1 Flowchart of the experimental work.

1) In addition, 101 fecal samples were extracted at NMBU, but were excluded from further analyses due to lack of information.

2.2 DNA Extraction

DNA extraction at NMBU was performed using the Mag™ midi DNA extraction kit (LGC Genomics, UK) and performed automatically by the KingFischer™ Flex robot (Thermo Scientific, USA). Fecal samples were resuspended in Cary-Blair medium, and prepared for mechanical lysis by bead-beating: 50 µl sample and 150 µl S.T.A.R (stool transport and recovery) buffer (Roche, Germany) was added to tubes containing ~0.2g acid washed 106 µm glass beads (Sigma-Aldrich, Germany). The S.T.A.R. buffer inactivates infectious organisms, prevents degradation of nucleic acids, and enhances binding of nucleic acids to magnetic beads (Espy et al., 2006). The samples were processed twice in the MagNAlyser (Roche, Germany) at 6500 rpm for 20 sec - with rest between runs. The samples were kept cold for 2 min before centrifuging at 13000 rpm for 5 min to collect supernatants.

The supernatants were then treated with proteinase to remove cellular proteins: 50 µl lysis buffer and 5 µl proteinase were added to 50 µl supernatant, followed by incubation at 55 °C for 10 min. Finally, the samples were loaded on the robot and the DNA was automatically purified by eluting from paramagnetic beads by following the manufacturer`s instructions. The DNA was stored at -20˚C until further use.

2.3 Polymerase Chain Reaction

Different primers used to amplify the target regions are presented in Table 2.2.

Table 2.2 An overview of the different primers and their properties used for PCR.

Primer	Sequence (5′-3′; 5′-3′)	Tm (˚C)	Target	Reference
***Int*1 F;** ***Int*1 R**	ACGAGCGCAAGGTTTCGGT; GAAAGGTCTGGTCATACATG	66; 53	The class 1 integron integrase gene	(Sørum, L'Abée-Lund, Solberg, & Wold, 2003)
16S F; **16S R**	TCCTACGGGAGGCAGCAGT; GGACTACCAGGGTATCTAATCCTGTT	59; 58	Conserved regions of the 16S rRNA gene	(Nadkarni, Martin, Jacques, & Hunter, 2002)
PRK341 F; **PRK806 R**	CCTACGGGRBGCASCAG; GGACTACYVGGGTATCTAAT	61; 60	The V3-V4 region of the 16S rRNA gene	(Y. Yu, Lee, Kim, & Hwang, 2005)
Plasmid F; **Integron R**	GCTCGGATCTCAGGACGAAG; TGCCTAGCATTCACCTTCCG	63; 62	The plasmid- integron	This thesis
Integron F; **Transposon R**	GGCCATTCCGACGTCTCTAC; GAAATGCGCCTGGTAAGCAG	61; 62	The integron-transposon	This thesis

2.3.1 Qualitative PCR

The reaction mix contained 1.25U HOT FIREPol® DNA polymerase (Solis BioDyne, Estonia), 1x HOT FIREPol ® buffer B2 (Solis BioDyne, Estonia), 2.5 mM $MgCl_2$ (Solis BioDyne, Estonia), 0.2 mM dNTP (Solis BioDyne, Estonia), 0.2 µM forward/reverse primer and nuclease-free water (Amresco, USA). The reaction volume was 25 µl per reaction, including 1 µl template. Thermal cycling was performed on a 2720 Thermal Cycler (Applied Biosystems, USA). The standard program started with initial denaturation at 95˚C for 15 min, followed by the desired number of cycles with 95˚C for 30 sec and annealing and extension adjusted according to the template. The final step was 72˚C for 7 minutes. This recipe has been used for *int*1-, PRK- and Illumina primers.

Gradient PCR was performed to detect the optimal annealing temperature of the primers targeting the plasmid-integron and the integron-transposon. The PCR was performed on Mastercycler® thermal cycler (Eppendorf, Germany) with 30 sec of annealing and extension at 72˚C for 45 sec. The gradient was set to ± 10 degrees, creating a temperature range from 40-60˚C. The PCR products were confirmed by Sanger sequencing.

2.3.2 Quantitative PCR

The reaction mix contained 1x HOT FIREPol® EvaGreen® qPCR mix (Solis BioDyne, Estonia), 0.2 µM forward/reverse primer and nuclease-free water. The reaction volume was 20 µl per reaction, including 1 µl template. Thermal cycling started with initial denaturation at 95˚C for 15 min, followed by 40 cycles of denaturation, annealing and eventually extension.

The samples were first amplified with primers targeting universally conserved regions of the 16S rRNA gene, and included denaturation at 95˚C for 30 sec and annealing at 60˚C for 1 min. Then the samples were amplified with primers targeting the integrase gene of the class 1 integron (*int*1), which included denaturation at 97˚C for 30 sec, annealing at 53˚C for 30 sec and extension at 72˚C for 30 sec. Thermal cycling for the plasmid- integron and the integron-transposon included denaturation at 95˚C for 30 sec, annealing at 60˚C for 30 sec and extension at 72˚C for 45 sec.

Fluorescence was measured by LightCycler® 480 (Roche, Germany), and included high resolution melting (HRM) curve analysis. The raw fluorescence data was exported into the LinRegPCR program (Ruijter et al., 2009), which performed baseline corrections and calculated mean PCR efficiency. The data was further exported to and processed in Microsoft Excel (Microsoft, USA). For the plasmid-integron and the integron-transposon primers, standard curves were included and used to calculate PCR efficiency.

2.3.3 Sequencing PCR

To investigate the microbial community, 16S rRNA metagenome sequencing was done. It was performed by following recommendations from Illumina.

A two-step PCR was done with FIREPol® DNA polymerase. The first PCR was done with PRK primers targeting the V3-V4 region of the 16S rRNA gene. Thermal cycling included 25 cycles with standard program, annealing at 50˚C for 30 sec and extension at 72 ˚C for 45 sec. The PCR products were cleaned with AMPure XP (Beckman Coulter, USA) - paramagnetic beads that bind DNA - in a ratio of 1:1. AMPure purification is needed to remove excess

primers, nucleotides and to select the desired fragment size. The cleaned PCR products were used in a second PCR.

The second PCR was performed with the same primers modified by addition of unique Illumina adapters (Figure 2.2). The library included 16 forward and 36 reverse primers (Appendix B), creating 576 possible combinations - only one present once and for one sample. The same thermal cycling conditions were used as above; expect changes initial denaturation at 95˚C for 5 min, followed by 10 cycles. The annealing time was increased to 1 min to ensure binding of extended primers.

F 5`-aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctagtcaaCCTACGGGRBGCASCAG -3`

R 5`-caagcagaagacggcatacgagatcgtgatgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT-3`

Figure 2.2 Illumina forward (F) and Illumina reverse (R) primer pair 1. The gene-specific region (capital) corresponds to the primer sequence of the gene to be amplified (in this case PRK). The colony-amplification region (blue) binds to a complementary sequence on the flowcell. Unique Illumina adapters (purple) make the separation and detection of colonies possible.

2.3.4 Primer Design

Primers were designed to investigate if the location of the integron was in a plasmid or a transposon. Primers were designed using the Geneious software version 8.0.5 (Biomatters, New Zealand) in cooperation with primer3 (Untergasser et al., 2012). A primer pair was designed to cover the plasmid and the integron (~300 bp) and another to cover the integron and the transposon (~300 bp).

2.4 DNA Qualification and Quantification

2.4.1 Gel Electrophoresis Qualification

The products were qualified on a 1 % agarose gel in 1x TAE buffer. The pores of the gel provides resistance that separates the smaller fragments from larger fragments during migration caused by electric voltage: Due to the DNA`s negative charge, the fragments move to the positive pole in an electric field. A 100 bp ladder (Solis BioDyne, Estonia) was used to determine the fragment sizes. GelRed™ (Biotium, USA) or peqGreen™ (Peqlab, Germany) - DNA binding dyes that fluoresces under exposure of UV light - was added to visualize the fragment bands using the Gel Doc™ XR+ System (Bio-Rad, USA).

2.4.2 Qubit Quantification

DNA was quantified by the Qubit® dsDNA (double-stranded DNA) HS (High Sensitivity) Assay Kit (Life Technologies, USA). The assay is highly selective for dsDNA and is designed to calculate DNA concentrations. The assay was performed by following manufacturer's recommendations using 2 µl of sample DNA and 198 µl working solution. The concentration was read using the Qubit® Fluorometer v 1.0.

2.5 Sequencing

2.5.1 Sanger Sequencing

Sanger sequencing was performed by using the BigDye Terminator v.1.1 Cycle Sequencing kit (Applied Biosystems). The PCR products were exposed to exonuclease 1 (Exo1) treatment to remove excess primers and small fragments. BigDye labeling included the *int*1 forward primer. BigDye PCR products were purified using Agencourt® CleanSEQ® (Beckman Coulter, USA) to remove excess BigDye labels. All the steps were performed by following the manufacturer's recommendations. The sequencing was done at the University College of Hedmark, Hamar, Norway.

2.5.2 16S rRNA Metagenome Sequencing

The PCR products were quantified and normalized based on gel electrophoresis. The samples were pooled according to the band strength on the agarose gel (1:1 because of the similarity). PCR cleanup was performed with AMPure XP with 0.6x bead concentration to 150 µl pooled sample and eluted in 30 µl 10 mM Tris pH 8.5 from Illumina. The sample concentration was measured with PerfeCta® NGS Quantification Kit for Illumina (Quanta BioSciences, USA) and diluted in Tris pH 8.5 to a 4 nM concentration. The sample was denaturated and loaded on the MiSeq (Illumina, USA) in a 6pM concentration spiked with 15% PhiX control (Illumina, USA). PhiX is necessary in low diversity samples for increasing the library nucleotide balance and making clusters easier to identify by the software.

2.5.3 Shotgun Metagenome Sequencing

For full metagenome investigation, shotgun metagenome sequencing was performed using the Nextera® XT DNA sample preparation kit (Illumina, USA). This kit uses a Nextera® XT transposome - an enzyme that simultaneously performs fragmentation and tagmentation of the DNA. This result in addition of unique adapter sequences, needed in a limited cycle PCR to amplify the insert DNA. Index sequences were added to the DNA to enable cluster formation during sequencing.

The samples were prepared by mainly following the *Nextera® XT DNA Sample Preparation Guide* from Illumina. Samples with high *int*1 gene quantification were selected for full metagenome sequencing. After tagmentation and PCR, PCR cleanup was performed as recommended with 0.6 x AMPure® XP bead concentration. Library normalization was done based on qPCR using the colony amplification primers from Illumina - not by bead-based normalization as recommended. The samples were pooled according to the Ct values. The sample concentration was measured with PerfeCta® NGS Quantification Kit for Illumina, and the pool was denaturated and loaded on the MiSeq in a 6pM concentration spiked with 5% PhiX control.

2.6 Data Analyses

Student`s t-Test (two tailed distribution assuming unequal variance) was used to calculate the p-values, if otherwise is not mentioned in the text.

2.6.2 16S rRNA Analyses

The 16S rRNA sequences were analyzed using the QIIME pipeline. The sequences were quality filtered and then clustered at 99% homology level using closed-reference uclust search against Greengenes database to create an OTU table. This table was used for the PLS analysis in MATLAB. The analysis included a receiver operating characteristic (ROC) curve to examine the sensitivity and specificity of the dataset, and generation of a VIP score to detect influential OTUs. The diversity in the samples was analyzed by rarefaction plots for the α-diversity; while Unifrac principal coordinates analyses (PCoA) was performed to assess the β-diversity.

2.6.1 Metagenome Analyses

The metagenome sequences from selected samples were aligned and assembled to reference genomes using Geneious pipeline and CLC main workbench. The sequences were also uploaded to and analyzed in MG-RAST. The default settings included maximum e-value cutoff of 1e-5 and minimum 60% identity. Taxonomic assignment was done by the M5 non redundant (M5NR) database, whereas functional assignment was done by the Subsystems database. For detection of resistance genes, the sequences were uploaded to and processed in the ResFinder 2.1 database. The default settings included a 98% threshold for identity and 60% minimum length.

3. Results

3.1 16S rRNA Metagenome Analyses

The microbial composition in the fecal samples was investigated using 16S rRNA metagenome sequencing by analyzing the sequences in QIIME. After quality filtering, a total of 14 621 559 sequences were detected in all the samples. The sequences were clustered with 99 % homology level using closed-reference uclust search against Greengenes database to construct an OTU table. This table was further processed; 6000 sequences/sample was randomly picked to ensure even sequence information, filtering away 68 samples. The final OTU table contained 192 samples with 528 OTUs belonging to 13 classes.

3.1.1 α-diversity Analyses

To investigate the species diversity in the samples, α-diversity analysis was performed. The α-diversity calculations from QIIME were used to make rarefaction plots with the average number of observed species to the amount of sequences per sample. The calculations were done for both NEC negative and NEC positive infants to compare the species diversity in the samples between the two groups (Figure 3.1A). No difference in diversity was detected between the two groups. The calculations were also done for the different hospitals; Boston, Chicago and Evanston (Figure 3.1B). Different diversities were detected between the hospitals, with the highest diversity difference in the samples from Evanston.

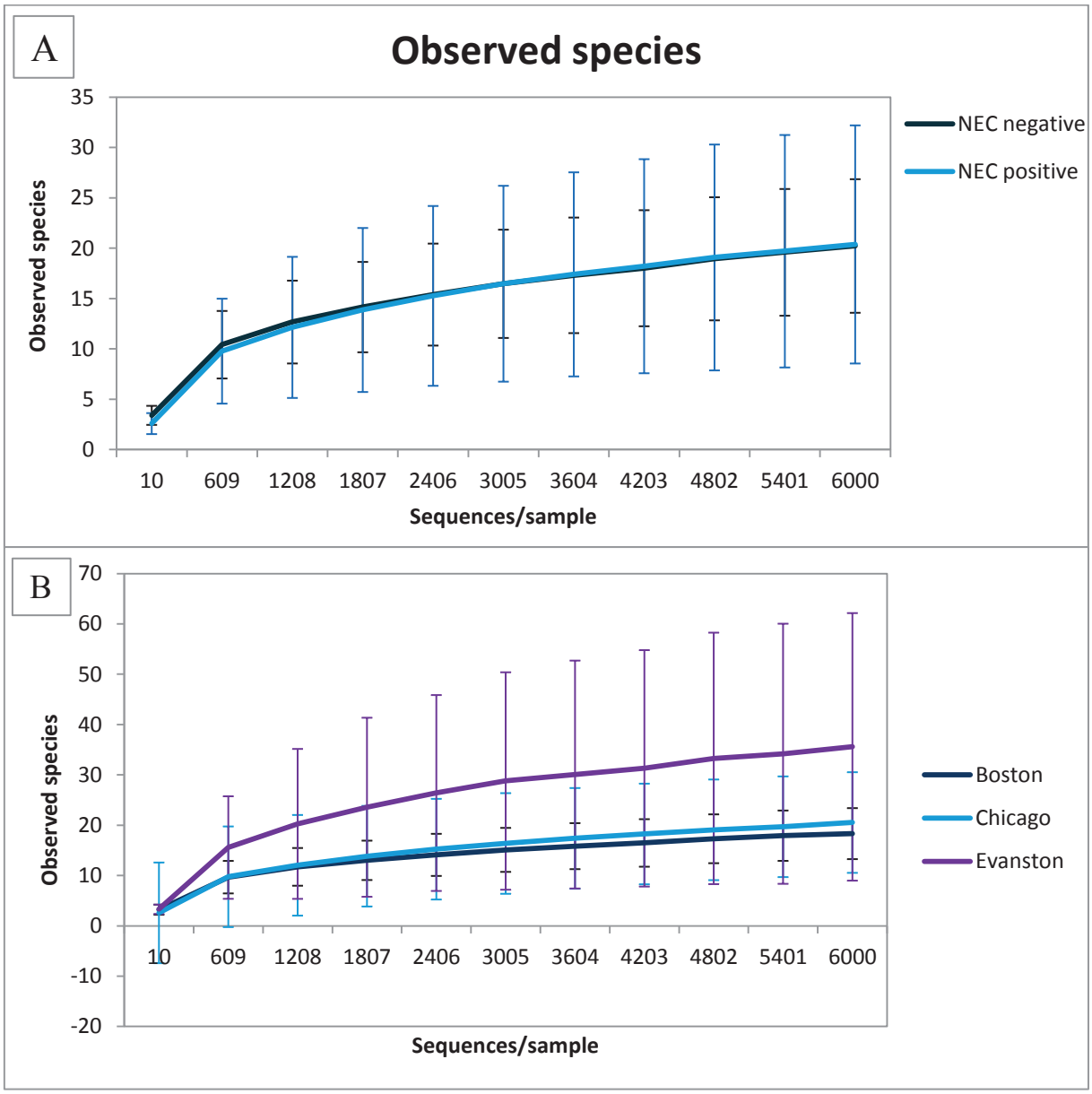


Figure 3.1 Rarefaction curves of observed species in the number of sequences per sample (average ± SEM). A: observed species between NEC negative and NEC positive patients. B: Observed species between the different hospitals.

3.1.2 β-diversity Analyses

The variation between the samples was analyzed by UniFrac principal coordinate's analysis (PCoA) plot. Unweighted UniFrac PCoA between the infants with and without NEC showed no difference (Appendix C). On the other hand, weighted UniFrac PCoA of this data showed a grouping of NEC positive samples (Figure 3.2A). Weighted UniFrac PCoA of samples belonging to the different hospitals showed no difference between the institutions (Figure 3.2B).

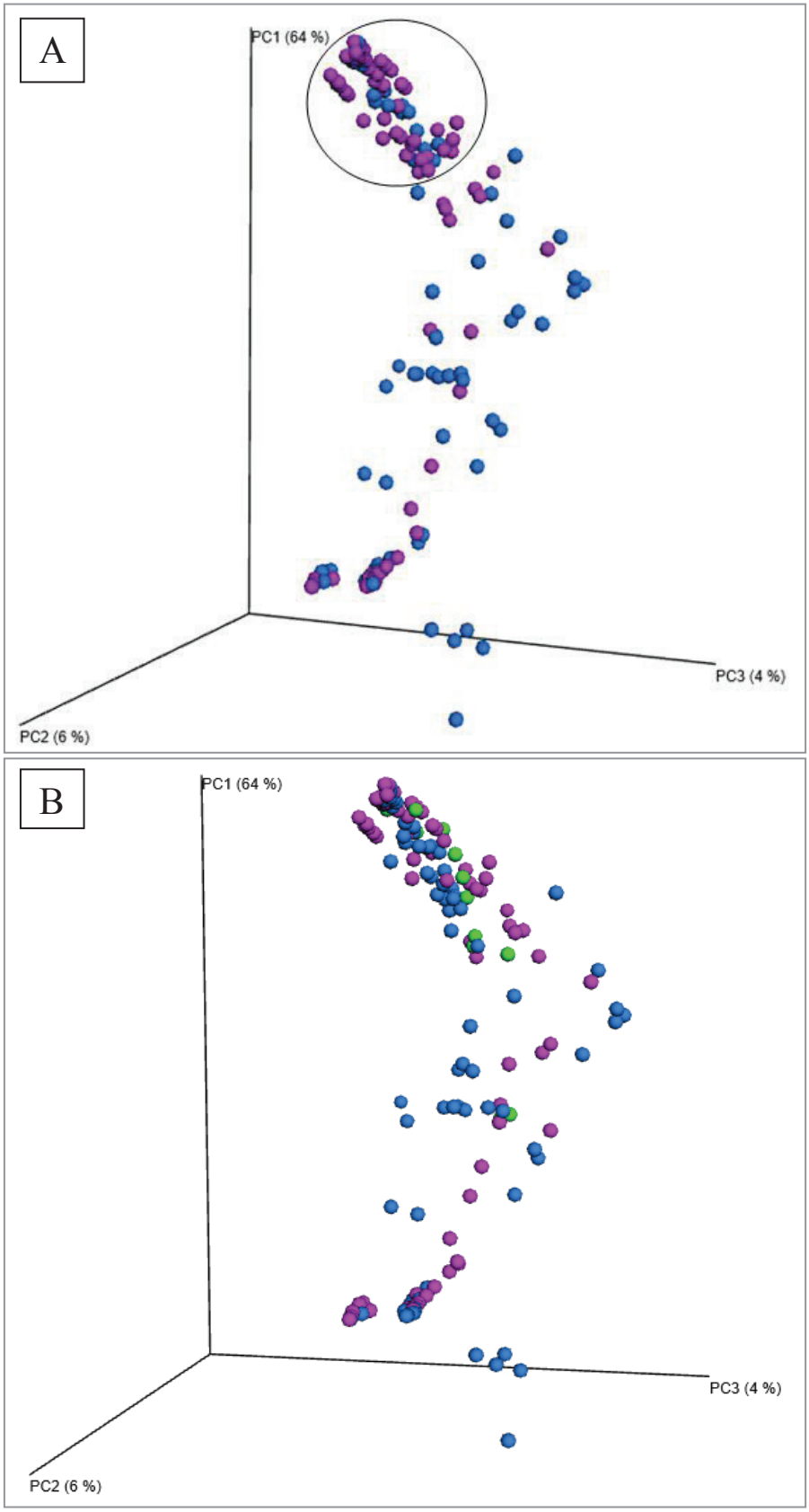


Figure 3.2 Weighted UniFrac PCoA plot. A: samples from patients with NEC (purple) and from patients without NEC (blue). The circle reveals a clustering of samples from NEC positive patients. B: samples from patients from Boston (blue), Chicago (Purple) and Evanston (green).

3.1.3 Taxonomic Analyses

The taxonomic composition at phylum, family and genus level was investigated in the samples. The taxonomic composition at the different levels was compared between preterm infants with and without NEC.

The analyses revealed that the phyla *Actinobacteria*, *Firmicutes* and *Proteobacteria* dominated in these infants (Appendix D). The abundance of *Actinobacteria* and *Firmicutes* were higher in infants without NEC, whereas the phylum *Proteobacteria* was more abundant in infants with NEC.

The taxonomic analysis on the family level revealed that *Enterobacteriaceae*, *Enterococcaceae* and *Staphylococcaceae* dominated in these infants (Figure 3.3A). The infants with NEC had a higher abundance of *Enterobacteriaceae* and *Enterococcaceae*, whereas the family *Staphylococcaceae* was more abundant in the healthy infants.

On the genus level, the dominating bacteria belonged to *Serratia*, an unclassified *Enterobacteriaceae*, *Enterococcus* and *Staphylococcus* (Figure 3.3B). *Serratia* and *Enterococcu*s were more abundant in infants with NEC, whereas healthy infants had a higher abundance of *Staphylococcus*. The abundance of the unclassified *Enterobacteriaceae* seemed to equal in the two groups.

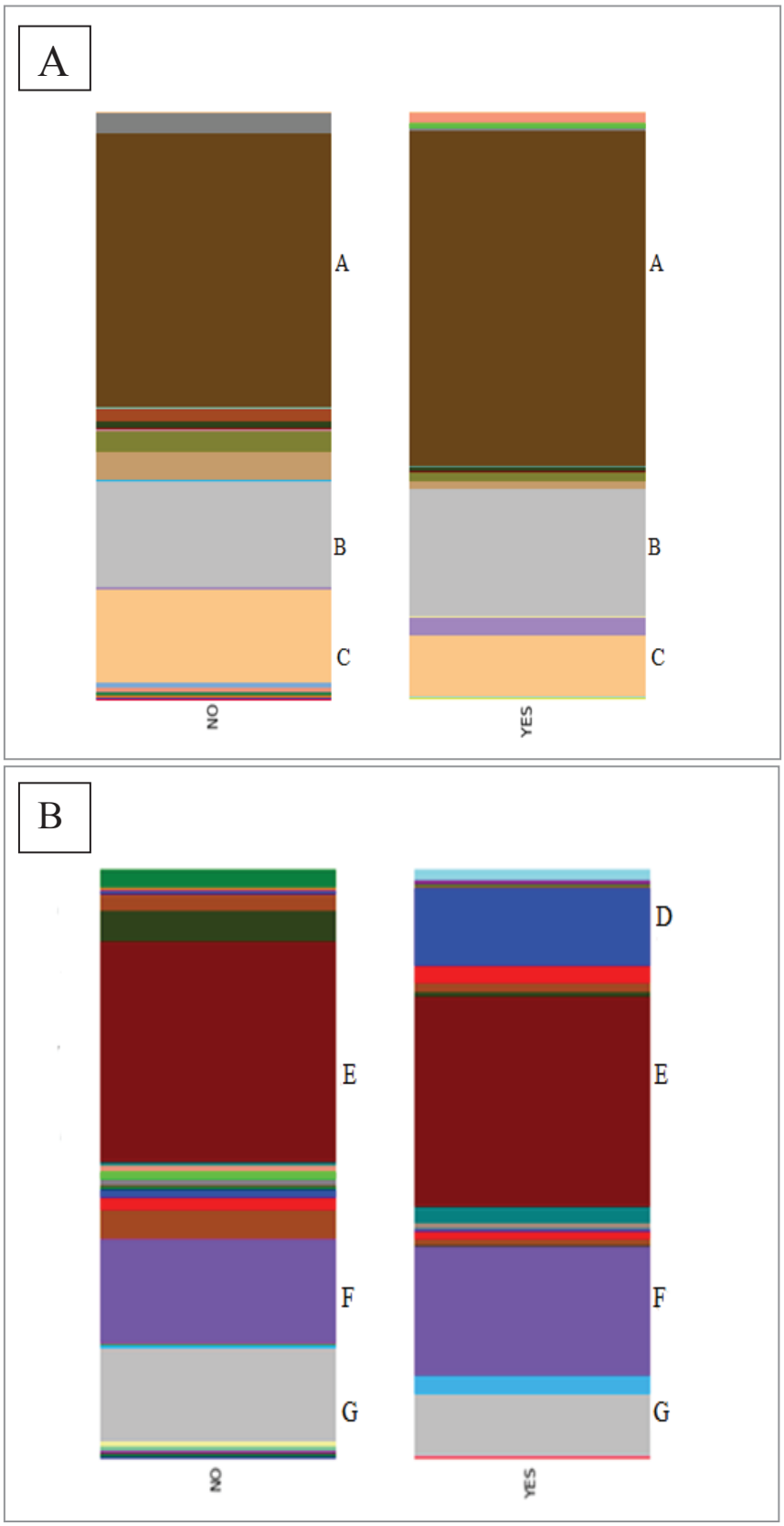


Figure 3.3 Taxonomic analysis of the samples from infants with (YES) and without (NO) NEC. Barchart A shows the taxonomic composition at family level. A= *Enterobacteriaceae*, B= *Enterococcaceae*, C= *Staphylococcaceae*. Barchart B shows the taxonomic composition at genus level. D= *Serratia*, E= *Enterobacteriaceae* (unclassified), F= *Enterococcus*, G=*Staphylococcus*.

3.1.4 Assigning OTUs According to NEC

Further, analyses were done to investigate if any OTUs could be linked to NEC. Due to statistical analyses, the average value of each OTU was calculated for all the samples that belonged to each patient. This table was used for PLS analysis (Conducted by Knut Rudi).

PLS analysis revealed six OTUs influential according to NEC; OTU2 (Unclassified *Enterobacteriaceae*), OTU4 (Unclassified *Enterobacteriaceae*), OTU9 (Unclassified *Enterobacteriaceae*), OTU13 (*Haemophilus parainfluenzae*), OTU22 (*Bifidobacterium*) and OTU25 (*Paenibacillus*).

Comparing the amount of these OTUs in patients with and without NEC (Figure 3.4) revealed that the amount of OTU2 was significantly higher in patients with NEC (p=0.03), whereas the amount of OTU13 was significantly higher in patients without NEC (p=0.04). The amount of OTU4 was slightly higher for patients with NEC, whereas the amount of OTU9, OTU22 and OTU25 was below 1% for in both groups.

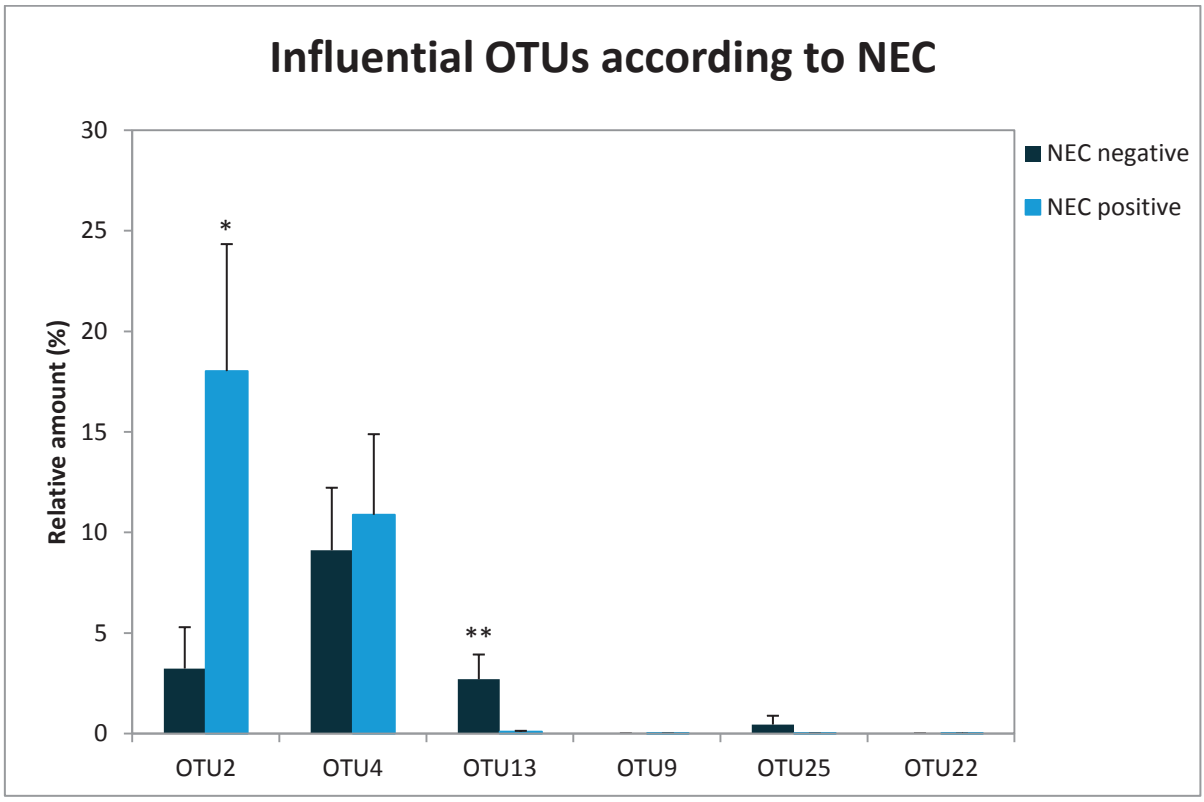


Figure 3.4 Influential OTUs according to NEC. Relative amounts of the six OTUs in NEC negative and NEC positive patients (average +SEM). *p=0.03, **p=0.04.

The sequences belonging to the six most influential OTUs were extracted. These were used to make a BLAST search at NCBI for better classification (Table 3.1). The classification turned out to be similar to that of QIIME for OTU 13, 22 and 25. The classification of OTU 2, 4 and 9 were at family level from QIIME, but got classification at genus and specie level from BLAST.

Table 3.1 The classification of the six OTU sequences by BLAST at NCBI. All the sequences were classified with 99% identity, 98% query cover and an E-value of 0.0.

OTU	Bacteria	Accession number
2	*Escherichia coli*	KP036624.1
4	*Klebsiella pneumonia*	KJ560981.1
9	*Pantoea sp.*	CP009866.1
13	*Haemophilus sp.*	KM873115.1
22	*Bifidobacterium longum*	LN824140.1
25	*Paenibacillus polymyxa*	CP000154.2

3.2 Detection of Integrons

The presence and abundance of integrons was analyzed using qPCR of the *int*1 gene. The *int*1 gene has a characteristic melting curve with complete denaturation at 94^{o}C and a fragment size ~530 bp, making the gene easily detectable. Based on these results, 44 possible *int*1 positive samples were detected showing an *int*1 melting curve, a ~530 bp fragment size or both. The HRM analysis and the agarose gel for these 44 samples are shown in Appendix E and Appendix F, respectively. Sanger sequencing of these samples gave BLAST hits to *int*1 sequences, which further confirm the presence of the *int*1 gene in these samples.

The *int*1 copy numbers were calculated relative to the 16S rRNA quantification. Average values were made for each of the 13 patients and the amount of *int*1 was compared between the NEC positive and NEC negative patient. The amount of *int*1 was close to be significantly higher in NEC positive patients (p=0.05).

Persistence of integrons was detected in some of the patients. This concerned four of the patients, in which three of these were NEC positive. Binomial testing was used to address the probability of randomly selecting 3/4 with NEC (p=0.19).

3.3 Metagenome Analyses

Based on the persistence of integrons and high *int*1 copy number, 15 samples were picked out for shotgun metagenome sequencing. These concerned four samples from both patient 9 and 13, three samples from both patient 58 and 62 and one sample from patient 49. More information about these patients can be found in Appendix A.

3.3.1. MG-RAST Analyses

The metagenome sequences were uploaded to and analyzed in MG-RAST. After quality control, 10 376 397 sequences were detected for all the samples, with an average number of sequences per sample of 691 759.

Taxonomic Analysis

Taxonomic information was generated by comparing the sequences to the M5NR database. The bacteria (89.24 ± 1.16%) [average ± SEM] dominated in the patients, followed by small fractions of viruses (0.28 ± 0.003%), eukaryota (0.12 ± 0.002%) and archaea (0.013 ± 0.0002%). Approximately 10% of the sequences were classified as unassigned.

The dominant bacteria phyla, families and genera were compared between the patients. *Proteobacteria* was the dominating phylum followed by *Firmicutes* (Figure 3.5A). At the family level, *Enterobacteriaceae* was the dominating group (Figure 3.5B). *Escherichia* and *Klebsiella* dominated at the genus level (Figure 3.5C).

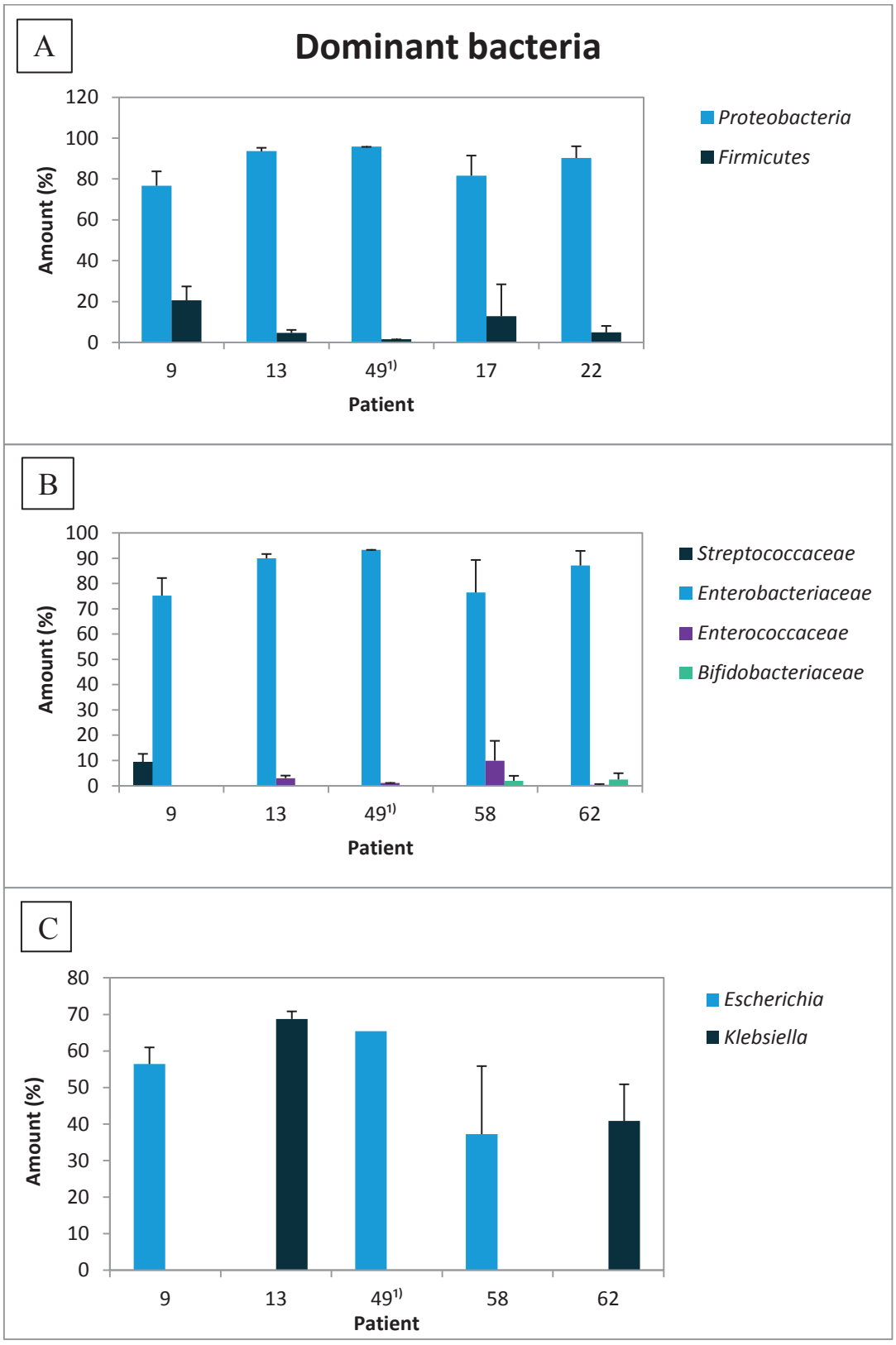


Figure 3.5 The dominant bacteria in the patients. The numbers are based on average values including SEM. A: dominant bacteria on the phylum level. B: dominant bacteria on the family level. C: dominant bacteria on the genus level.

1) Only one sample.

Functional Analysis

The functional analyses were performed by the Subsystems database, which detected a total of 28 subsystems. The subsystems that compromised of less than 1% were grouped in the subsystem other. The standard deviation of the subsystems varied slightly in the samples, but did not exceed 3.4%. The dominating subsystems belonged to Carbohydrates and Clustering-based subsystems (Appendix G). The subsystem Virulence, Disease and Defense had coverage of 3%, in which the subsystem Resistance to Antibiotics and Toxic Compounds accounted for ~70%. This subsystem was further dominated by the group Multidrug Resistance Efflux pumps, which accounted for 19% (Figure 3.6).

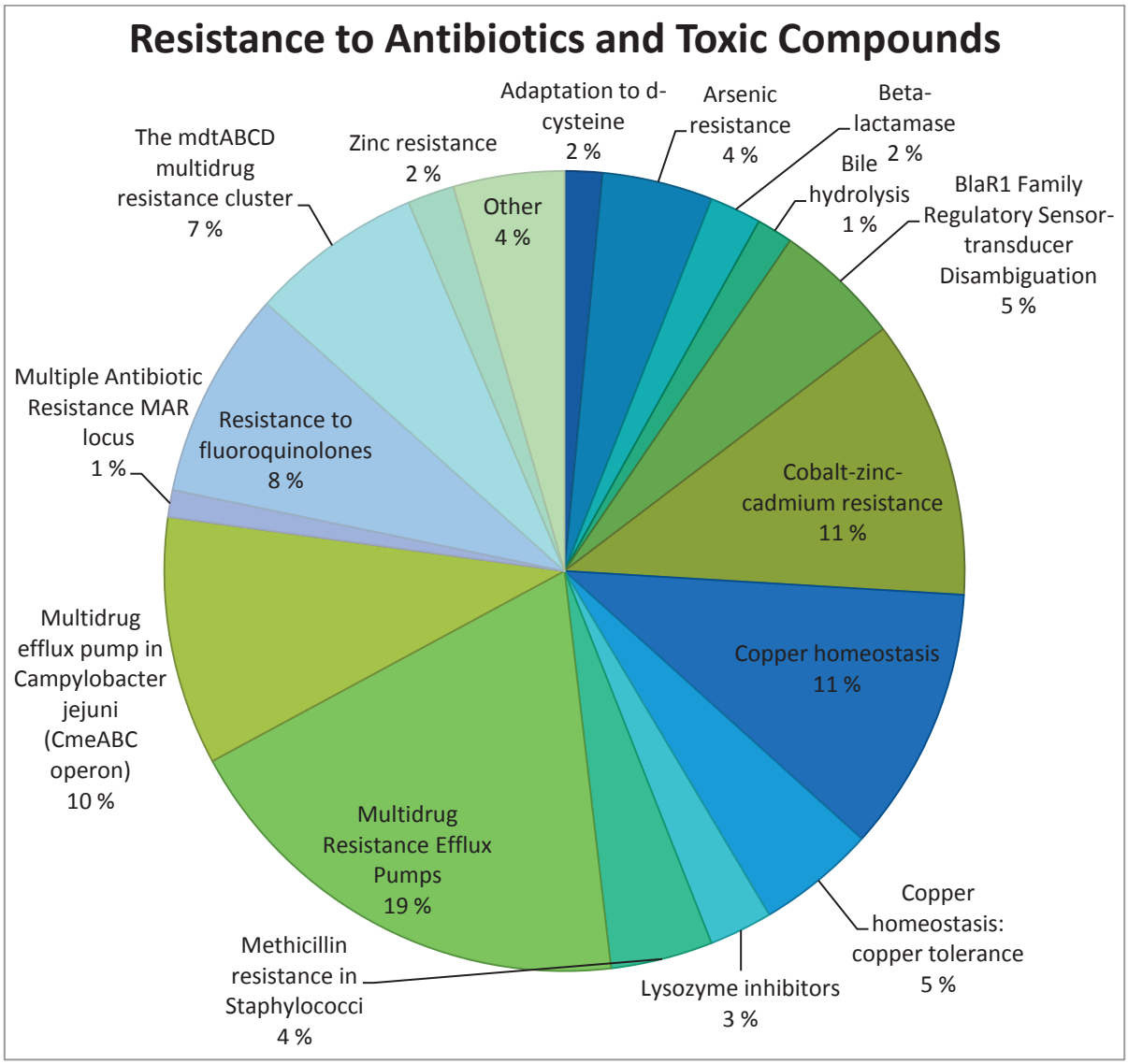


Figure 3.6 The subsystem Resistance to Antibiotics and Toxic Compounds. The groups are relatively distributed in the subsystem and based on average values for all the samples.

Abundance of Integrons

All the samples showed presence of integrons with varying abundance. The highest abundance of integrons was detected by the PATRIC server and this was normalized according to the total number of sequences. These relative values were made for each patient (Figure 3.7). The highest relative abundance of integrons was found in patient 9, 49 and 62.

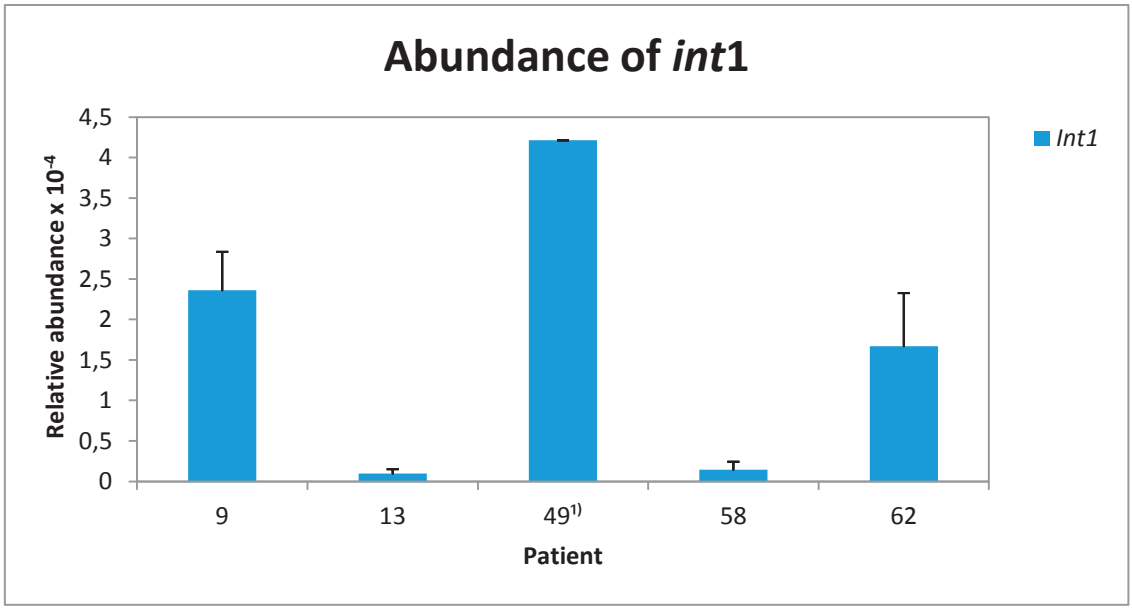


Figure 3.7 Abundance of integrons in the different patients (average + SEM).

1) Only one sample.

3.3.2. Detection of Resistance Genes

The raw metagenome sequences were uploaded to the ResFinder database for detection of resistance genes. Resistance to ten different antibiotics was detected and several resistance genes were detected in all the samples (Appendix H). The results were scored according to the abundance of the resistance genes in samples from each patient (Table 3.2). The abundance and type of resistance genes were different for each patient, but beta-lactam resistance was found in every patient.

Table 3.2. Different resistance genes detected in the different patients.

P[1]	I[2]	NEC	N[3]	Beta-lactam	Macrolide	Sulphonamide	Aminoglycoside	Trimethoprim	Fosfomycin	Tetracycline	Phenicol	Macrolide, Lincosamide and Streptogramine B	Quinolone
9	B	No	4	++[4]	++	++	++	++	-	-	-	-	-
13	B	Yes	4	+[5]	+	-	-	-	++	-	-	-	-
49	E	Yes	1	++	-[6]	++	++	++	-	++	++	-	-
58	C	Yes	3	+	+	-	+	-	-	+	-	+	-
62	C	Yes	3	+	+	++	++	+	++	-	-	-	++

1) P = Patient

2) I = Institution; B= Boston E= Evanston, C= Chicago.

3) N = Number of samples

4) ++ = Resistance gene found in all samples

5) + = Resistance gene found in some samples

6) - = Resistance gene not detected

3.3.3 Assembly of Integron Containing Contigs

The metagenome sequences were made into contigs and assembled to reference genomes (Conducted by Anuradha Ravi). The contigs were assembled to the *Escherichia coli* plasmid p1658/97 (GenBank accession number: AF550679), which contained the *int*1 gene in addition to a transposon (Appendix I). The contigs from all the samples mapped to the plasmid, the *int*1 gene and the transposon.

To investigate if the integron was linked to the plasmid or the transposon, primers were designed for a diagnostic PCR. Primers covering the plasmid and the integron were designed from ORF35 of the plasmid to the *sul*1 gene of the integron (~300bp). Primers covering the transposon and the integron were designed from the *int*1 gene of the integron to the truncated transposon AIS26 (~300bp).

Gel electrophoresis of the PCR product showed that integron is more likely connected to the plasmid rather that the transposon (Figure 3.8). That is due to expected fragment size in addition to presence in every sample. The PCR product that connected the integron to the transposon showed background amplifications with varying fragment sizes.

In addition, Sanger sequencing of the integron-plasmid amplicons matched with the metagenome sequences, whereas the sequence covering the transposon and the integron mapped to different areas around the target (Appendix J).

Figure 3.8 Gel electrophoresis of the qPCR products. The upper gel shows the integron-plasmid samples, whereas the bottom gel shows the integron-transposon samples. Positive (+) and negative (-) controls were included. The negative controls were added in the picture as they were originally included in the standard curve. The contrast is increased for better visualization of the fragments.

3.4 Technical Validation

3.4.1 DNA Extraction

After DNA extraction of the fecal samples, Qubit measurements of a few random selected samples were done to ensure presence of DNA. Even though the measurements varied for different samples (0.125-11.2 µg/ml), all the samples contained DNA.

3.4.2 Quantitative PCR

The qPCR was used both for normalization, quantification of genes and to verify that the DNA could be amplified. A qPCR with the colony amplification primers was used to normalize the metagenome samples. The Ct values varied from 16.474 - 22.643 and the samples was normalized according to the lowest Ct value.

A qPCR of the 16S rRNA gene was done both for verification and for quantification of the gene. The Ct values varied between the samples (10.991 - 0.000). Samples with Ct values above the negative PCR controls (< 32) and zero were diluted 1:10 and included in a second PCR. All together, only two samples did not amplify. The Ct values were used to calculate the relative quantity of the *int*1.

In the qPCR of the *int*1 gene, a positive control (*E. coli*) was added in addition to the negative PCR control. The positive control was used to compare Ct values, HRM curves and fragments visualized by gel electrophoresis. The Ct values varied greatly between the samples (16.980-0.000) and many samples did not amplify. All the samples with Ct values below the negative PCR control, including the positive control, were used HRM analysis and gel electrophoresis in order to detect as many *int*1 positive samples as possible.

The qPCR with the primers covering the plasmid-integron and the integron-transposon included a positive (*E. coli*) and a negative PCR control. Standard curves for the primer pairs were added to compare the PCR efficiency and gel electrophoresis were included to compare the fragments.

HRM analysis was included for all the qPCRs in order to detect unspecific amplification by comparing the melting curves.

3.4.3 Sequencing

Sanger Sequencing

After the *int*1 PCR of the positive *int*1 samples, gel electrophoresis was performed to ensure amplification. A positive (*E. coli*) and negative control were included in the PCR and the following reactions including sequencing.

16S rRNA Sequencing

The PRK primers gave the expected fragment size (~500bp) by gel electrophoresis. Samples that did not amplify were diluted 1:10 and included in a second PCR. Even though, 76 samples did not amplify. The PCR products were purified by AMPure and some samples were randomly selected for Qubit measurements before and after purification to check that the purification was performed. All the samples had lower DNA concentration after purification. The normalized sample was AMPure purified and included in a gel electrophoresis along with the supernatant to control the purification. The concentration of the normalized sample was quantified by Quanta qPCR and Qubit to be 74.2 nM and 93.4 nM, respectively.

After sequencing, the OTUs from duplicate samples isolated with different DNA extraction method were included in regression analysis (Appendix K). Variations were detected between some of the samples. The R^2 values for the samples extracted with GA method versus magnetic beads varied from 0.015 to 1, with an average value of 0.73 ± 0.09 (SEM). The R^2 values for the samples extracted with GA method versus Qiagen varied from 0.0001 to 1, with an average value of 0.69 ± 0.07 (SEM).

4. Discussion

4.1 Presence and Persistence of Integrons

The amount of *int*1, relative to the 16S rRNA, was close to significantly higher in the patients with NEC. Therefore, it is possible that high amounts of *int*1 in a patient are associated with NEC. A reason for this explanation might be that infants with NEC are more exposed to antibiotics, since the use of antibiotics can increase the abundance of class 1 integrons (Gillings et al., 2014). As we know, this study is the first to associate *int*1with NEC.

Persistence of integrons was also found in several of these patients, as several samples taken at different time periods from the same patient contained integrons. All these samples showed high copy number of *int*1, and all the patients except one were NEC positive. These samples, along with another sample also showing high *int*1 copy number, were chosen for metagenome sequencing. Integrons were detected in all the metagenome sequences, but the abundance varied between the patients.

Assembly of integron containing contigs detected that the integrons could be associated with mobile genetic elements such as a plasmid and a transposon. This hypothesis was strengthened by the diagnostic PCR, which showed that the integrons was associated with the plasmid. The association of the integrons to mobile genetic elements is important, as these easily transfers to other bacteria. This may explain the high prevalence and persistence in some of the patients. Studies have proposed that mobile genetic elements can spread in the neonatal intensive care unit and potentially contribute to NEC (Raveh-Sadka et al., 2015; Stewart & Cummings, 2015). Therefore, it is possible that this also concerns integrons. In addition, other studies have found that integrons are widespread in commensally *E. coli* from healthy humans (Li et al., 2014; Vinue et al., 2008). It is therefore possible that the integron can be transferred by these bacteria to the infants through human contact.

4.2 Diversity of Antibiotic Resistance Genes

Several resistance genes were detected in all the metagenome samples by the ResFinder database. The abundance of resistance genes detected was different for each patient. Beta-lactam resistance was found in every patient, whereas macrolide and aminoglycoside resistance were detected in four of five patients. Sulphonamide and trimethoprim resistance were found in three of five patients, while fosfomycin and tetracycline were detected in two

of five patients. Phenicol, quinolone and macrolide, lincosamide and streptogramine B resistance were only detected in one of five patients.

The prevalence of resistance genes varied between the patients. Some of the patients showed resistance towards three antibiotics, while others showed resistance towards as many as seven antibiotics. Many factors could be a part of this explanation. The health of the infants plays an important role, as different patients may have needed different treatments and have been exposed to different antibiotics causing resistance. The maternal microbiota also plays an important role, as antibiotic resistance genes can be transferred from mother to child (Zhang et al., 2011). Another factor may be increased exposure to the hospital environment, e.g. by caesarian section or prolonged hospitalization, since antibiotic resistance genes exists and can be transferred in such environments (Davies & Davies, 2010).

Some of the resistance genes were detected in all the samples belonging to a patient, whereas some of the resistance genes were only detected in some of the samples belonging to a patient. As the time frame between the sampling of these samples were short - and as the microbiota tends to be quite stable in the preterm gut - antibiotic exposure might be the reason. The antibiotic exposure might have created disturbance in the microbiota causing the variations, or changes in the antibiotic treatment itself might be the reason.

The diversity of resistance genes in the metagenome samples were also provided through functional analyses in MG-RAST. The subsystem Resistance to Antibiotics and Toxic Compounds contained several determinants for multidrug resistance, such as multidrug resistance efflux pumps, a multidrug resistance cluster and a multiple resistance locus. Together, these constituted 37 % of the subsystem. Unfortunately, the kind of resistance provided by these was not available. However, resistance to some single antibiotics was also detected in this subsystem. These included resistance to fluoroquinolones, methicillin and beta-lactam.

4.3 Microbiota Associated with NEC

Disagreements exist whether there is a difference in the microbial diversity in the gut between preterm infants with NEC and healthy preterm infants. Differences in specie richness within the samples, the α-diversity, were detected between the hospitals. However, no difference was observed between infants with NEC and healthy infants. Therefore, we could not detect a reduced microbial diversity in the infants with NEC, as in accordance with the study of Mai et al. (2011) and Torrazza et al. (2013).

On the other hand, β-diversity differences were detected between the samples. Weighted UniFrac PCoA showed a clustering of NEC positive samples. However, no clustering of the samples from the different hospitals was detected. Weighted UniFrac emphasizes the abundance of different species. An explanation of the clustering of the NEC positive samples could therefore be that some species were influential according to NEC. This was confirmed by further analyses.

Regards the microbial composition in the gut, preterm infants are thought to be mainly colonized with gram positive *Firmicutes* and gram negative *Proteobacteria* (Schwiertz et al., 2003; Y. Wang et al., 2009). These were also found to be the dominating phyla in this study. *Proteobacteria* was the dominating phylum in both infants with NEC and healthy infants, but the abundance of this phylum was higher in the patients with NEC. This increase of Proteobacteria in NEC cases has also been reported in other studies (Mai et al., 2011; McMurtry et al., 2015; Torrazza et al., 2013; Y. Wang et al., 2009).

The family Enterobacteriaceae dominated in the infants and was more abundant in infants with NEC. These bacteria are associated with NEC (Hunter et al., 2008) and they contains several genera, such as *Escherichia* and *Klebsiella*, that is commonly found in the gut of hospitalized preterm infants (Schwiertz et al., 2003). Some genera from this family were classified as "unclassified", which was the dominating group in both the infants with NEC and the healthy infants. Based on the results from both the classification from MG-RAST and the BLAST classification together with the relative abundance of influential OTUs, *Klebsiella* and *Escherichia* is likely to be the dominating bacteria in this group. A genus from this family that was classified, and was more abundant in the infants with NEC, was *Serratia*. This bacterium is not considered as a part of the normal microbiota in infants, but have been associated with hospital infections and found in stool samples from infants with NEC (Hällström, Eerola, Vuento, Janas, & Tammela, 2004).

Other dominating families in the infants were the *Enterococcaceae* and the *Staphylococaceae*. The *Enterococcaceae* family with the genus *Enterococcus* was found to be more abundant in infants with NEC. These bacteria has been described as persistent colonizers of the preterm gut (Aujoulat et al., 2014) and is commonly found in hospitalized preterm infants (Schwiertz et al., 2003). The *Staphylococcaceae* family with the genus *Staphylococcus* had a lower abundance in the infants with NEC. These bacteria is regarded as common early colonizers of

the preterm gut, but has also been reported in cases of NEC (Aujoulat et al., 2014; Torrazza et al., 2013).

The microbial colonization is thought to be important in the pathogenesis of NEC. Several pathogens have been associated with NEC, but none has been defined as definitively causative of NEC (Hunter et al., 2008; McMurtry et al., 2015; Neu & Walker, 2011). The family *Enterobacteriaceae* has been commonly associated with NEC. This is in line with our findings as two of the OTUs that were found to be influential according to NEC belonged to this family. These OTUs belonged to the genera *Escherichia* and *Klebsiella*. The amount of the OTU assigned as *Escherichia* constituted approximately 20 % of all the OTUs in infants with NEC, and was significantly higher in the patients with NEC compared to those without. *Escherichia* is one of the pathogens that is associated with NEC, but is also considered one of the first bacteria to colonize the gut (Wallace et al., 2011).

The amount of *Klebsiella* was slightly higher in infants with NEC. Strains of beta-lactam resistant *Klebsiella pneumoniae* have previously been associated with sepsis and outbreaks of NEC in hospitals (Cotton, Pieper, Kirsten, Orth, & Theron, 2001; Gregersen et al., 1999; Grishin, Papillon, Bell, Wang, & Ford, 2013). High amounts of *Klebsiella* in both groups can be explained as hospitalized preterm infants are more commonly colonized with these bacteria (Hunter et al., 2008). Other bacteria such as *Cronobacter sakazakii* and some clostridia strains have been associated with NEC (Grishin et al., 2013), but these associations were not detected in our study.

Absence of different bacteria has also been associated with NEC. McMurtry et al. (2015) proposed that absence of *Clostridia* could be associated with the development of NEC, as their immunoregulatory functions could be important to prevent an inflammatory response (McMurtry et al., 2015). In addition, absence of *Bifidobacteria* may also contribute to the development of NEC, as they may inhibit some clostridia to become pathogenic (Grishin et al., 2013). Neither of these results was in line with our findings, even though one OTU influential to NEC was assigned as *Bifidobacterium*. However, one OTU was detected showing a positive correlation to the healthy infants. This OTU was assigned as *Haemophilus*, and the amount of this OTU was significantly higher in the healthy patients. *Haemophilus* belongs to the gram negative bacteria and includes both commensally and pathogenic bacteria. That this bacterium was absent in the infants with NEC may be because it was more susceptible to the antibiotics or to be outcompeted by resistant or pathogenic bacteria.

4.4 Methodological Considerations

The *int*1 gene was used as a proxy for antibiotic resistance. There was several reasons for this; the gene is commonly linked to antibiotic resistance genes, it exits in both pathogenic and commensally human bacteria, its abundance can change rapidly both due to the hosts rapid generation time and because it easily transfers through horizontal gene transfer and, at last, because it is found in a range of different genetic elements (Gillings et al., 2014).

The DNA used for sequencing was extracted by different methods. This may create differences in the bacterial composition (Kennedy et al., 2014). This was addressed by using regression analysis for duplicate samples extracted with different methods. Most of these samples were similar, but to exclude the differences between the samples, all the statistics was based on average values for all samples belonging to each patient. Therefore, we believe in the significant results in this study.

Even though next generation sequencing technologies are considered accurate techniques, they are not infallible. Incomplete extension or addition of multiple nucleotides can lead to unsynchronized incorporation of additional bases (Michael L. Metzker, 2010). This causes fluorescence noise, base calling errors and shorter reads (Erlich, Mitra, delaBastide, McCombie, & Hannon, 2008). The most common error of the Illumina sequencing is substitution, as sequencing happens base by base (Huang, Li, Myers, & Marth, 2012). This error rate has been reported to be 0.1 substitutions per 100 bases on the MiSeq (Loman et al., 2012), which is considered quite low. Therefore, it is unlikely that sequencing errors caused by the MiSeq would influence our results.

Taxonomic assignment was done for both the 16S rRNA sequences and the metagenome sequences in this thesis. The same dominating bacteria were detected in both methods, but with small differences in abundance. However, the analyses done in this thesis are based on the taxonomic assignment for the 16S rRNA sequences. There are several reasons for this. First, the sequencing methods use different primers. The PRK primers used for the 16S rRNA sequencing are designed for classification and to cover both bacteria and archaea. In addition, they have a matching efficiency close to 87% (Y. Yu et al., 2005). Second, different applications were used for the taxonomy assignment. The Greengenes database assigned taxonomy to the 16S rRNA sequences through QIIME (J Gregory Caporaso et al., 2010), whereas the M5NR database assigned taxonomy to the metagenome sequences through MG-RAST (Meyer et al., 2008). QIIME is considered to give improved taxonomical classification

to that of MG-RAST (D'Argenio, Casaburi, Precone, & Salvatore, 2014) and is therefore more reliable.

Different determinants for antibiotic resistance genes were detected by ResFinder and MG-RAST. The results provided by ResFinder are considered more reliable and have therefore been prioritized in this study. The ResFinder database detects specific acquired resistance genes in the metagenome sequences using BLAST, whereas MG-RAST uses BLAT (BLAST-like alignment tool) to detect sequences in the metagenome dataset homologous to sequences in M5NR, and functional genes such as antibiotic resistance genes, is detected by comparing the homology of the genes against the Subsystems database. BLAT is less sensitive than BLAST (K. Yu & Zhang, 2013), and the subsystem containing antibiotic resistance genes provided by the Subsystems database contains incomplete information (Z. Wang et al., 2013).

4.5 Future Work

The preterm gut as a reservoir for antibiotic resistance genes and integrons is an important research field that has to be elaborated, as this may play an important role in the pathogenesis of NEC. Several antibiotic resistance genes were detected in this study, and it would be of interest to investigate how many of these that is actually located on the integron. Further, as the integron was associated with a plasmid, conjugation experiments would be useful to address if the plasmid can be transferred to other bacteria. In addition, to provide more information about the patients included in this study, such as mode of delivery, type of feeding and antibiotic use, can be valuable as these factors may relate to NEC.

5. Conclusion

Our results suggest that integrons are associated with NEC, as the abundance of integrons was close to be significantly higher in the preterm infants with NEC. Persistence of integrons was found in several patients and the integrons were associated with mobile genetic elements as transposons and plasmids. It is therefore possible that the integrons can be transferred to other bacteria in the microbiota and hence are persistent in the preterm gut. In addition, the bacterium *Escherichia coli* and had a significantly higher abundance in patients with NEC. However, the understanding of the pathogenicity of NEC is still incomplete. The role of the preterm gut as a reservoir for antibiotic resistance genes and integrons deserves future investigation, as this information may be useful for the prevention and treatment of NEC.

6. References

Aagaard, K., Ma, J., Antony, K. M., Ganu, R., Petrosino, J., & Versalovic, J. (2014). The placenta harbors a unique microbiome. *Sci Transl Med, 6*(237), 237ra265. doi: 10.1126/scitranslmed.3008599

Alexander, V. N., Northrup, V., & Bizzarro, M. J. (2011). Antibiotic Exposure in the Newborn Intensive Care Unit and the Risk of Necrotizing Enterocolitis. *The Journal of Pediatrics, 159*(3), 392-397. doi: http://dx.doi.org/10.1016/j.jpeds.2011.02.035

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., . . . Bork, P. (2011). Enterotypes of the human gut microbiome. *Nature, 473*(7346), 174-180. doi: 10.1038/nature09944

Aujoulat, F., Roudiere, L., Picaud, J. C., Jacquot, A., Filleron, A., Neveu, D., . . . Jumas-Bilak, E. (2014). Temporal dynamics of the very premature infant gut dominant microbiota. *BMC Microbiol, 14*(1), 2320. doi: 10.1186/s12866-014-0325-0

Avershina, E., Storro, O., Oien, T., Johnsen, R., Wilson, R., Egeland, T., & Rudi, K. (2013). Bifidobacterial succession and correlation networks in a large unselected cohort of mothers and their children. *Appl Environ Microbiol, 79*(2), 497-507. doi: 10.1128/aem.02359-12

Baharoglu, Z., Garriss, G., & Mazel, D. (2013). Multiple Pathways of Genome Plasticity Leading to Development of Antibiotic Resistance. *Antibiotics, 2*(2), 288-315. doi: 10.3390/antibiotics2020288

Bell, M. J., Ternberg, J. L., Feigin, R. D., Keating, J. P., Marshall, R., Barton, L., & Brotherton, T. (1978). Neonatal necrotizing enterocolitis. Therapeutic decisions based upon clinical staging. *Ann Surg, 187*(1), 1-7.

Blair, J. M., Webber, M. A., Baylay, A. J., Ogbolu, D. O., & Piddock, L. J. (2014). Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology*.

Buermans, H. P. J., & den Dunnen, J. T. (2014). Next generation sequencing technology: Advances and applications. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 1842*(10), 1932-1941. doi: http://dx.doi.org/10.1016/j.bbadis.2014.06.015

Cabezón, E., Ripoll-Rozada, J., Peña, A., de la Cruz, F., & Arechaga, I. (2014). Towards an integrated model of bacterial conjugation. *FEMS Microbiology Reviews*, n/a-n/a. doi: 10.1111/1574-6976.12085

Candela, M., Perna, F., Carnevali, P., Vitali, B., Ciati, R., Gionchetti, P., . . . Brigidi, P. (2008). Interaction of probiotic Lactobacillus and Bifidobacterium strains with human intestinal epithelial cells: adhesion properties, competition against enteropathogens and modulation of IL-8 production. *Int J Food Microbiol, 125*(3), 286-292. doi: 10.1016/j.ijfoodmicro.2008.04.012

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., . . . Gordon, J. I. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature methods, 7*(5), 335-336.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., . . . Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A, 108*(Suppl 1), 4516-4522. doi: 10.1073/pnas.1000080107

Collins, S. M. (2014). A role for the gut microbiota in IBS. *Nat Rev Gastroenterol Hepatol, 11*(8), 497-505. doi: 10.1038/nrgastro.2014.40

Collins, S. M., Surette, M., & Bercik, P. (2012). The interplay between the intestinal microbiota and the brain. *Nat Rev Micro, 10*(11), 735-742.

Cotton, M. F., Pieper, C. H., Kirsten, G. F., Orth, H., & Theron, D. C. (2001). Necrotising enterocolitis as an infectious disease--evidence from an outbreak of invasive disease due to extended-spectrum beta-lactamase-producing Klebsiella pneumoniae. *S Afr Med J, 91*(2), 133-135.

Cox, G., & Wright, G. D. (2013). Intrinsic antibiotic resistance: Mechanisms, origins, challenges and solutions. *International Journal of Medical Microbiology, 303*(6–7), 287-292. doi: http://dx.doi.org/10.1016/j.ijmm.2013.02.009

D'Argenio, V., Casaburi, G., Precone, V., & Salvatore, F. (2014). Comparative metagenomic analysis of human gut microbiome composition using two different bioinformatic pipelines. *BioMed research international, 2014*.

Davies, J., & Davies, D. (2010). Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev, 74*(3), 417-433. doi: 10.1128/mmbr.00016-10

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., . . . Turner, S. (2009). Real-time DNA sequencing from single polymerase molecules. *Science, 323*(5910), 133-138. doi: 10.1126/science.1162986

Erlich, Y., Mitra, P. P., delaBastide, M., McCombie, W. R., & Hannon, G. J. (2008). Alta-Cyclic: a self-optimizing base caller for next-generation sequencing. *Nat Meth, 5*(8), 679-682. doi: http://www.nature.com/nmeth/journal/v5/n8/suppinfo/nmeth.1230_S1.html

Espy, M. J., Uhl, J. R., Sloan, L. M., Buckwalter, S. P., Jones, M. F., Vetter, E. A., . . . Smith, T. F. (2006). Real-Time PCR in Clinical Microbiology: Applications for Routine Laboratory Testing. *Clinical Microbiology Reviews, 19*(1), 165-256. doi: 10.1128/CMR.19.1.165-256.2006

Fleming, A. (1929). On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. *British journal of experimental pathology, 10*(3), 226.

Frost, L. S., Leplae, R., Summers, A. O., & Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. *Nat Rev Micro, 3*(9), 722-732.

Gaustad, P. (2001). Mekanismer for utvikling av antibiotikaresistente bakterier. *Tidsskrift for Den norske legeforening, 26*. doi: 121:3090 – 4

Gillings, M. R. (2014). Integrons: past, present, and future. *Microbiol Mol Biol Rev, 78*(2), 257-277. doi: 10.1128/MMBR.00056-13

Gillings, M. R., Gaze, W. H., Pruden, A., Smalla, K., Tiedje, J. M., & Zhu, Y. G. (2014). Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *Isme j*. doi: 10.1038/ismej.2014.226

Goldenberg, R. L., Culhane, J. F., Iams, J. D., & Romero, R. (2008). Epidemiology and causes of preterm birth. *The lancet, 371*(9606), 75-84.

Gregersen, N., Van Nierop, W., Von Gottberg, A., Duse, A., Davies, V., & Cooper, P. (1999). Klebsiella pneumoniae with extended spectrum beta-lactamase activity associated with a necrotizing enterocolitis outbreak. *Pediatr Infect Dis J, 18*(11), 963-967.

Grishin, A., Papillon, S., Bell, B., Wang, J., & Ford, H. R. (2013). The Role of the Intestinal Microbiota in the Pathogenesis of Necrotizing Enterocolitis. *Seminars in pediatric surgery, 22*(2), 69-75. doi: 10.1053/j.sempedsurg.2013.01.002

Guerin, É., Cambray, G., Sanchez-Alberola, N., Campoy, S., Erill, I., Da Re, S., . . . Mazel, D. (2009). The SOS Response Controls Integron Recombination. *Science, 324*(5930), 1034. doi: 10.1126/science.1172914

Higuchi, R., Dollinger, G., Walsh, P. S., & Griffith, R. (1992). Simultaneous amplification and detection of specific DNA sequences. *Biotechnology (N Y), 10*(4), 413-417.

HMP. (2012). Structure, function and diversity of the healthy human microbiome. *Nature, 486*(7402), 207-214. doi:

http://www.nature.com/nature/journal/v486/n7402/abs/nature11234.html#supplementary-information

Hu, Y., Yang, X., Qin, J., Lu, N., Cheng, G., Wu, N., . . . Zhu, B. (2013). Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. *Nat Commun, 4*. doi: 10.1038/ncomms3151

Huang, W., Li, L., Myers, J. R., & Marth, G. T. (2012). ART: a next-generation sequencing read simulator. *Bioinformatics, 28*(4), 593-594. doi: 10.1093/bioinformatics/btr708

Huddleston, J. R. (2014). Horizontal gene transfer in the human gastrointestinal tract: potential spread of antibiotic resistance genes. *Infect Drug Resist, 7*, 167-176. doi: 10.2147/IDR.S48820

Hunter, C. J., Upperman, J. S., Ford, H. R., & Camerini, V. (2008). Understanding the Susceptibility of the Premature Infant to Necrotizing Enterocolitis (NEC). *Pediatr Res, 63*(2), 117-123.

Hällström, M., Eerola, E., Vuento, R., Janas, M., & Tammela, O. (2004). Effects of mode of delivery and necrotising enterocolitis on the intestinal microflora in preterm infants. *European Journal of Clinical Microbiology and Infectious Diseases, 23*(6), 463-470.

Jiménez, E., Marín, M. L., Martín, R., Odriozola, J. M., Olivares, M., Xaus, J., . . . Rodríguez, J. M. (2008). Is meconium from healthy newborns actually sterile? *Research in Microbiology, 159*(3), 187-193. doi: http://dx.doi.org/10.1016/j.resmic.2007.12.007

Johnsborg, O., Eldholm, V., & Håvarstein, L. S. (2007). Natural genetic transformation: prevalence, mechanisms and function. *Research in Microbiology, 158*(10), 767-778. doi: http://dx.doi.org/10.1016/j.resmic.2007.09.004

Johnston, C., Martin, B., Fichant, G., Polard, P., & Claverys, J.-P. (2014). Bacterial transformation: distribution, shared mechanisms and divergent control. *Nat Rev Micro, 12*(3), 181-196. doi: 10.1038/nrmicro3199

http://www.nature.com/nrmicro/journal/v12/n3/abs/nrmicro3199.html#supplementary-information

Kennedy, N. A., Walker, A. W., Berry, S. H., Duncan, S. H., Farquarson, F. M., Louis, P., . . . Hold, G. L. (2014). The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing. *PLoS ONE, 9*(2), e88982. doi: 10.1371/journal.pone.0088982

Kohanski, M. A., Dwyer, D. J., & Collins, J. J. (2010). How antibiotics kill bacteria: from targets to networks. *Nat Rev Micro, 8*(6), 423-435.

Kubista, M., Andrade, J. M., Bengtsson, M., Forootan, A., Jonak, J., Lind, K., . . . Zoric, N. (2006). The real-time polymerase chain reaction. *Mol Aspects Med, 27*(2-3), 95-125. doi: 10.1016/j.mam.2005.12.007

Labbate, M., Case, R. J., & Stokes, H. W. (2009). The integron/gene cassette system: an active player in bacterial adaptation. *Methods Mol Biol, 532*, 103-125. doi: 10.1007/978-1-60327-853-9_6

Li, B., Zhao, Z.-c., Wang, M.-h., Huang, X.-h., Pan, Y.-h., & Cao, Y.-p. (2014). Antimicrobial resistance and integrons of commensal Escherichia coli strains from healthy humans in China. *Journal of Chemotherapy (Florence, Italy), 26*(3), 190-192. doi: 10.1179/1973947813Y.0000000113

Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J., & Pallen, M. J. (2012). Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotech, 30*(5), 434-439. doi: 10.1038/nbt.2198

http://www.nature.com/nbt/journal/v30/n5/abs/nbt.2198.html#supplementary-information

Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K., & Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature*. doi: 10.1038/nature11550

Mai, V., Young, C. M., Ukhanova, M., Wang, X., Sun, Y., Casella, G., . . . Neu, J. (2011). Fecal Microbiota in Premature Infants Prior to Necrotizing Enterocolitis. *PLoS ONE, 6*(6), e20647. doi: 10.1371/journal.pone.0020647

Marques, T. M., Wall, R., Ross, R. P., Fitzgerald, G. F., Ryan, C. A., & Stanton, C. (2010). Programming infant gut microbiota: influence of dietary and environmental factors. *Current Opinion in Biotechnology, 21*(2), 149-156. doi: http://dx.doi.org/10.1016/j.copbio.2010.03.020

Mazel, D. (2006). Integrons: agents of bacterial evolution. *Nat Rev Microbiol, 4*(8), 608-620. doi: 10.1038/nrmicro1462

McMurtry, V. E., Gupta, R. W., Tran, L., Blanchard, E. E., Penn, D., Taylor, C. M., & Ferris, M. J. (2015). Bacterial diversity and Clostridia abundance decrease with increasing severity of necrotizing enterocolitis. *Microbiome, 3*, 11. doi: 10.1186/s40168-015-0075-8

Metzker, M. L. (2005). Emerging technologies in DNA sequencing. *Genome research, 15*(12), 1767-1776.

Metzker, M. L. (2010). Sequencing technologies [mdash] the next generation. *Nat Rev Genet, 11*(1), 31-46.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., . . . Wilke, A. (2008). The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC bioinformatics, 9*(1), 386.

Morrow, A. L., Lagomarcino, A. J., Schibler, K. R., Taft, D. H., Yu, Z., Wang, B., . . . Newburg, D. S. (2013). Early microbial and metabolomic signatures predict later onset of necrotizing enterocolitis in preterm infants. *Microbiome, 1*(1), 13. doi: 10.1186/2049-2618-1-13

Nadkarni, M. A., Martin, F. E., Jacques, N. A., & Hunter, N. (2002). Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology, 148*(1), 257-266.

Neu, J., & Walker, W. A. (2011). Necrotizing Enterocolitis. *New England Journal of Medicine, 364*(3), 255-264. doi: doi:10.1056/NEJMra1005408

Olsen, G. J., & Woese, C. R. (1993). Ribosomal RNA: a key to phylogeny. *Faseb j, 7*(1), 113-123.

Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A., & Brown, P. O. (2007). Development of the human infant intestinal microbiota. *PLoS biology, 5*(7), e177.

Panigrahi, P. (2006). Necrotizing enterocolitis: a practical guide to its prevention and management. *Paediatr Drugs, 8*(3), 151-165.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., . . . Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature, 464*(7285), 59-65. doi: http://www.nature.com/nature/journal/v464/n7285/suppinfo/nature08821_S1.html

Raveh-Sadka, T., Thomas, B. C., Singh, A., Firek, B., Brooks, B., Castelle, C. J., . . . Banfield, J. F. (2015). Gut bacteria are rarely shared by co-hospitalized premature infants, regardless of necrotizing enterocolitis development. *eLife, 4*, e05477. doi: 10.7554/eLife.05477

Round, J. L., & Mazmanian, S. K. (2009). The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol, 9*(5), 313-323.

Ruijter, J. M., Ramakers, C., Hoogaars, W. M., Karlen, Y., Bakker, O., van den Hoff, M. J., & Moorman, A. F. (2009). Amplification efficiency: linking baseline and bias in the

analysis of quantitative PCR data. *Nucleic Acids Res, 37*(6), e45. doi: 10.1093/nar/gkp045

Salonen, A., Salojarvi, J., Lahti, L., & de Vos, W. M. (2012). The adult intestinal core microbiota is determined by analysis depth and health status. *Clin Microbiol Infect, 18 Suppl 4*, 16-20. doi: 10.1111/j.1469-0691.2012.03855.x

Salyers, A. A., Gupta, A., & Wang, Y. (2004). Human intestinal bacteria as reservoirs for antibiotic resistance genes. *Trends in Microbiology, 12*(9), 412-416. doi: http://dx.doi.org/10.1016/j.tim.2004.07.004

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A, 74*(12), 5463-5467.

Schwiertz, A., Gruhl, B., Lobnitz, M., Michel, P., Radke, M., & Blaut, M. (2003). Development of the Intestinal Bacterial Composition in Hospitalized Preterm Infants in Comparison with Breast-Fed, Full-Term Infants. *Pediatr Res, 54*(3), 393-399.

Segata, N., Boernigen, D., Tickle, T. L., Morgan, X. C., Garrett, W. S., & Huttenhower, C. (2013). *Computational meta'omics for microbial community studies* (Vol. 9).

Sekelja, M., Berget, I., Naes, T., & Rudi, K. (2011). Unveiling an abundant core microbiota in the human adult colon by a phylogroup-independent searching approach. *Isme j, 5*(3), 519-531. doi: 10.1038/ismej.2010.129

Sommer, F., & Backhed, F. (2013). The gut microbiota--masters of host development and physiology. *Nat Rev Microbiol, 11*(4), 227-238. doi: 10.1038/nrmicro2974

Stalder, T., Barraud, O., Casellas, M., Dagot, C., & Ploy, M. C. (2012). Integron involvement in environmental spread of antibiotic resistance. *Front Microbiol, 3*, 119. doi: 10.3389/fmicb.2012.00119

Stewart, C. J., & Cummings, S. P. (2015). Gut bacteria and necrotizing enterocolitis: cause or effect? *Trends in Microbiology*(0). doi: http://dx.doi.org/10.1016/j.tim.2015.04.005

Sørum, H., L'Abée-Lund, T. M., Solberg, A., & Wold, A. (2003). Integron-Containing IncU R Plasmids pRAS1 and pAr-32 from the Fish Pathogen Aeromonas salmonicida. *Antimicrobial Agents and Chemotherapy, 47*(4), 1285-1290. doi: 10.1128/AAC.47.4.1285-1290.2003

Thompson, A. M., & Bizzarro, M. J. (2008). Necrotizing Enterocolitis in Newborns: Pathogenesis, Prevention and Management. *Drugs, 68*(9), 1227-1238.

Torrazza, R. M., Ukhanova, M., Wang, X., Sharma, R., Hudak, M. L., Neu, J., & Mai, V. (2013). Intestinal Microbial Ecology and Environmental Factors Affecting Necrotizing Enterocolitis. *PLoS ONE, 8*(12), e83304. doi: 10.1371/journal.pone.0083304

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The human microbiome project. *Nature, 449*(7164), 804-810. doi: 10.1038/nature06244

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Res, 40*(15), e115. doi: 10.1093/nar/gks596

van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics, 30*(9), 418-426. doi: http://dx.doi.org/10.1016/j.tig.2014.07.001

Vinue, L., Saenz, Y., Somalo, S., Escudero, E., Moreno, M. A., Ruiz-Larrea, F., & Torres, C. (2008). Prevalence and diversity of integrons and associated resistance genes in faecal Escherichia coli isolates of healthy humans in Spain. *J Antimicrob Chemother, 62*(5), 934-937. doi: 10.1093/jac/dkn331

Wallace, T. C., Guarner, F., Madsen, K., Cabana, M. D., Gibson, G., Hentges, E., & Sanders, M. E. (2011). Human gut microbiota and its relationship to health and disease. *Nutrition Reviews, 69*(7), 392-403. doi: 10.1111/j.1753-4887.2011.00402.x

Wang, Y., Hoenig, J. D., Malin, K. J., Qamar, S., Petrof, E. O., Sun, J., . . . Claud, E. C. (2009). 16S rRNA gene-based analysis of fecal microbiota from preterm infants with and without necrotizing enterocolitis. *Isme j, 3*(8), 944-954. doi: 10.1038/ismej.2009.37

Wang, Z., Zhang, X.-X., Huang, K., Miao, Y., Shi, P., Liu, B., . . . Li, A. (2013). Metagenomic Profiling of Antibiotic Resistance Genes and Mobile Genetic Elements in a Tannery Wastewater Treatment Plant. *PLoS ONE, 8*(10), e76079. doi: 10.1371/journal.pone.0076079

Westerbeek, E. A. M., van den Berg, A., Lafeber, H. N., Knol, J., Fetter, W. P. F., & van Elburg, R. M. (2006). The intestinal bacterial colonisation in preterm infants: A review of the literature. *Clinical Nutrition, 25*(3), 361-368. doi: http://dx.doi.org/10.1016/j.clnu.2006.03.002

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

Yu, K., & Zhang, T. (2013). Construction of Customized Sub-Databases from NCBI-nr Database for Rapid Annotation of Huge Metagenomic Datasets Using a Combined BLAST and MEGAN Approach. *PLoS ONE, 8*(4), e59831. doi: 10.1371/journal.pone.0059831

Yu, Y., Lee, C., Kim, J., & Hwang, S. (2005). Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. *Biotechnology and Bioengineering, 89*(6), 670-679. doi: 10.1002/bit.20347

Yunwei Wang, Jeanette D Hoenig, Kathryn J Malin, Sanaa Qamar, Elaine O Petrof, Jun Sun, . . . Claud, E. C. (2009). 16S rRNA gene-based analysis of fecal microbiota from preterm infants with and without necrotizing enterocolitis. *Isme j*. doi: doi:10.1038/ismej.2009.37

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., . . . Larsen, M. V. (2012). Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother, 67*(11), 2640-2644. doi: 10.1093/jac/dks261

Zhang, L., Kinkelaar, D., Huang, Y., Li, Y., Li, X., & Wang, H. H. (2011). Acquired antibiotic resistance: are we born with it? *Appl Environ Microbiol, 77*(20), 7134-7141. doi: 10.1128/aem.05087-11

7. Appendix

Appendix A: Patient Information

Table A.1 Patient information.

Patient	N[1)]	Institution	Gender[2)]	Gestational age (weeks)	Birth weight (g)	NEC[3)]	Abx 1[4)]	Abx 2[5)]
1	1	Boston	M	27,1	770	Y	Yes	Yes
2	1	Boston	M	27,1	970	N	No	Yes
3	1	Boston	M	27,3	630	N	No	Yes
4	7	Boston	F	25,0	710	Y	Yes	Yes
5	6	Boston	F	25,2	740	N	No	Yes
6	5	Boston	M	24,6	720	N	No	Yes
7	4	Boston	F	32,1	1235	Y	No	Yes
8	4	Boston	F	32,3	1325	N	No	Yes
9	4	Boston	F	32,5	1865	N	No	Yes
10	4	Boston	M	28,4	960	Y	No	Yes
11	4	Boston	M	28,0	1015	N	No	Yes
12	3	Boston	F	29,3	1075	N	No	Yes
13	5	Boston	F	29,3	1090	Y	No	Yes
14	5	Boston	M	29,4	1515	N	No	Yes
15	5	Boston	F	30,4	1485	N	Yes	Yes
16	7	Boston	F	31,6	1560	Y	Yes	Yes
17	6	Boston	F	31,6	1905	N	No	Yes
18	6	Boston	F	31,6	1860	N	No	Yes
19	3	Boston	F	27,6	1190	Y	No	Yes
20	2	Boston	F	27,6	1110	N	No	Yes
21	3	Boston	M	27,4	470	N	No	Yes
22	1	Chicago	M	23,2	1110	Y	Yes	N.d
23	1	Chicago	M	23,0	650	N	Yes	N.d
24	1	Chicago	M	24,1	680	N	No	N.d
25	1	Chicago	M	23,5	700	Y	No	N.d
26	1	Chicago	F	24,0	660	N	Yes	N.d
27	1	Chicago	M	24,3	905	N	No	N.d
28	1	Chicago	M	24,0	620	Y	Yes	N.d
29	1	Chicago	M	24,0	640	N	No	N.d
30	1	Chicago	M	24,1	680	N	No	N.d
31	1	Chicago	M	25,2	850	Y	Yes	N.d
32	1	Chicago	M	25,0	710	N	No	N.d
33	1	Chicago	M	25,6	820	N	No	N.d

34	1	Chicago	M	28,3	810	Y	No	N.d
35	1	Chicago	M	28,4	945	N	No	N.d
36	1	Chicago	M	28,5	1360	N	Yes	N.d
37	1	Chicago	M	30,2	1389	Y	No	N.d
38	1	Chicago	M	30,6	1526	N	No	N.d
39	1	Chicago	F	30,0	1260	N	No	N.d
40	1	Chicago	F	24,4	600	Y	Yes	N.d
41	1	Chicago	M	24,1	680	N	Yes	N.d
42	1	Chicago	M	24,3	905	N	No	N.d
43	1	Evanston	N.d	N.d	N.d	Y	N.d	N.d
44	1	Evanston	N.d	N.d	N.d	N	N.d	N.d
45	1	Evanston	N.d	N.d	N.d	N	N.d	N.d
46	1	Evanston	N.d	N.d	N.d	Y	N.d	N.d
47	1	Evanston	N.d	N.d	N.d	N	N.d	N.d
48	1	Evanston	N.d	N.d	N.d	N	N.d	N.d
49	1	Evanston	N.d	N.d	N.d	Y	N.d	N.d
50	1	Evanston	N.d	N.d	N.d	N	N.d	N.d
51	1	Evanston	N.d	N.d	N.d	N	N.d	N.d
52	5	Boston	F	32,0	1835	Y	No	Yes
53	5	Boston	F	32,2	2035	N	Yes	Yes
54	5	Boston	F	32,2	1620	N	No	Yes
55	5	Chicago	N.d	27,0	N.d	Y	Yes	N.d
56	6	Chicago	N.d	25,2	N.d	Y	No	N.d
57	7	Chicago	N.d	29,3	N.d	Y	Yes	N.d
58	4	Chicago	N.d	24,1	N.d	Y	No	N.d
59	3	Chicago	N.d	24,0	N.d	Y	Yes	N.d
60	6	Chicago	N.d	30,5	N.d	Y	No	N.d
61	4	Chicago	N.d	30,5	N.d	Y	Yes	N.d
62	5	Chicago	N.d	26,5	N.d	Y	No	N.d
63	N.d	Chicago	F	24,2	675	N	N.d	N.d
64	N.d	Chicago	F	39,4	2820	N	N.d	N.d
65	3	Chicago	F	24,1	530	Y	N.d	N.d
66	2	Chicago	M	25,0	690	Y	N.d	N.d
67	3	Boston	M	29.2	770	N	N.d	N.d
68	3	Boston	F	25.0	710	Y	Yes	Yes

1) N = number of samples, N.d = Not defined

2) M = Male, F = Female

3) Y = Yes, N = No

4) Abx 1: Received antibiotics 1-3 days prior to sampling

5) Abx 2: Have previously received antibiotics.

Appendix B: PRK Illumina Primers

PRK Illumina forward primers (5' - 3'):

1. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctagtcaaCCTACGGGRBGCASCAG
2. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctagttccCCTACGGGRBGCASCAG
3. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctatgtcaCCTACGGGRBGCASCAG
4. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctccgtccCCTACGGGRBGCASCAG
5. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtagagCCTACGGGRBGCASCAG
6. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtccgcCCTACGGGRBGCASCAG
7. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtgaaaCCTACGGGRBGCASCAG
8. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtggccCCTACGGGRBGCASCAG
9. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtttcgCCTACGGGRBGCASCAG
10. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctcgtacgCCTACGGGRBGCASCAG
11. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgagtggCCTACGGGRBGCASCAG
12. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctggtagcCCTACGGGRBGCASCAG
13. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctactgatCCTACGGGRBGCASCAG
14. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctatgagcCCTACGGGRBGCASCAG
15. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctattcctCCTACGGGRBGCASCAG
16. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctcaaaagCCTACGGGRBGCASCAG

PRK Illumina reverse primers (5' - 3'):

1. caagcagaagacggcatacgagatcgtgatgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
2. caagcagaagacggcatacgagatacatcggtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
3. caagcagaagacggcatacgagatgcctaagtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
4. caagcagaagacggcatacgagattggtcagtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
5. caagcagaagacggcatacgagatcactctgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
6. caagcagaagacggcatacgagatattggcgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
7. caagcagaagacggcatacgagatgatctggtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
8. caagcagaagacggcatacgagattcaagtgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
9. caagcagaagacggcatacgagatctgatcgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
10. caagcagaagacggcatacgagataagctagtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
11. caagcagaagacggcatacgagatgtagccgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
12. caagcagaagacggcatacgagattacaaggtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
13. caagcagaagacggcatacgagatttgactgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
14. caagcagaagacggcatacgagatggaactgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
15. caagcagaagacggcatacgagattgacatgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
16. caagcagaagacggcatacgagatggacgggtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT

17. caagcagaagacggcatacgagatctctacgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
18. caagcagaagacggcatacgagatgcggacgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
19. caagcagaagacggcatacgagattttcacgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
20. caagcagaagacggcatacgagatggccacgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
21. caagcagaagacggcatacgagatcgaaacgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
22. caagcagaagacggcatacgagatcgtacggtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
23. caagcagaagacggcatacgagatccactcgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
24. caagcagaagacggcatacgagatgctaccgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
25. caagcagaagacggcatacgagatatcagtgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
26. caagcagaagacggcatacgagatgctcatgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
27. caagcagaagacggcatacgagataggaatgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
28. caagcagaagacggcatacgagatcttttggtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
29. caagcagaagacggcatacgagattagttggtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
30. caagcagaagacggcatacgagatccggtggtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
31. caagcagaagacggcatacgagatatcgtggtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
32. caagcagaagacggcatacgagattgagtggtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
33. caagcagaagacggcatacgagatcgcctggtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
34. caagcagaagacggcatacgagatgccatggtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
35. caagcagaagacggcatacgagataaaatggtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT
36. caagcagaagacggcatacgagattgttgggtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT

Appendix C: Unweighted UniFrac Plot

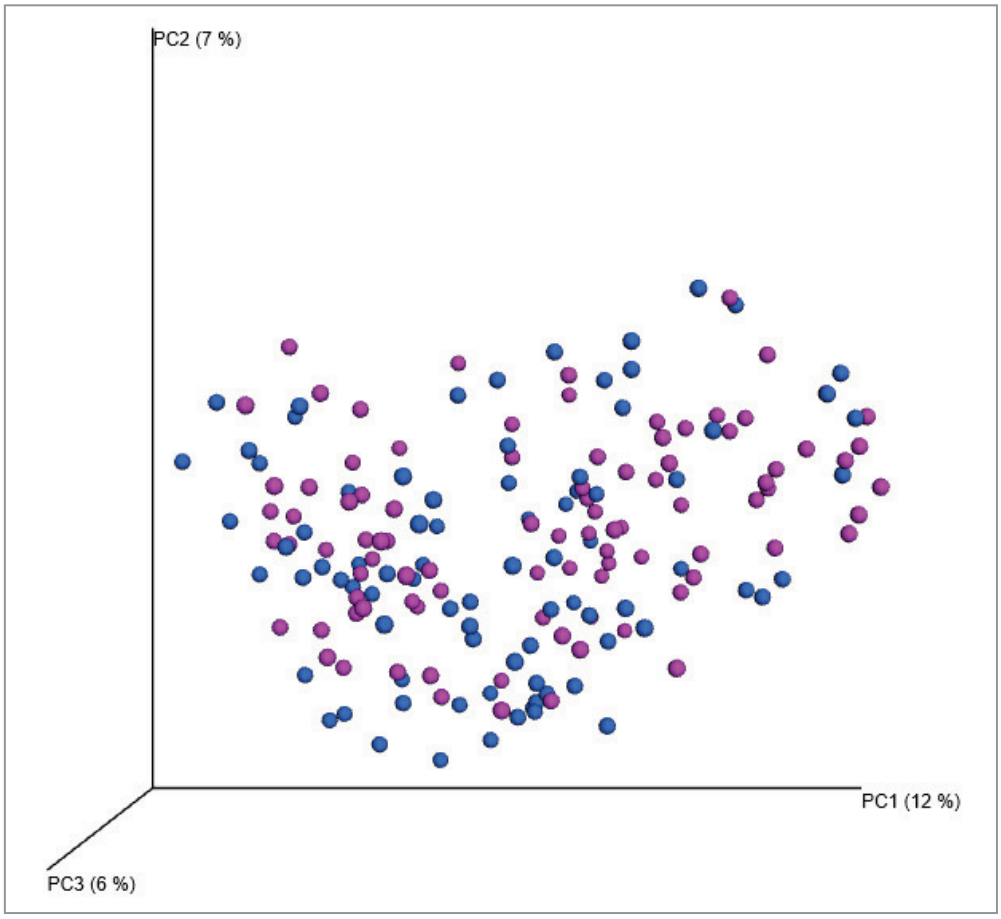


Figure A.1 Unweighted UniFrac PCoA plot. Samples from patients with NEC are purple and samples from patients without NEC are blue.

Appendix D: Bacteria Composition at Phylum Level

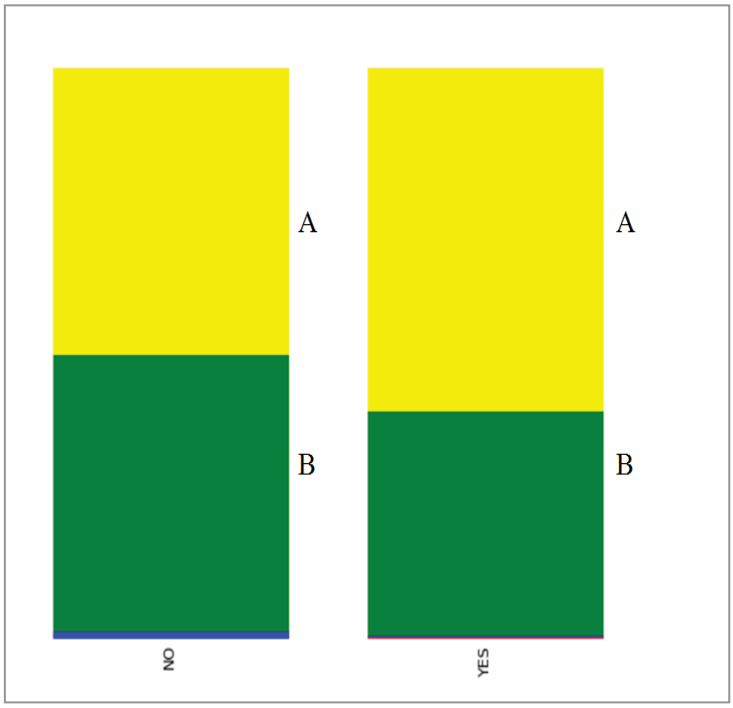


Figure A.2 Taxonomic composition at phylum level. Samples belonging to patients with (Yes) and without (No) NEC are compared. A = *Proteobacteria*, B = *Firmicutes*.

Appendix E: HRM Analysis of *int*1 Positive Samples

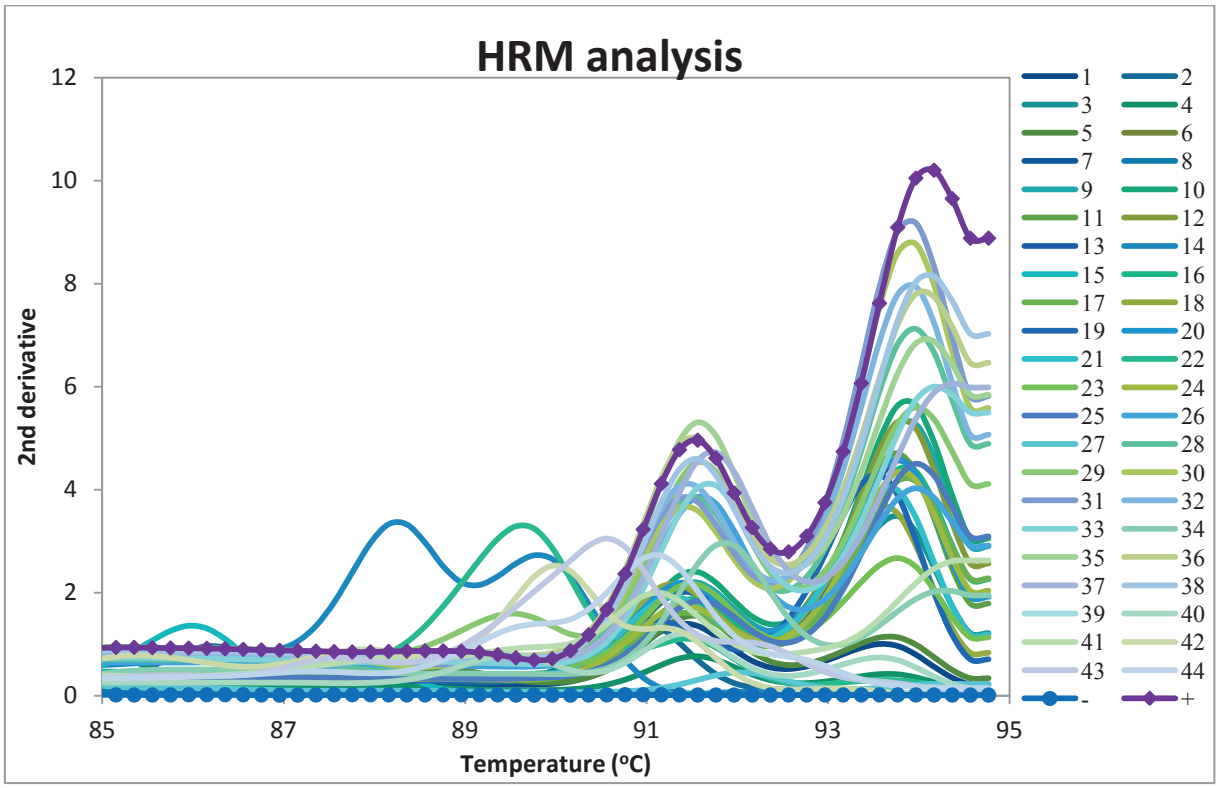


Figure A.3 HRM analysis of possible *int*1 positive samples. The negative control (-) and the positive (*E. coli*) control (+) is indicator marked. As showed in the positive control, the characteristic *int*1 melting curve has a peak at ~92°C followed by complete denaturation at 94°C.

Appendix F: Agarose Gel of *int*1 Positive Samples

Figure A.4 Gel electrophoresis of the *int*1 PCR products for the possible *int*1 positive samples. A negative control (-) and a positive (*E. coli*) control (+) were included. The positive control shows the *int*1 fragment (~530 bp) highlighted by the arrow. The contrast is increased for better visualization of the fragments.

Appendix G: Functional Subsystems

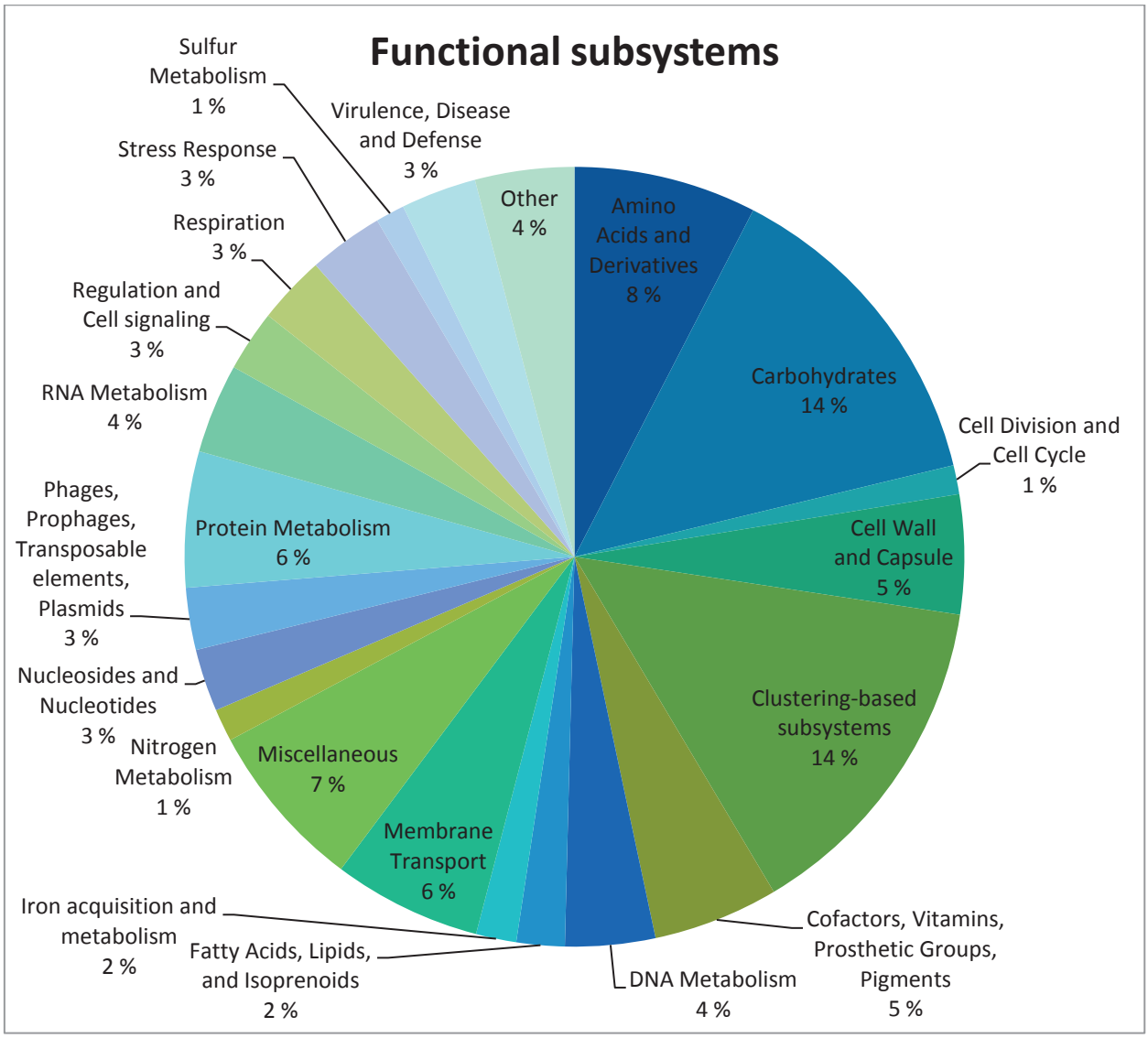


Figure A.5 Functional subsystems annotated by the Subsystems database (average).

Appendix H: Resistance Genes

Table A.2 Resistance genes detected with ResFinder in metagenome samples from different patients.

P[1)]	S[2)]	I[3)]	NEC[4)]	Gene	Identity	Query/HSP	Resistance	Accession number
9	9	B	N	blaTEM-1B	100.00	861/861	Beta-lactam	JF910132
				mph(A)	99.67	921/922	Macrolide	U36578
				blaCTX-M-15	100.00	876/784	Beta-lactam	DQ302097
				sul1	100.00	927/927	Sulphonamide	CP002151
				aadA5	100.00	789/789	Aminoglycoside	AF137361
				dfrA17	100.00	474/474	Trimethoprim	FJ460238
9	112	B	N	mph(A)	100.00	906/906	Macrolide	D16251
				blaTEM-1B	100.00	861/861	Beta-lactam	JF910132
				dfrA17	100.00	474/474	Trimethoprim	FJ460238
				aadA5	100.00	789/789	Aminoglycoside	AF137361
				blaCTX-M-15	100.00	876/876	Beta-lactam	DQ302097
				sul1	100.00	927/927	Sulphonamide	CP002151
9	113	B	N	blaTEM-1B	100.00	861/861	Beta-lactam	JF910132
				mph(A)	99.67	921/922	Macrolide	U36578
				blaCTX-M-15	100.00	876/876	Beta-lactam	DQ302097
				sul1	100.00	927/927	Sulphonamide	CP002151
				aadA5	100.00	789/789	Aminoglycoside	AF137361
				dfrA17	100.00	474/474	Trimethoprim	FJ460238
9	114	B	N	mph(A)	100.00	906/906	Macrolide	D16251
				blaTEM-1B	100.00	861/861	Beta-lactam	JF910132
				blaCTX-M-15	100.00	876/775	Beta-lactam	DQ302097
				sul1	100.00	927/927	Sulphonamide	CP002151
				aadA5	100.00	789/789	Aminoglycoside	AF137361
				dfrA17	100.00	474/474	Trimethoprim	FJ460238
13	13	B	Y	fosA	99.76	420/420	Fosfomycin	NZ_ACWO01000079
				blaSHV-11	99.88	861/861	Beta-lactam	EF035558
13	65	B	Y	fosA	99.76	420/420	Fosfomycin	NZ_ACWO01000079
				erm(C)	100.00	735/449	Macrolide	Y09001
13	66	B	Y	fosA	99.76	420/420	Fosfomycin	NZ_ACWO01000079
				blaSHV-11	99.88	861/861	Beta-lactam	EF035558
13	67	B	Y	fosA	99.76	420/420	Fosfomycin	NZ_ACWO01000079
49	49	E	Y	sul1	100.00	927/927	Sulphonamide	CP002151
				aadA1	99.75	792/792	Aminoglycoside	JQ414041
				blaOXA-1	100.00	831/831	Beta-lactam	J02967
				blaTEM-1B	100.00	861/861	Beta-lactam	JF910132
				strB	100.00	837/837	Aminoglycoside	M96392
				tet(B)	100.00	1206/1206	Tetracycline	AF326777
				strA	100.00	804/804	Aminoglycoside	M96392
				sul2	100.00	816/816	Sulphonamide	GQ421466
				catA1	99.85	660/660	Phenicol	V00622

				dfrA5	100.00	474/474	Trimethoprim	X12868
58	151	C	Y	ant(6)-Ia	100.00	909/610	Aminoglycoside	AF330699
				mph(C)	100.00	900/816	Macrolide	AF167161
				msr(A)	98.60	1467/931	Macrolide, Lincosamide, Streptogramin B	X52085
				erm(B)	100.00	738/738	Macrolide	U18931
				tet(J)	99.25	1197/1197	Tetracycline	ACLE01000065
				aph(3')-III	99.87	795/795	Aminoglycoside	M26832
				lsa(A)	98.93	1497/1497	Macrolide	AY225127
58	152	C	Y	blaDHA-1	99.78	1140/912	Beta-lactam	Y16410
				blaTEM-1B	100.00	861/861	Beta-lactam	JF910132
58	153	C	Y	blaTEM-1B	100.00	861/861	Beta-lactam	JF910132
62	169	C	Y	aadA2	100.00	792/792	Aminoglycoside	JQ364967
				dfrA12	100.00	498/498	Trimethoprim	AB571791
				fosA	99.05	420/420	Fosfomycin	NZ_ACWO01000079
				aadA1	100.00	792/792	Aminoglycoside	JQ414041
				oqxA	99.57	1176/1176	Quinolone	EU370913
				oqxB	98.69	2450/2450	Quinolone	EU370913
				sul1	100.00	927/927	Sulphonamide	CP002151
				blaSHV-11	99.88	861/861	Beta-lactam	FJ483937
62	170	C	Y	aadA1	100.00	792/792	Aminoglycoside	JQ414041
				oqxB	98.69	2450/2450	Quinolone	EU370913
				blaACT-15	99.74	1146/1146	Beta-lactam	JX440356
				oqxA	99.57	1176/1176	Quinolone	EU370913
				sul1	100.00	927/927	Sulphonamide	CP002151
				fosA	99.05	420/420	Fosfomycin	NZ_ACWO01000079
				blaSHV-11	99.88	861/861	Beta-lactam	FJ483937
62	171	C	Y	fosA	99.05	420/420	Fosfomycin	NZ_ACWO01000079
				sul1	100.00	927/927	Sulphonamide	CP002151
				oqxA	99.57	1176/1176	Quinolone	EU370913
				dfrA12	100.00	498/325	Trimethoprim	AB571791
				oqxB	98.81	2450/1764	Quinolone	EU370913
				erm(B)	100.00	738/738	Macrolide	U86375
				QnrB19	99.84	645/645	Quinolone	HM146784
				erm(X)	99.61	762/763	Macrolide	X51472
				aadA1	100.00	792/792	Aminoglycoside	JQ414041

1) P = Patient

2) S = Sample

3) I = Institution, B = Boston, E = Evanston, C = Chicago

4) N = No, Y = Yes

Appendix I: *Escherichia coli* Plasmid p1658/97

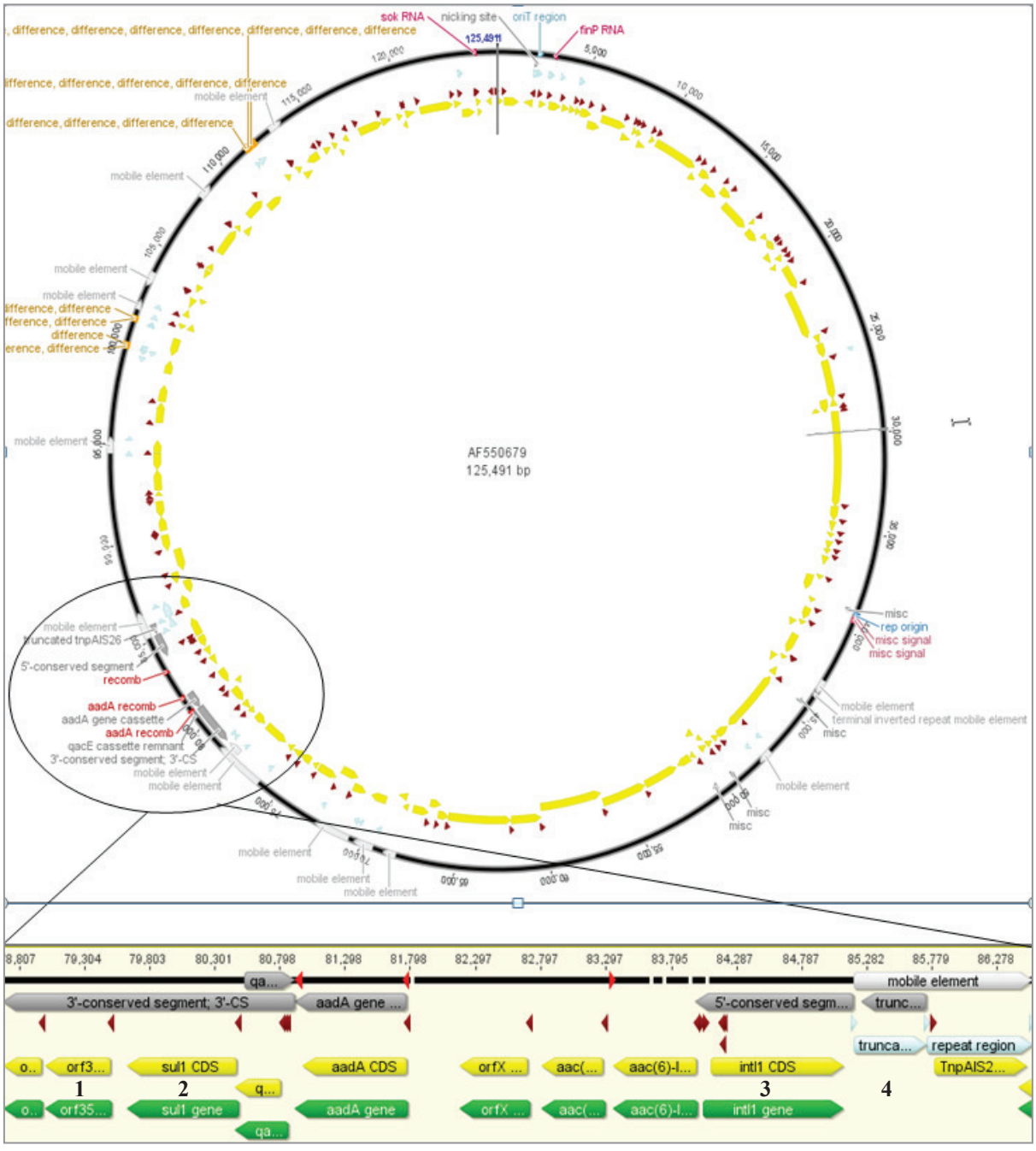


Figure A.6 The *E. coli* plasmid p1658/97. The plasmid contains ORFs (1), the sulfonamide (*sul*1) gene (2), the class 1 integron integrase gene (*int*1) (3) and a truncated transposon (4). The highlighted area was maximized and added beneath the plasmid (Estensmo, this thesis).

Appendix J: Sequence Alignment to *Escherichia coli* Plasmid p1658/97

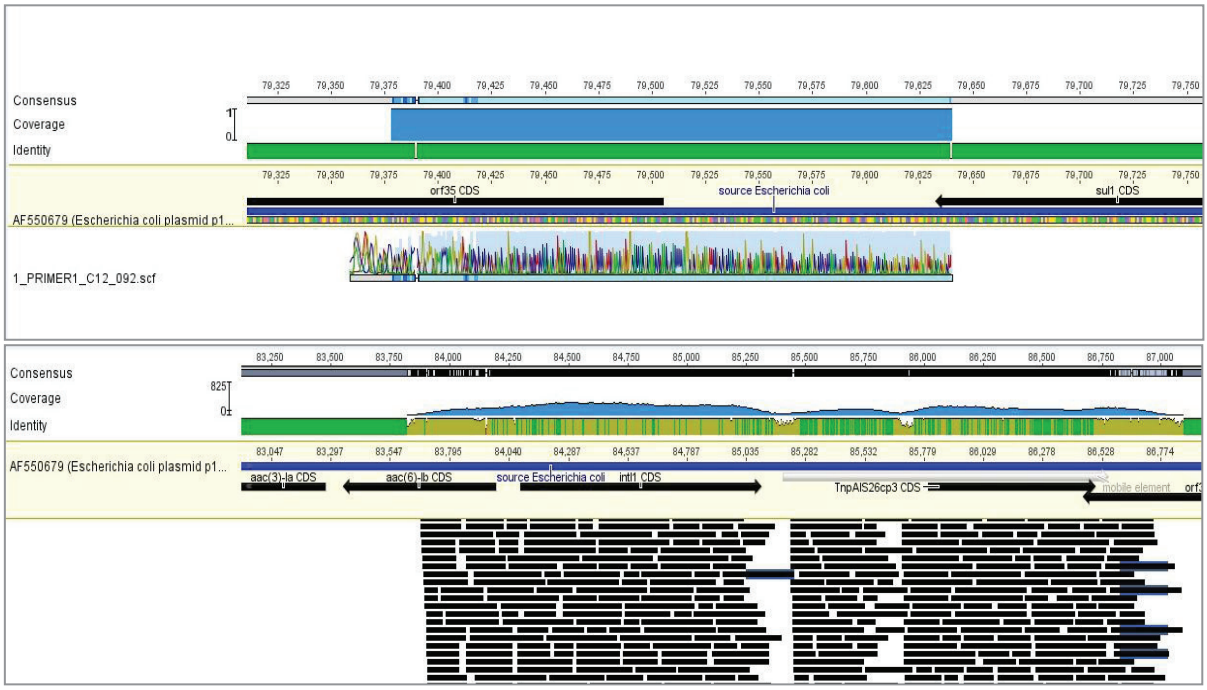
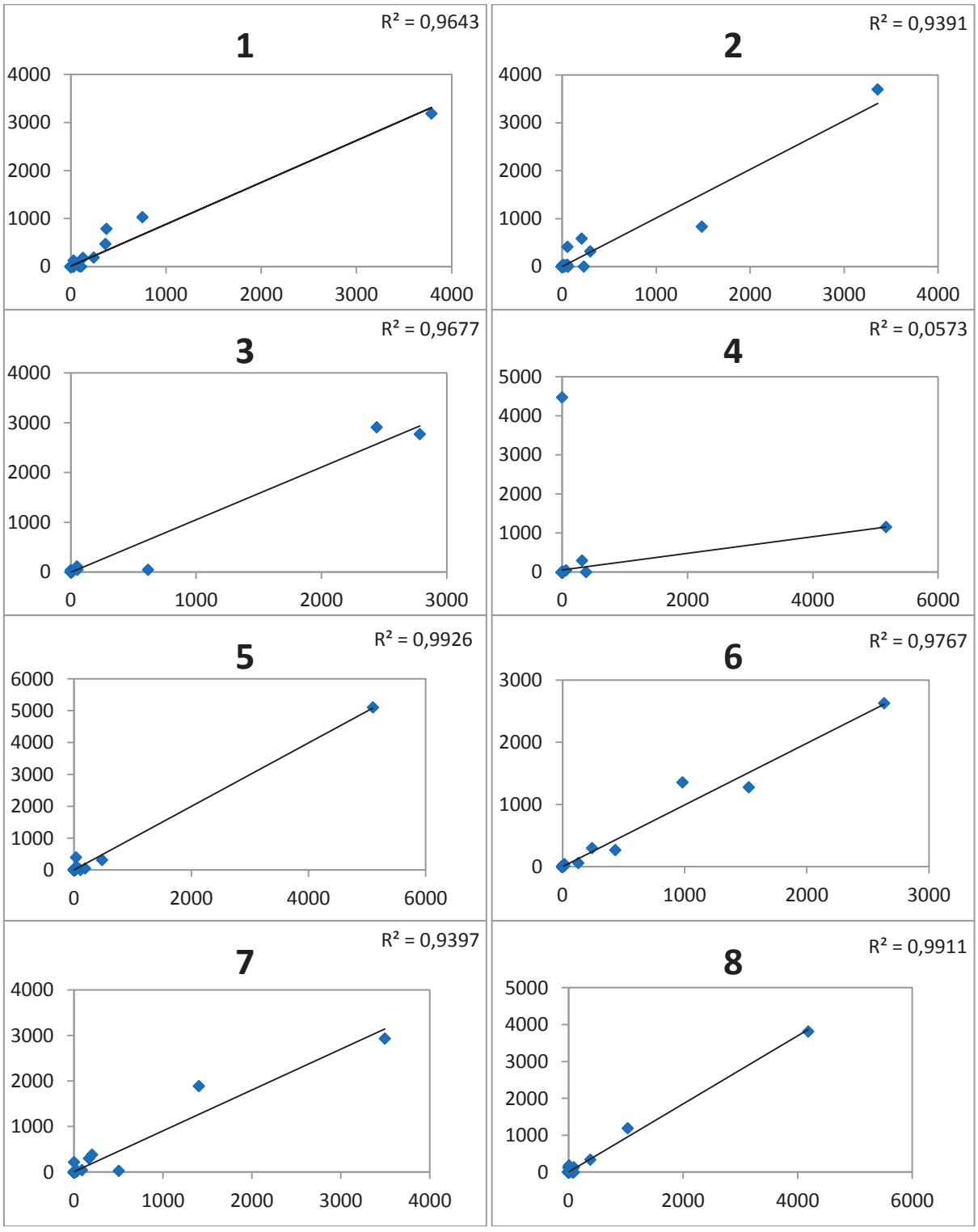
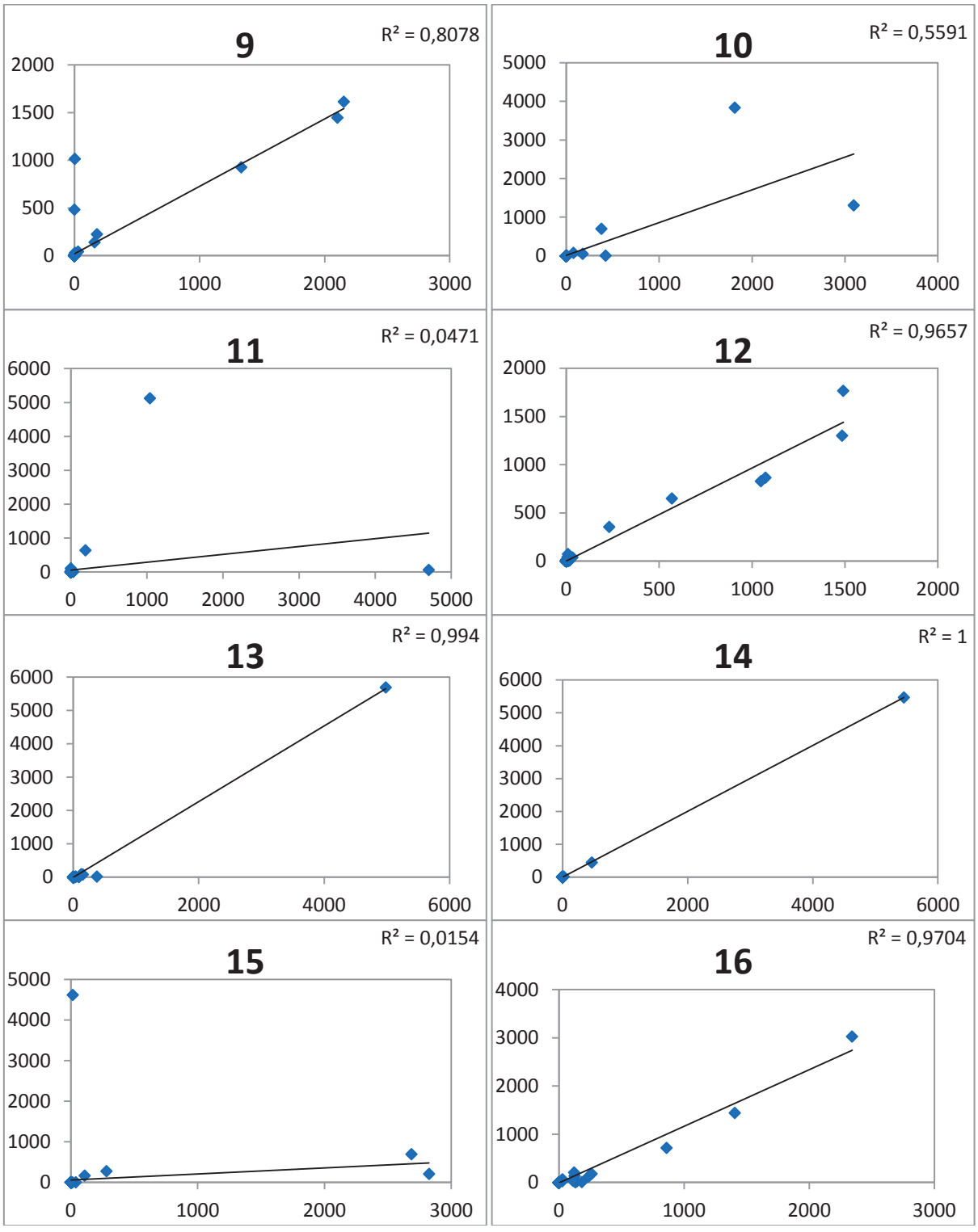
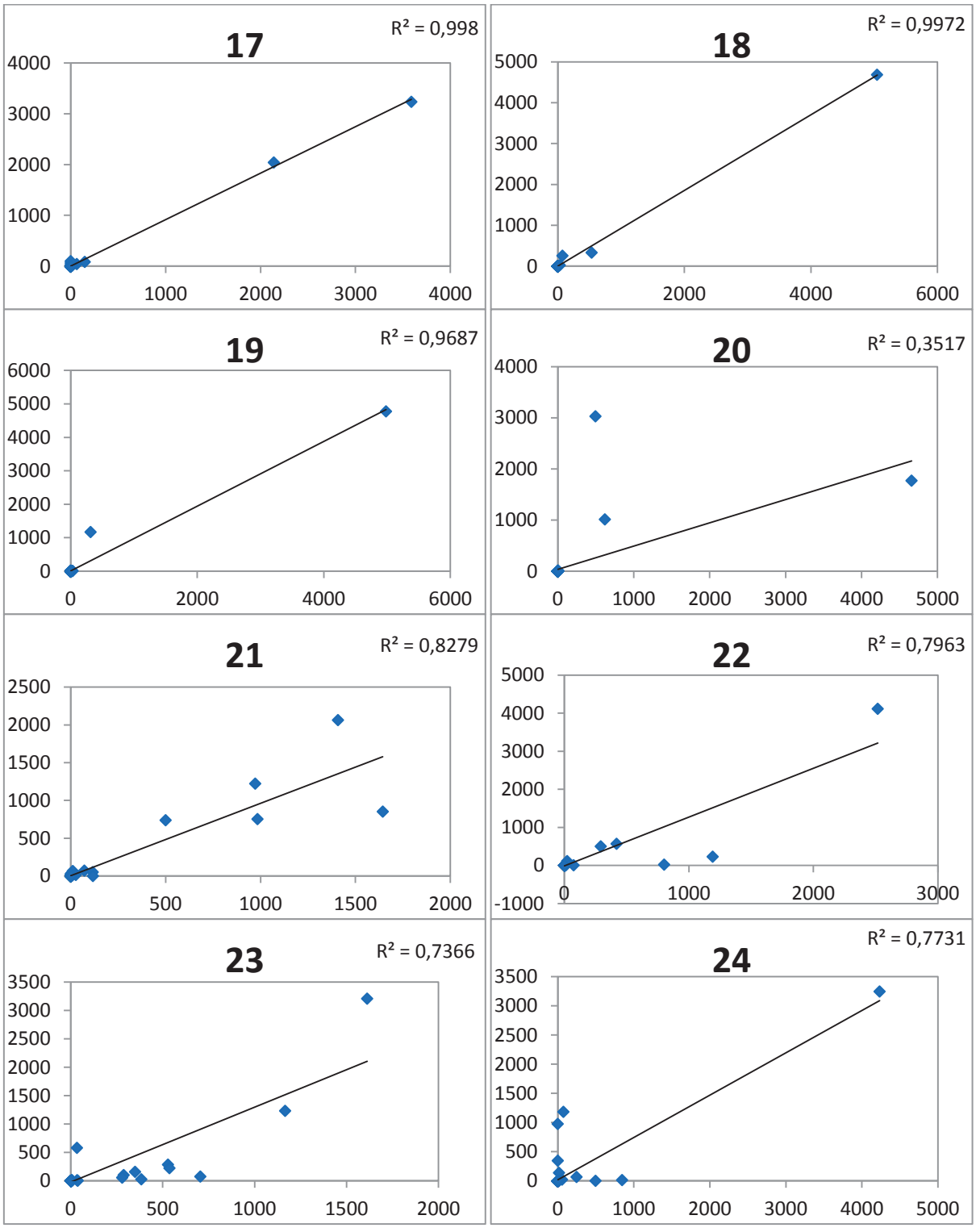


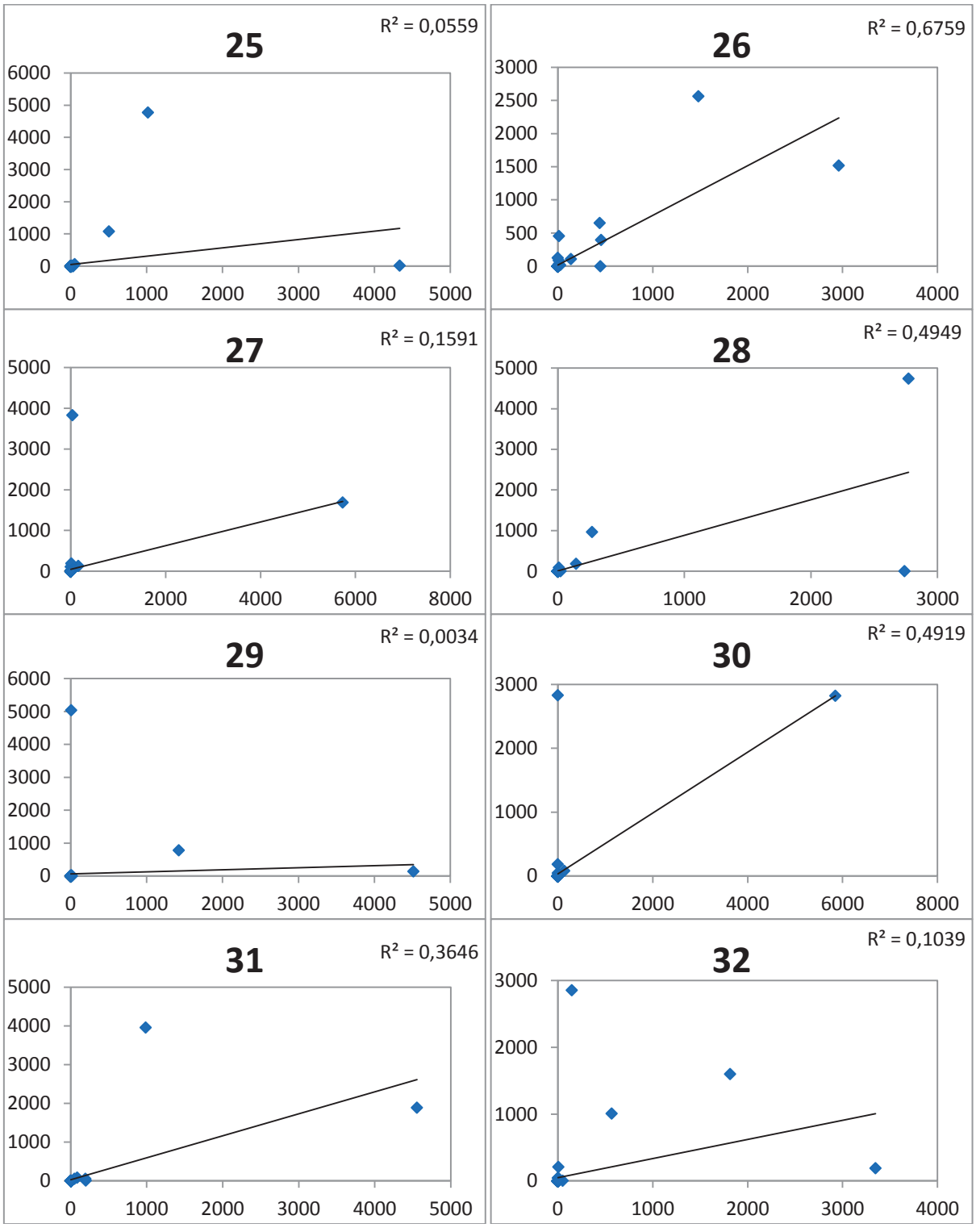
Figure A.7 Sequence alignment to the *E. coli* plasmid p1658/97. The upper picture shows alignment of the sequence covering the plasmid and the integron, while the bottom picture shows alignment of the sequence covering the integron and the transposon (Estensmo, this thesis).

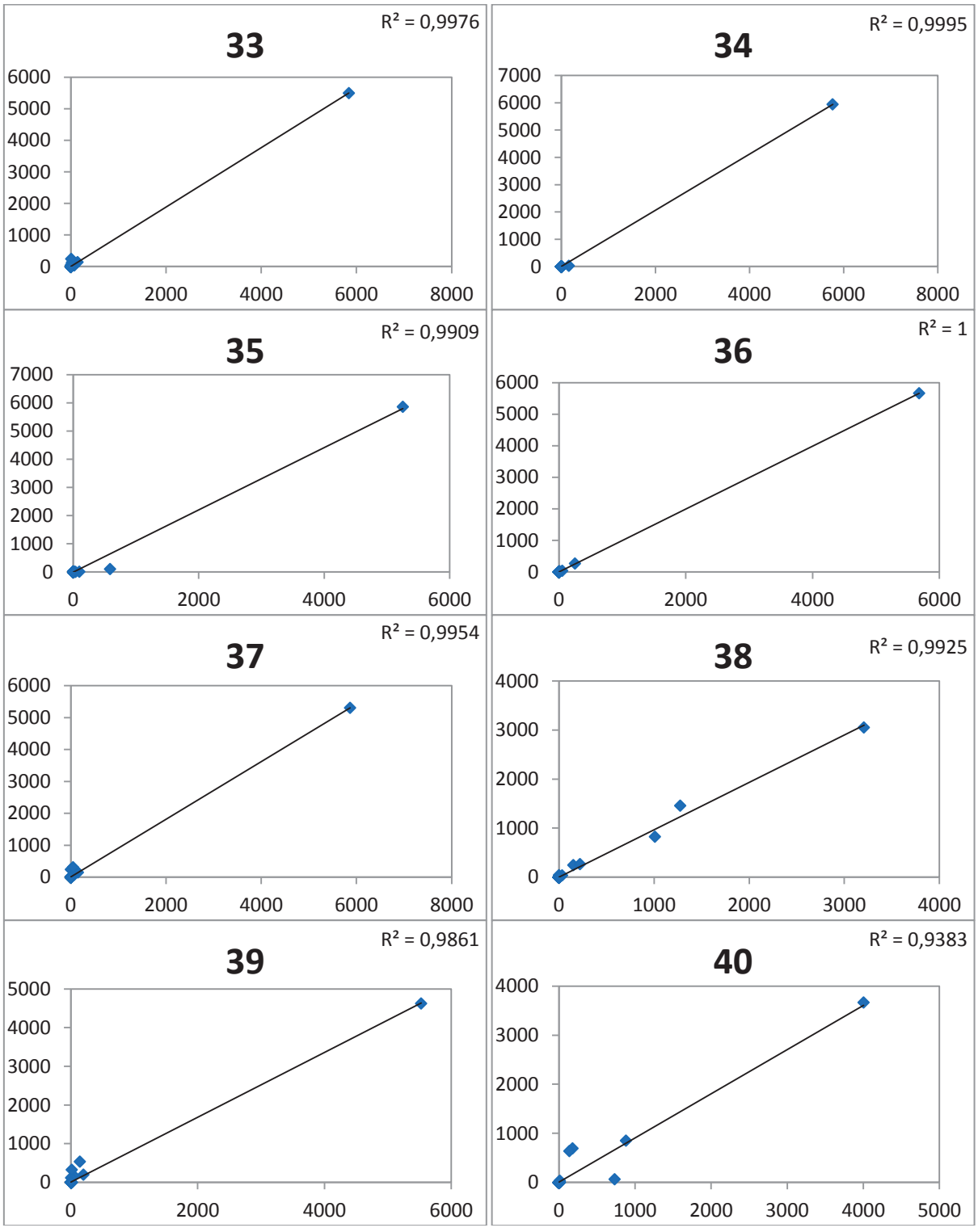
Appendix K: Regression Analysis of Duplicates




9
R² = 0,8078
10
R² = 0,5591
11
R² = 0,0471
12
R² = 0,9657
13
R² = 0,994
14
R² = 1
15
R² = 0,0154
16
R² = 0,9704


17
R² = 0,998
18
R² = 0,9972
19
R² = 0,9687
20
R² = 0,3517
21
R² = 0,8279
22
R² = 0,7963
23
R² = 0,7366
24
R² = 0,7731


25
R² = 0,0559
26
R² = 0,6759
27
R² = 0,1591
28
R² = 0,4949
29
R² = 0,0034
30
R² = 0,4919
31
R² = 0,3646
32
R² = 0,1039


33
R² = 0,9976
34
R² = 0,9995
35
R² = 0,9909
36
R² = 1
37
R² = 0,9954
38
R² = 0,9925
39
R² = 0,9861
40
R² = 0,9383

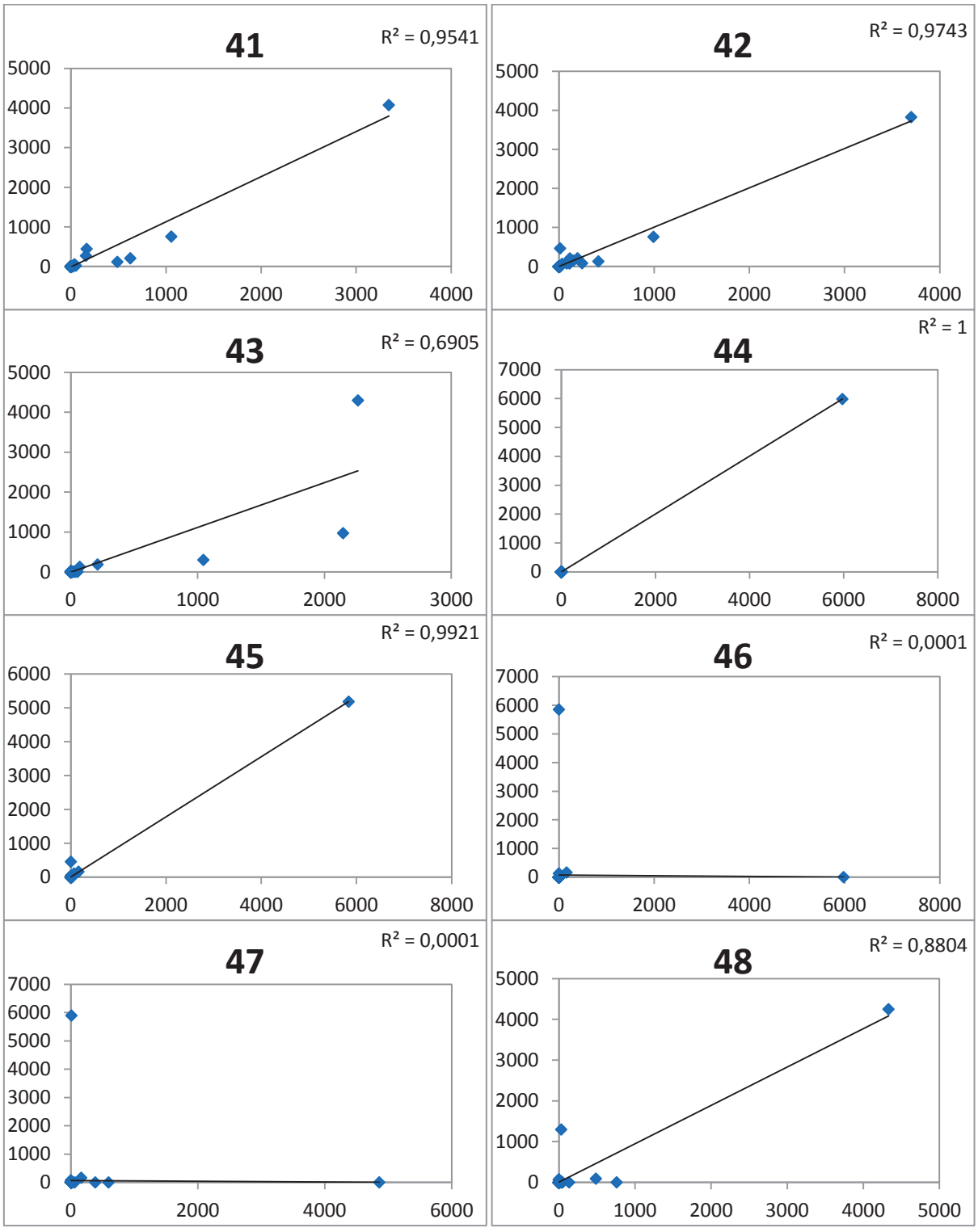
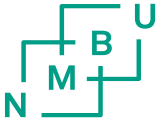


Figure A.8 Regression analysis of duplicate samples. Each figure represents a duplicate pair of samples isolated by different DNA extraction methods: GA, magnetic beads or Qiagen. The number of sequences belonging to the OTUs has been compared. Figure 1-19: GA vs. magnetic beads. Figure 20-48: GA vs. Qiagen.



Norwegian University
of Life Sciences

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