Development of the preterm gut microbiota in infants at risk of necrotising enterocolitis and sepsis

CHRISTOPHER JAMES STEWART

PhD

2014

Development of the preterm gut microbiota in infants at risk of necrotising enterocolitis and sepsis

CHRISTOPHER JAMES STEWART

A thesis submitted in partial fulfilment of the requirements of the University of Northumbria at Newcastle for the degree of Doctor of Philosophy

Research undertaken in the School of Life Sciences and in collaboration with the Royal Victoria Infirmary and Freeman Hospitals in Newcastle upon Tyne.

March 2014

In loving memory of

Irene Stewart (1928 – 2012)

Abstract

The gut microbiota comprises all the microorganisms colonising the gastrointestinal tract. It is a complex and dynamic community influenced by genetic and environmental factors. While the gut microbiota has crucial roles in micronutrient production and immunomodulation, it has also been associated with necrotising enterocolitis (NEC) and sepsis in preterm infants, which can exist exclusively or concurrently. As the number of babies born preterm continues to rise, so too will the incidence of these disease states. Exploring the development of the preterm gut microbiota longitudinally may offer important insights into the role of modern clinical practises in shaping the community and its subsequent role in disease pathogenesis.

To explore the development of the preterm gut microbiota we compared routine culture data with denaturing gradient gel electrophoresis (DGGE). Both techniques revealed differential profiles between patients with NEC and sepsis, compared to healthy controls. This was due, in part, to an increased abundance of *Staphylococcus* spp. identified in patients with NEC and sepsis. Based on these findings we explored the differential community development utilising a more extensive molecular approach, advancing on previous studies by exploring both the bacterial and fungal communities and also exploring the viability of each organism. For the fungal community, only non-viable fungal species were detected but showed no significant association with NEC or sepsis. Conversely, the viable bacterial community largely corresponded to that of the total community and showed *Sphingomonas* sp. was significantly associated with NEC. Interestingly, antifungal treatment had a significantly effect on the bacterial community and antibiotics limited the bacterial diversity which may have important consequences in the pathogenesis of disease.

We further analysed a twin cohort to investigate the role of host genetics in influencing the development of the gut microbiota and the subsequent risk of disease. Twins

showed comparable gut microbiota development with antibiotics attributable for major shifts in the community. A twin discordant for NEC showed a reduction in diversity and prevalence of an *Escherichia* sp. prior to the diagnosis which was not observed in the control twin. To further explore the discrepancies in the organisms associated with NEC and sepsis, overcoming the limitations of previous studies, we utilised next generation sequencing (NGS) in a large cohort with regular sampling pre and post disease diagnosis, matched to controls. Gestational age was shown to have important influences on the community development. No consistent associations between reduced diversity or increased dominance prior to disease diagnosis were observed, although *Escherichia coli* was prevalent prior to diagnosis of NEC. The organism identified in sepsis cases was present in the gut microbiota and was usually a dominant member. A diverse community seems to be important to the health of a neonate supporting the notion that a stable and diverse gut microbiota is important for preterm neonatal health.

Table of contents

Table of figures

Table of tables

Acknowledgements

First and foremost, I would like to thank Prof. Stephen Cummings for proving the opportunity and encouraging me to pursue the PhD. His belief and support throughout my time as his student has been invaluable and what I achieve in my career as a scientist will be due in no small part to the advice and inspiration he has bestowed in me.

The team at the Royal Victoria Infirmary have been fundamental to the project through their constant support, enthusiasm, and hard work. I cannot thank Dr Janet Berrington enough for all the hard work and support she has provided in getting this project off the ground and in the generation of manuscripts. Her enthusiasm for the research and passion for improving the lives of preterm neonates is inspiring. I owe sincere gratitude to Dr Nicholas Embleton whose input into both this project and applications for further funding have been invaluable. Thanks must also go to the nursing staff involved in processing samples and patient notes, particularly Eileen Turnbull who dealt with much of my correspondence. A sincere thank you must go to Tiny Lives, a fantastic charity I will contine to support. Thanks also to the parents and babies who contributed samples.

I would like to express my gratitude to the team at the Freeman hospital. I am grateful to Prof. John Perry for his comments in the generation of manuscripts and his advice on the clinical microbiology aspects of the project. I am especially thankful to Emma Marrs who generated the culture data in chapters 3 and 4, as well as processing and storage of samples.

I have been lucky in my time at Northumbria to work with some excellent scientists both past and present. I thank David Scribbins for assisting in the fungal work in chapter 5 and Giles Holt for running the samples on the MiSeq in chapter 7. Special mention must go to Dr Andrew Nelson who facilitated my progression into an independent researcher and remains a close friend. Rachael Dack has been there for me

during both undergraduate and PhD more than any other person and I could not have done it without her support.

Without the love and support from my family this thesis would not be possible. To my parents I feel I owe everything for always believing in me. Throughout my school years they gave me the confidence and at times kick up the backside I needed to be all I could be. I am thankful to my sister and fierce competitor for setting a great academic example. I dedicate this thesis to my nanna, Irene Stewart, who sadly passed away during my PhD. Knowing she was always proud of me was a great source of comfort during the more challenging times.

Lastly I need to thank my friends, of which there are too many to name. Suffice to say I would not have enjoyed the last 3 years as much as I have without them. They have kept me grounded when science was going well and picked me up when it was not. Those weekends in the pub were at times as valuable as the weekends in the lab. So for everyone who has believed in and supported me through my PhD, thank you.

Declaration

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. The work was done in collaboration with the Microbiology department at the Freeman hospital, Newcastle upon Tyne and the special care baby unit at the Royal Victoria Infirmary, Newcastle upon Tyne.

Any ethical clearance for the research presented in this thesis has been approved. Approval has been sought and granted by the NHS national research ethics service on 07/03/2011.

Name:

Signature:

Date:

Abbreviations

ANOVA	One-way analysis of variance
ATP	Adenosine triphosphate
bp	Base pair
BPD	Bronchopulmonary dysplasia
BSA	Bovine serum albumin
CCA	Canonical correspondence analysis
CCD	Charge coupled device
cDNA	Complementary DNA
CMV	Cytomegalovirus
CoNS	Coagulase negative *Staphylococcus*
CRP	C-reactive protein
DCA	Detrended correspondence analysis
DGGE	Denaturing gradient gel electrophoresis
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
DOL	Day of life
DPI	Dots per inch
E	Species evenness
EBM	Expressed breast milk
ELBW	Extremely low birth weight

EOS	Early onset sepsis
g	Gram(s)
GIT	Gastrointestinal tract
H'	Shannon diversity index
H_2O	Water
IBD	Inflammatory bowel disease
Ig	Immunoglobulin
IL	Interleukin
LB	Luria-Bertani medium
LOS	Late onset sepsis
LPS	Lipopolysaccharide
M	Molar
mA	Milliamps
MALDI-TOF	Matrix-assisted laser desorption/ionization time-of-flight
MAMPs	Microbial-associated molecular patterns
$MgCl_2$	Magnesium chloride
mRNA	Messenger RNA
MS	Mass spectrometry
NEC	Necrotising enterocolitis
NF-κB	Nuclear factor kappa-B
NGS	Next generation sequencing
NICU	Neonatal intensive care unit

NTC	No template control
OTU	Operational taxonomical unit
PAF	Platelet activating factor
PCA	Principal component analysis
PCoA	Principal coordinate analysis
PCR	Polymerase chain reaction
PCT	Procalcitonin
PLS-DA	Partial least squares discriminant analysis
PRR	Pattern recognition receptor
QIIME	Quantitative insights into microbial ecology
qPCR	Quantitative PCR
R	Species richness
RDA	redundancy discriminate analysis
Rf	Retention factor
RNA	Ribonucleic acid
ROP	Retinopathy of prematurity
rRNA	Ribosomal RNA
RT-PCR	Reverse transcription polymerase chain reaction
SAA	Serum amyloid A
SOP	Standard operating procedure
spp.	Species
TAE	tris-acetate-ethylenediaminetetraacetic acid

TEMED	N,N,N',N'-Tetramethylethylenediamine
TGGE	Temperature gradient gel electrophoresis
TIFF	Tagged image file format
TLR	Toll-like receptor
TNF	Tumor-necrosis factor
TTTS	Twin to twin transfusion syndrome
UV	Ultraviolet
V	Volts
v/v	Volume per volume
VLBW	Very low birth weight
w/w	Weight per weight
QSM	Quorum sensing molecule

1. Introduction

Microorganisms have long been recognised as fundamental to the cause and prevention of human disease, as demonstrated by the early work of Pasteur, Lister, and Koch. Indeed, Louis Paster is quoted as saying:

"The role of the infinitely small in nature is infinitely great"

This is particularly true of the microbial communities present in the gastrointestinal tract (GIT) of mammals, termed the gut microbiota, which has received significant interest over the last decade. There is an increasing understanding of the role of the gut microbiota in maintaining health through immunomodulation, protection, nutrition and metabolism and in contributing to disease through inflammation, diabetes, autism, obesity and allergy (Scanlan *et al.* 2006; Neish 2009; Sekirov *et al.* 2010). As a consequence, many studies have looked at the development of the gut microbiota at all stages of life from birth to old age and related how the community structure changes due to demographic (patient) variables (Echarri et al. 2011; Jalanka-Tuovinen et al. 2011; Arboleya et al. 2012; Claesson et al. 2012). While the gut microbiota may contain a variety of microorganisms, including species of fungi, archaea, and virus, in this thesis the focus is on assessing the bacterial community.

The human GIT represents a highly complex ecosystem. At birth, an infant's GIT is regarded as sterile but rapidly becomes colonised with bacteria derived from the maternal and environmental flora. The initial development of the gut microbiota is critical for the development and maturation of the adaptive immune system, protection against pathogens, and metabolism of otherwise indigestible nutrients (Sekirov *et al.* 2010). Modelling the gut microbiota in *term* infants suggests that competition between three phyla: Bacteroidetes, Proteobacteria and Firmicutes, explain most community

dynamics (Palmer *et al.* 2007; Trosvik *et al.* 2010). However, studies typically assess the total community and thus shifts within the viable, that is the living or metaboloically active portion of the gut microbiota remain undetected. While analysis of the total gut microbiota provides a phylogenetic picture of the community, it does not reflect viability as the deoxyribonucleic acid (DNA) extracted and sequenced could originate from metabolically active, dormant, lysed, or dead cells (Tannock *et al.* 2004). The viable gut microbiota has been shown to differ to the total community. Statistical analysis of adult populations indicated a decrease in viable Bacteroidetes and an increase in Firmicutes, which was masked if only the diversity of the total community was considered (Peris-Bondia *et al.* 2011). There is no comparable study exploring the viability of the gut microbiota in the neonate, however, these observations suggest it is important to distinguish how the viable and total communities differ in the early stages of life.

During the first year of life the infants gut microbiota is assembled and undergoes a series of significant changes associated with life events such as feeding (with formula or breast milk), the adoption of solid foods, exposure to the home environment, and antibiotic treatment for infection (Koenig *et al.* 2010; Morowitz *et al.* 2011). After the first year of life the gut microbiota profile will resemble that of the adult for composition, although it will still be unique to the individual (Adlerberth & Wold 2009). In comparison, the adult gut microbiota has much greater temporal stability and is estimated to contain between 400 and 1000 bacterial species, primarily to the phyla Bacteroidetes and Firmicutes (Peris-Bondia *et al.* 2011). The shift from neonate to adult gut microbiota is perhaps defined most by the increased abundance of the genus *Bacteroides* and a substantial reduction in the Proteobacteria. The Bacteroidetes are specialised in the breakdown of complex plant polysaccharides, which when introduced

into the diet at weaning probably explains the increased contribution of Bacteroidetes to the gut microbiota (Sellitto *et al.* 2012).

In adults the gut microbiota influences a diverse range of health outcomes from obesity, diabetes, asthma and allergy to seemingly 'remote' diseases like Parkinson's disease (Turnbaugh *et al.* 2007). In preterm infants, establishment of the gut microbiota is also of importance for key morbidities like sepsis (specifically late onset sepsis) and necrotising enterocolitis (NEC), both significant causes of mortality (Berrington *et al.* 2012). Many episodes of sepsis are with gut derived organisms and changes in the intestinal barrier contribute to both NEC and sepsis (Vergnano *et al.* 2011). The gut microbiota is key to developing barrier function, integrity, and mucosal and systemic immune function. It also 'educates' the gut associated lymphoid tissue, allowing the establishment of a 'tolerant' state between microbiota and the immune system, affecting intestinal function including tight junction structure and immune function (Rakoff-Nahoum & Paglino 2004; Bäckhed 2011; Bevins & Salzman 2011). Furthermore, patterns of initial colonisation affect host metabolic function: fat deposition, circulating leptin levels, and insulin resistance (Bäckhed 2011).

Due to the complexity and variability of community development in a niche like the GIT, where environmental, biological and genetic backgrounds are significantly different and in some cases are constantly changing between and within subjects, studies that have attempted to control for such variation where possible are most informative. One such variable relates to the host genetic predisposition. Twin studies, therefore, offer important insights into the significance of the host genetic background in affecting GIT microbiota development. Healthy twins have been shown to develop a comparable gut microbiota when compared following term birth (Palmer *et al.* 2007), in childhood (Stewart *et al.* 2005) and in adulthood (Zoetendal *et al.* 2001), suggesting genetic factors shape the gut microbiota. However, due to a lack of similarity with

siblings from a different birth and the high similarity observed in both monozygotic and dizygotic twins, it is conceivable that the environment is more influential in shaping the gut microbiota (Palmer *et al.* 2007).

In the preterm gut, structural and immunological immaturity contributes to inflammatory necrosis and abnormal bacterial colonisation, termed dysbiosis. This may result in a limited microbial diversity and an increased inflammatory response exacerbated by an immature innate immune response that increases the risk of diseases like NEC or sepsis. An improved understanding of the microbiota of infants cared for in neonatal intensive care units (NICUs) and how this is affected by current practices may allow clinicians to promote more 'healthy' gut microbiota patterns, thus resulting in reductions in mortality and improvements in long term outcomes (Caicedo *et al.* 2005). The microbiota of a full term vaginally delivered neonate, who receives its own mother's breast milk and remains healthy, is regarded as the ideal 'gold standard'. However, the optimum microbial colonisation process can be disrupted by premature birth, mode of delivery, diet, and antibiotic and antifungal administration.

1.1 Factors influencing neonatal gut microbiota development

1.1.2 Premature birth

The development of the gut microbiota differs between premature and full term neonates. In general, neonates harbour a very simple community at birth that increases in diversity over time. In full-term infants the pioneering bacteria detected are taxa belonging to enterococci, streptococci, and *Enterobacteriaceae* (Favier *et al.* 2002). The bacterial community is initially very dynamic and studies that have modelled the gut

microbiota in full-term infants have shown that competition between three phyla, Bacteroides, Proteobacteria and Firmicutes, exerts the greatest effects on the community dynamics (Trosvik *et al.* 2010).

In contrast, preterm infants are cared for in hygienically controlled intensive care units and have a relatively simple gut microbiota compared to term babies (Berrington *et al.* 2013). This is likely due to a combination of several environment factors relating to the care of preterm infants including sterile practise, housing in incubators, increased antibiotic and antifungal administration, and enteral feeding by catheter (no skin contact). Like full-term infants, the pioneering bacteria in preterm infants commonly include members of enterococci, streptococci, and *Enterobacteriaceae* (Cilieborg *et al.* 2012). However, preterm infants show subsequent delayed colonisation with potentially important 'beneficial bacteria' such as bifidobacteria and lactobacilli. In healthy breast fed term infants bifidobacteria dominate by day seven, but not in preterm infants (Butel *et al.* 2007). It is postulated that the degree of gut maturation at birth may be important to colonisation with these organisms (MacDonald & Baker 1998; Feja *et al.* 2005).

1.1.2 Mode of delivery

Vaginally delivered infants initially harbour bacterial communities which resemble the vaginal microbiota. Thus, for infants whose birth involves passage through the birth canal, the dominant bacteria are initially composed of the genera *Lactobacillus*, *Prevotella*, *Atopobium*, and *Sneathia* (Dominguez-Bello *et al.* 2010). Moreover, the infant gut microbiota is more similar to its own mothers vaginal microbiota than that of non-related mothers. Higher levels of Bifidobacteria and *Bacteroides*, with less *Clostridium* (importantly *Clostridium difficile*), have been reported in vaginally delivered infants (Penders *et al.* 2006). Caesarean delivery is more common in preterm

infants and has been shown to result in delayed colonisation with a gut initially dominated by environmental bacteria, specifically a high prevalence of *Clostridium*, *Escherichia*, *Streptococcus* and *Staphylococcus* (Thompson-Chagoyán *et al.* 2007). Interestingly, infants born by caesarean section initially harbour a gut microbiota reflective of adult skin communities; but noteworthy is a lack of distinct similarity with the respective maternal skin microbiota (Dominguez-Bello *et al.* 2010).

1.1.3 Feeding

Breast feeding has been associated with an increased abundance of Lactobacilli and Bifidobacteria, with a reduction in *Clostridium* spp. compared to formula feeding (Penders *et al.* 2005). Formula fed infants typically show greater diversity with increased levels of facultative and obligate anaerobes, particularly *Bacteroides*, *Clostridium*, and *Enterococcus* compared to breast fed infants. Feeding very preterm infants with artificial formula rather than expressed maternal breast milk (EBM) increases the risk of NEC and sepsis. Interestingly, after weaning (introduction of solid food), the gut microbiota of breast fed infants changes to reflect that of formula fed infants. This occurs as a result of a significant increase in the abundance of the genera *Enterococcus* and *Enterobacter*, and the appearance of *Bacteroides*, *Clostridium*, and other anaerobic Streptococci (Adlerberth & Wold 2009).

1.1.4 Antibiotics and antifungals

To reduce fungal infections, preterm neonates may undergo a short course of antifungal prophylaxis, usually with nystatin or fluconazole (Manzoni *et al.* 2011). Fluconazole remains the most reported and useful antifungal for use in prophylaxis, although the

direct contribution of antifungal prophylaxis in the development of the neonatal gut microbiota remains elusive. Fluconazole prophylaxis has been shown to be effective in the prevention of *Candida* colonisation and infection in neonates. Although little work exists on the antibacterial activity of fluconazole, as a member of the imidazoles it is feasible that the antifungal will also have bacteriostatic and bactericidal properties, with particular action against Gram positive bacteria (Samuelson 1999). Conversely, it is known that frequent use of broad-spectrum antibiotics, particularly third generation cephalosporins, increase the risk of colonisation and infection with multidrug-resistant bacteria and fungi (Kaufman & Fairchild 2004; Feja *et al.* 2005; Chapman 2007).

Concerns of antibiotic and antifungal treatment related to the gut microbiota include the spread of antibiotic resistance among pathogens and that alteration of the microbiota will interfere with human–microbe interactions that are fundamental to human development. Antibiotics have been demonstrated to play a significant and long term role in altering the bacterial composition within the gut microbiota (Dethlefsen & Relman 2011; Pérez-Cobas *et al.* 2012). Antibiotics are commonly prescribed to neonates, particularly preterm infants where standard practise is to administer antibiotics for 48 hours unless proven infection. This is likely to result in the delayed development of a diverse gut microbiota in preterm infants (Cotton 2009). Understanding the impact of antibiotic administration on the developing gut microbiome is extremely challenging. The type of antibiotics, the combination used, their dosage and length of time of administration vary hugely between individual infants reflecting the patients' needs and the preference and experience of the clinicians treating the individual. These variables mean that monitoring the exact effects on the gut microbiota *in vivo* is extremely difficult. This is particularly true for neonates where robust sampling, which will be key to tracking the response of the gut microbiota to each antibiotic, is problematic. Thus, the majority of studies to date lack the statistical

power to convincingly guide clinical practise. Nonetheless, when the role of ceftriaxone was studied in term breast fed infants a decreased count of total bacteria, particularly *Enterobacteriaceae*, enterococci and lactobacilli, was observed (Savino *et al.* 2011). Ceftriaxone was also shown to cause a disappearance of *Bifidobacterium* spp. with a preservation of potentially pathogenic *Streptococcus* spp. and *Staphylococcus* spp. Frequent use of broad-spectrum antibiotics, particularly third generation cephalosporins, increase the risk of colonisation and infection with multidrug-resistant bacteria and fungi. These results highlight the need for greater research into the role of different types of antibiotics on the gut microbiota. Briefly, in adults, the effect of ciprofloxacin on the gut microbiota was profound and rapid, with a loss of diversity and a shift in community composition after 3 days. However, 1 week following treatment communities began to return to their initial state, but importantly the return was often incomplete and varied between patients (Dethlefsen & Relman 2011).

1.2 Techniques to explore the microbiota

Pioneering studies exploring the vast ecosystem of the gut microbiota relied on culture based approaches. However, the advent of molecular approaches has revolutionised microbiomic studies by enabling the contribution and identification of uncultivable organisms to be explored. The increasing reliance on sequence data to generate data on microbial composition presents new challenges. One of which is the adherence to Linnean classification (seven hierarchical taxa from kingdom to species) which relied on pragmatic definitions of species, integrating phenotypic, biochemical and phylogenetic data. However, for prokaryotes there is no accepted conceptual definition for a species (Sutcliffe *et al.* 2012). Table 1.1 summarises common current molecular

methodologies and standard culture approaches. These approaches and the importance of sampling are discussed in detail in subsequent sections.

Table 1.1 Comparison of techniques

Technique	Standard culture	Fingerprinting methods	qPCR	Next Generation Sequencing
Example	Culture on blood MacConkeys agar for Gram-negative species	Temperature or Denaturing Gel electrophoresis (T/DGGE)	SYBR Green I florescent dye; *Taq*Man probes	454 pyrosequencing (Roche); sequence by synthesis (Illumina)
Brief description	Viable bacteria grown in a laboratory on specific media with specific conditions; physical/chemical characteristics used to speciate	DNA extracted and amplified. Separation by sequence differences within amplicons using temperature or chemical gradient	DNA extracted and amplified with the number of copies of the target gene shown real-time	DNA extracted and sequenced; utilises extensive databases of 16S gene to identify species
Advantages	Lots of experience in technique; cheap; limited equipment needed; target key organisms	Greater depth of analysis than culture; relatively cheap (for molecular technique)	Accurately quantifies the presence of a gene in a sample; quick	Very detailed information; complex communities can be examined; quick
Disadvantages	~90% species remain undetected; need to already know species of interest	Time consuming; PCR bias; sequence identification is time consuming and not robust	Can be costly; PCR bias; requires highly skilled technician; data must conform to very stringent criterion	Costly; PCR bias; specialised equipment needed, available in few localities; enormous amounts of data generated require storing and specialist analysis

1.2.1 Sampling strategy

Studies exploring the gut microbiota typically utilise stool samples as a convenient non-invasive means of exploring its diversity. However, the gut microbiota is not uniformly distributed throughout the GIT due to the changing physicochemical conditions, exerting selective pressures on the community. These physicochemical conditions include intestinal motility, pH, redox potential, nutrient supplies, water content, and host secretions such as hydrochloric acid, digestive enzymes, bile and mucus (Booijink *et al.* 2007). While it is well known the gut microbiota is not homogeneous throughout the GIT, the exact differences in diversity have not been definitively determined (Gerritsen *et al.* 2011). This has implications when exploring disease states using stool samples, for example NEC, where the origin of inflammation is known to be the terminal ileum and proximal colon. Eckburg *et al.* (2005) postulated that the stool microbiota represented a combination of shed mucosal bacteria and a separate nonadherent luminal population. Further work assessing the biostructure of microbiota in adult stool showed a clear structure from the outside of stool (closest to the mucosa) toward the centre (luminal bacteria), which was distorted in patients with idiopathic diarrhoea (Swidsinski *et al.* 2008). However, microbiota structure did not differ when comparing the front and end of a stool pellet, thus a section of a stool pellet is an accurate representative of the whole stool (Mai *et al.* 2010). In contrast, there is only limited insight in the composition of the microbiota that resides in the small intestine, particularly the ileum, which is difficult to sample (Booijink *et al.* 2007). Due to the relatively large amounts of gastric acid, bile and pancreatic secretions in this region creating a harsh niche for microbial growth, it is feasible that microbial diversity will be low compared to other regions. Indeed, ileal effluent from ileostomy patients was shown to clearly differ from the stool microbiota with lower diversity and stability and a higher relative abundance of potentially pathogenic species (Booijink *et al.* 2010). Due

to the ease of utilising stool samples in clinical research, the majority of publications referenced in this thesis are based on results from stool samples. Future studies will continue to utilise stool when analysing the gut microbiota, but more detail on how the stool microbiota differs from the distinct niches throughout the GIT is warranted.

Storage of samples is an important factor in ensuring consistency and scientific accuracy in many biological experiments. Studies exploring the effect of storage, for example in the postage of sputum samples for cystic fibrosis (CF) analysis, have shown that microbial profiles can be significantly altered with an overall reduction in diversity when left at room temperature for 24 hours compared to immediate cold storage. Importantly in the case of CF, an increase in the abundance of *Pseudomonas* spp. and a reduction in *Haemophilus influenzae* were observed in samples left at room temperature (mimicking postage), leading to a misrepresentation of the original and true profiles (Nelson *et al.* 2010). Freeze-thaw cycles are known to lyse cells and so it especially important for viable work, such as culture based experiments, to avoid this and where possible prepare the culture from freshly collected material (Sharma *et al.* 2006). However, it has been shown that storage of stool samples for up to two weeks at room temperature does not significantly alter the observed bacterial community in molecular based approaches when exploring total community using extracted DNA (Lauber *et al.* 2010). Nonetheless, preservation at -80 ºC and avoidance of repeated freeze-thaw cycles is recommended to prevent loss of diversity.

1.2.2 Culture based techniques

Initial studies exploring gut microbiota in neonates relied on the ability to cultivate different species and carry out subsequent colony counts and biochemical methods to identify organisms (Blakey *et al.* 1982; Stark & Lee 1982; Westra-Meijer *et al.* 1983). It

has since been estimated that only 20% of the gut microbiota is cultivable (Eckburg *et al.* 2005). Successful culture of microorganisms allows their phenotypic and metabolic capacities and behaviour in co-culture to be analysed. Importantly for the GIT, which consists of many distinct niches, co-culture studies have shown that many ecological niches can only be filled by syntrophic consortia of different microbes. For instance, co-culture of the xylanolytic *Roseburia intestinalis* with H_2-utilizing *Ruminococcus hydrogenotrophicus* dramatically increased butyrate production (Chassard & Bernalier-Donadille 2006).

Although culture dependent approaches are regarded by many as inadequate for exploring the full diversity of the gut microbiota, there have been efforts to optimise this technique with the implementation of selective media and anaerobic incubation. Advances in mass spectrometry (MS) techniques coupled with matrix-assisted laser desorption ionisation time-of-flight (MALDI-TOF) have also improved the accuracy and time needed to identify large numbers of colonies (Seng *et al.* 2009). Furthermore, there is evidence to suggest that culture may still offer important insights. For example, a patient with low bacterial load had more bacterial species identified by culture dependent approaches compared to next generation sequencing (NGS) 454 pyrosequencing (Dubourg *et al.* 2013).

As greater knowledge is gained on the composition of niches by molecular techniques, it may directly improve the ability to cultivate organisms previously deemed uncultivable. More insight into what other organisms are present, particularly relating to the metabolites they produce as well as other available nutrients, might allow the simulation of specific growth conditions leading to successful isolation. What is important in the modern era, when researchers are increasingly realising the power of molecular techniques, is that the power of culture based techniques are not forgotten.

1.2.3 Culture independent techniques

Since the 1990s, molecular techniques have been increasingly applied to clinical investigations as they offer much greater coverage of the microbial community compared to culture dependant approaches (Petrosino *et al.* 2009). Molecular techniques allow the gut microbiota to be characterised and monitored without prior knowledge of its structure or composition. The basis of molecular techniques is the extraction of nucleic acid (DNA and/or RNA) from samples (Nechvatal 2008).

Both DNA and ribonucleic acid (RNA) can be used to analyse microbial communities in stool by molecular approaches. DNA analysis is more common as a result of the added difficultly and potential for degradation when working with RNA. RNA is unstable at room temperature and thus requires stringent sample handling procedures to ensure the sample is transported to the laboratory without degradation or a shift in microbial profiles. To prevent this degradation during transit it is paramount the sample is stored in an RNA stabilisation reagent (Hernandez *et al.* 2009). Carrying out subsequent analysis on extracted RNA allows the characterisation of the metabolically active community. This is of particular importance when aetiology is explored as the causative agent is most likely to be metabolically active. It is feasible that an organism is present in a dormant state and so not metabolically active, but when environmental conditions shift they can become active and play a part in the community function (Prosser *et al.* 2007). Similarly, to achieve full coverage of potential viral communities then both DNA and RNA need to be examined. Thus, in order to gain the most comprehensive insight into the pathology of diseases like NEC, analysis of both DNA and RNA is warranted.

After nucleic acid extraction from biological samples, subsequent diversity studies require amplification by polymerase chain reaction (PCR), typically based on the 16S

ribosomal RNA (rRNA) gene. These amplified regions (amplicons) are then differentiated into groups that share a pre-defined similarity to each other known as operational taxanomic units (OTUs). The 16S rRNA gene is a component of the small subunit in the prokaryotic ribosome. This ribosomal gene is functionally conserved and it is ubiquitously expressed in bacteria. It represents a rational target for comparative studies between bacterial communities and is also the most widely used gene for bacterial classification and identification (Isenbarger *et al.* 2008). The gene contains nine 'hypervariable regions' showing considerable sequence diversity, flanked by 'conserved regions' of homologous sequence (Chakravorty *et al.* 2007) (Figure 1.1). This allows universal PCR primers, complimentary to conserved regions, to be designed facilitating the amplification of the intervening hypervariable regions. Bacterial classification can then be determined by comparing the amplicons to homologous 16S rRNA sequences from characterised bacteria within open access sequence databases.

The copy number of the 16S rRNA gene varies between species from 1 to 15 (Klappenbach *et al.* 2001) with the number of nucleotides that are different between any pair ranging from 0 to 19 (Coenye & Vandamme 2003). While copy numbers are generally species specific, there is reported variation among strains of the same species (Acinas & Marcelino 2004). This is believed to be a life strategy by bacteria owing the rRNA copy number correlating with the ability to respond to favourable growth conditions (Klappenbach *et al.* 2000). Noteworthy is that multiple copies of the same target gene in PCR analysis is not limited to the bacterial kingdom. The linking of cell abundance and PCR amplicons abundance is also limited by the multiple copy nature and intragenomic variability of the common molecular marker of fungi; the ITS region of the rRNA gene from the 28S subunit (Amend *et al.* 2010). The specific limitations conferred by both heterogeneous and homogeneous copies of the target gene in PCR are discussed in more detail in the subsequent sections.

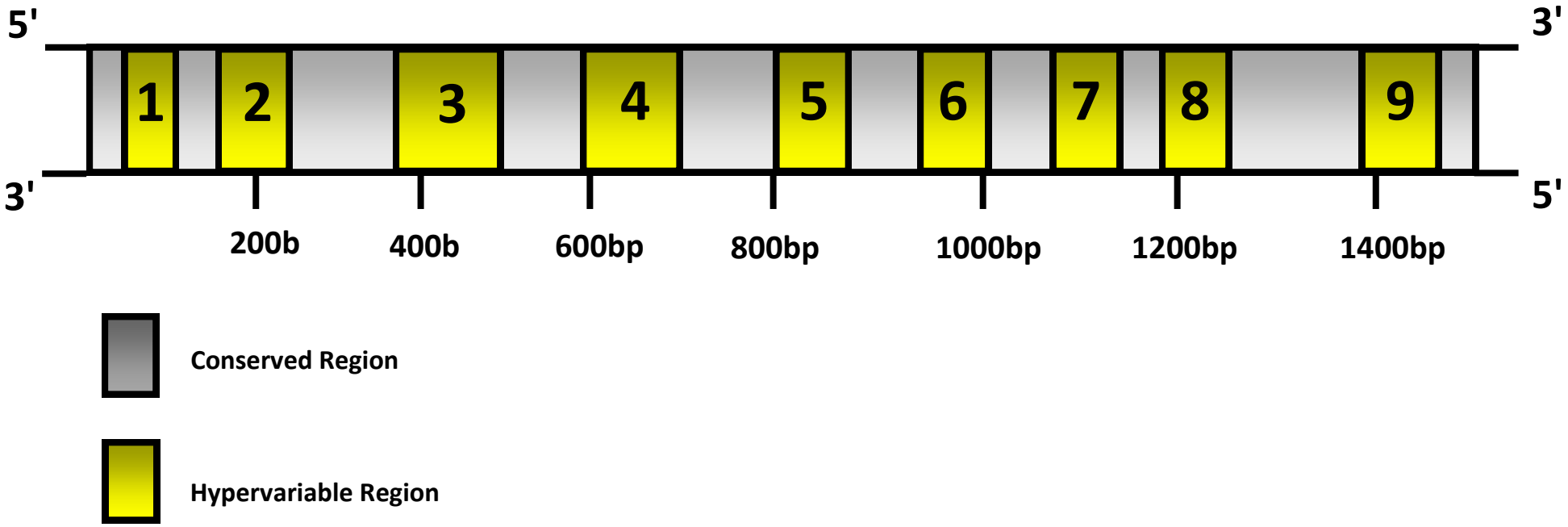


Figure 1.1 – Schematic representation demonstrating the 16S ribosomal RNA gene used in molecular studies.

1.2.4.1 Polymerase Chain Reaction (PCR)

PCR was originally described by Mullis *et al.* (1986) and has become one of the most widely used techniques in molecular biology. It is fundamental for the majority of molecular approaches including molecular fingerprinting techniques and NGS. PCR involves adding template DNA to a reaction containing primers (synthetic oligonucleotides complementary to a part of the target sequence), each deoxynucleoside triphosphate (dNTPs; dATP, dCTP, dGTP and dTTP), polymerase (enzyme responsible for incorporating the dNTPs into the complementary sequence), and cofactors needed by the polymerase such as buffer containing stabilising salts and magnesium chloride ($MgCl_2$). Polymerase, including the frequently used *Taq* polymerase, is derived from thermophilic microbes such as the *Thermus aquaticus.* The reaction is principally carried out at 3 temperature ranges in subsequent cycles of 1) high temperature for denaturation of the DNA, 2) low temperature based on the annealing temperature of the primers for annealing of the primers, and 3) medium temperature for extension of the complementary sequence by addition of free dNTPs. The number of amplicons, which is the sequence generated based on the fragment of sequence between the forward and reverse primers, increases exponentially after each cycle of the reaction (Mullis *et al.* 1986).

Like all techniques, PCR is not without limitations and bias. The primers used in PCR reactions are inherently biased toward certain OTUs. Bias is introduced by universal primers through differential annealing in the amplification of a heterogeneous template, affecting the amplification efficiency of some OTUs (Wang & Qian 2009). Ideally, universal primers will amplify all bacteria in a mixed population but this is impossible. For instance, the longest number of conserved nucleotides in the 16S rRNA gene that are 100% conserved is 11 (Baker *et al.* 2003). Primers which match 95% of sequences in the Ribosomal Database Project (RDP) from typically dominant gut organisms have

been found to miss specific OTUs; the reverse primer at position 1492 (*Escherichia coli* 16S rDNA position) detects only 61% of Actinobacteria and 54% of Proteobacteria (Hamady & Knight 2009). To overcome this bias, sites of primer mismatch can be identified and a degenerate base can be incorporated into the primer sequence. By addition of a degenerate base pair to the -357F/926R primer set (Muyzer *et al.* 1993), Sim *et al.* (2012) were able to increase the bifidobacteria detection rate in stool samples, without diminishing the amplification of other OTUs. The potential for improved detection of bifidobacteria has important consequences in the analysis of neonatal samples where bifidobacteria is regarded as an important species for health, but is often found to only constitute a minor component of the gut microbiota (Picard *et al.* 2005; Palmer *et al.* 2007). Artifacts may also arise due to the formation of heteroduplex and chimeric sequences. A heteroduplex is formed in PCR by the cross-hybridization of heterologous sequences, giving rise to sequences for non-existent genes (Thompson *et al.* 2002). While it is known to occur during mixed template PCR using universal primers, it can be decreased by the addition of *Taq* polymerase and reducing the number of PCR cycles (Michu *et al.* 2010). A chimera on the other hand is formed by the incomplete extension of a primer and template switching which also gives rise to artificial gene diversity (Shuldiner *et al.* 1989; Patel *et al.* 1996).

1.2.4.2 Quantitative PCR (qPCR)

Quantitative PCR (qPCR) is another PCR based technique which enables the quantification of the target gene in real-time. qPCR is a robust, highly reproducible and sensitive method which can be used to quantitatively track phylogenetic and functional gene changes (Smith & Osborn 2009). In microbial ecology, this technique can be used to accurately quantify bacteria, typically at the family or genus levels (Palmer *et al.*

2007; Bucher *et al.* 2011). The technique relies on the detection of fluorescence, where the signal intensity is relative to the number of amplicons generated on completion of each cycle in the PCR. The concentration of target DNA in the sample is then calculated based on the exponential phase of the PCR when reaction components are in abundance by calculating the C_t (cycle threshold) value (Smith & Osborn 2009).

Two different fluorescent based systems are commonly used for qPCR; SYBR Green I which is based on fluorescent chemistry and *Taq*Man technology which is based on a reporter-quencher system. Firstly, owing to its affordability, SYBR Green I is the most commonly used dye for non-specific detection. SYBR Green I is a double-stranded DNA (dsDNA) intercalating dye which emits fluorescence when bound to the DNA (Valasek & Repa 2005). Thus, as the target is amplified the dye will bind to the amplicons and the amount of fluorescence will be proportional to the amount of target gene generated. As this is non-specific binding, the SYBR Green I will bind to any dsDNA in the reaction including primer dimers, heteroduplex formations, and chimera sequences which will introduce bias into the quantification (Gibson 2006). For this reason it is important that the qPCR has been optimised, which can be confirmed by a well-defined peak on the melt curve (Smith & Osborn 2009). The second commonly used system is the *Taq*Man probe. Here, a fluorophore is attached to the 5′ end of the probe and a quencher to the 3′ end. The technique relies on FRET (Fluorescence Resonance Energy Transfer) from the fluorophore (high energy) and the quencher (low energy) (Gibson 2006). The probe is designed to anneal at a sequence downstream of one of the PCR primers and binds to the amplicon during each annealing step in the PCR. While the fluorophore and the quencher are in close proximity, the fluorescence is quenched and not detectable. However, as the new strand is synthesised from the primer by *Taq* polymerase, the 5′ exonuclease activity of the enzyme cleaves the labelled 5′ nucleotide of the probe, releasing the reporter from the probe (Smith & Osborn 2009).

As cleavage continues the remaining probe melts off the amplicon releasing the fluorophore and quencher into solution, spatially separating them, leading to an increase in the detection of fluorescence (Valasek & Repa 2005).

In an attempt to overcome the lack of sufficient experimental detail in publications reporting qPCR analysis, the minimum information for publication of quantitative PCR experiment (MIQE) guidelines have been proposed (Bustin *et al.* 2009). The idea of the guidelines is "to help ensure the integrity of the scientific literature, promote consistency between laboratories, and increase experimental transparency" (Bustin *et al.* 2009).

1.2.4.3 Molecular fingerprinting

First generation molecular fingerprinting techniques include denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE). These techniques are only capable of separating short amplicons, not exceeding 500 bp (Temmerman & Masco 2003). For this reason the variable 3 (V3) region within the 16S rRNA gene (Figure 1.1) is ideally suited to this analysis with primers targeting the conserved regions at positions 341 to 518 (*E. coli* 16S rDNA position) (Muyzer *et al.* 1993). Molecular fingerprinting approaches exploit the different number of intermolecular hydrogen bonds between G-C and A-T base pairs. G-C base pairs have three hydrogen bonds and A-T have two, thus G-C base pairs require greater denaturant concentration to dissociate the extra hydrogen bond. Therefore, following electrophoresis, fragments with a higher G-C content will travel further through a denaturing gel, separating amplicons on the basis of their base pair content (Muyzer *et al.* 1993). The gels can be stained and observed under UV to visualise the position of

the amplicons, which appear as bands. DGGE is a semi-quantitative technique so the intensity of the band positively correlates with the abundance of the OTU in the sample.

While in theory the technique is capable of separating amplicons with a single base pair difference, in practise the resolution obtainable from gels is limited. There are two main limitations of molecular fingerprinting techniques. The first is multiple sequences migrating to the same position on a gel and as a result only one band is detected. This leads to an underrepresentation of the true number of unique OTUs in a mixed population sample, masking the true diversity (Muyzer & Smalla 1998). A single band which incorporates mixed sequences will also lead to false classification of abundance, potentially leading to the assumption of a highly abundant OTU. The second important limitation involves a single organism containing multiple copies of the target gene. If the multiple copies are homogenous then this may lead to an overestimation of the true abundance of an organism, as described above. However, if the multiple copies of the target gene are heterogeneous then multiple bands from the same organism will appear on the gel at different positions (Kang *et al.* 2010). This will lead to an over representation of the true diversity of samples. Moreover, if sequence information is required for taxonomic classification from the molecular fingerprint then individual bands can be excised and sequenced. This process can be time consuming and the added expense can be relatively high. Furthermore, the hypervariable regions utilised typically generate short amplicons of around 200 bp which limits the resolution of taxonomic identification (Temmerman & Masco 2003).

1.2.4.4 Metagenomics (next generation sequencing)

Metagenomic technology has revolutionised studies of microbioal diversity. The advent of NGS has facilitated the generation of unprecedented amounts of data, at a depth and resolution far greater than previously possible. There are several NGS platforms

available with each involving a complex interplay of enzymology, chemistry, high-resolution optics, hardware, and software engineering. Multiplexing of samples is typically used to reduce costs. Here, barcodes unique to each sample are incorporated into the primers resulting in the generation of barcoded amplicons, which can be mapped back to the original sample during bioinformatic processing (Siqueira *et al.* 2012). Ultimately each technology aims to amplify single fragments and perform sequencing reactions on the amplified fragments.

454 pyrosequencing was the first NGS platform available as a commercial product (Margulies *et al.* 2005). This platform requires a 454 GenomeSequencer FLX instrument (Roche Applied Science). In this system, DNA fragments are ligated with adapters which cause the binding of one fragment to a 28 μm bead. Emulsion PCR is then carried out for fragment amplification, which is necessary to obtain sufficient light signal intensity for reliable detection in the subsequent sequencing by synthesis steps (Ansorge 2009). Following emulsion PCR, beads are treated with denaturant (removing untethered strands) before hybridisation based enrichment for amplicon bearing beads. A sequencing primer is then hybridised to the universal adaptor and sequencing is performed based on the pyrosequencing method (Ronaghi & Karamohamed 1996). Beads are deposited into a picotiter plate, which is designed to hold a single bead in each well. Smaller beads containing the enzymes adenosine triphosphate (ATP) sulfurylase and luciferase are also added. The picotiter plate is positioned opposite a charge coupled device (CCD) camera. The first four nucleotides on the adapter fragment are TCGA which correspond to the sequential flow of nucleotides into the flow cell. This allows the 454 base calling software to calibrate the light emitted by the incorporation of each nucleotide as the dNTPs are sequentially added (Mardis 2008). Unlabelled dNTPs continue to be added to the reaction sequentially, when this results in an incorporation event pyrophosphate (PPi) is released. The PPi is subsequently

converted to ATP by sulfurylase which provides the energy to luciferase to oxidiase luciferin into oxyluciferin, generating of a burst of visible light which is detected by the CCD camera. Remaining nucleotides are degraded by apyrase before the next dNTP is added to the reaction system and the process is repeated (Ronaghi 2001). Because nucleotides are added sequentially, it is known which dNTP was incorporated into the sequence to produce the light signal. The amount of light signal emitted corresponds to the number of a specific nucleotide incorporated at that cycle. However, the base calling software is unable to accurately interpret long homopolymer runs (E. Mardis 2008). It has been shown that a maximum of 10 identical adjacent nucleotides can be incorporated in the presence of apyrase, but this may require specific software algorithms (Ronaghi 2001).

More recently the sequencing by synthesis chemistry (Bentley & Balasubramanian 2008) used in the Illumina sequencer has gained popularity with the release of the HiSeq 2000 in 2010 followed by the benchtop MiSeq in 2011 (Loman *et al.* 2012). Sequence by synthesis is based on the sequence information being obtained during the synthesis of a DNA strand. First, DNA fragments are ligated at both ends to adapters, denatured into single strands, and immobilised at one end to a flow cell which is coated densely with the adapters and the complementary adapters (Ansorge 2009). The flow cell is an 8-channel sealed glass microfabricated device which ensures amplified sequences from the template library remain locally tethered near the point of origin. Each single stranded fragment immobilised at one end to the flow cell hybridises with its free end to the complementary adapter on the flow cell, creating a bridge structure (Ansorge 2009). The adapters on the surface act as primers and, in the presence of the PCR amplification reagents, amplified sequence features are generated by bridge PCR (Adessi *et al.* 2000). Following PCR, each clonal cluster contains ~1,000 copies of the sequence from the template library (Shendure & Ji 2008). After cluster generation the

clusters are denatured into single stranded molecules prior to sequencing. The reaction mixture for the DNA synthesis and sequencing is supplied onto the surface of the flow cell which contains: primers flanking the target gene, DNA polymerase, and four reversible terminator nucleotides that carry a base-unique fluorescent label and have the 3′-OH chemically inactivated to ensure only a single base is incorporated per cycle (Mardis 2008). Imaging follows each nucleotide incorporation step where the specific terminator nucleotide and its position on the flow cell is detected and identified via its base-unique fluorescent label by the CCD camera (Ansorge 2009). The terminator group at the 3′-end of the base is chemically removed allowing the incorporation of the next fluorescent nucleotide in the sequence by DNA polymerase (Mardis 2008). Because another nucleotide cannot be incorporated into the sequence until the previous nucleotide has been recorded, the Illumnia HiSeq and MiSeq platforms do not suffer from errors in homopolymer runs. The MiSeq was also recently compared to other bench top high throughput sequencing platforms and found to have the highest throughput and lowest error rate (Loman *et al.* 2012). However, in the same study, the 454 pyrosequencing bench top instrument (454 GS Junior) generated the longest reads and best assemblies (Loman *et al.* 2012).

A problem currently limiting all NGS platforms is the short length of reads generated and the tendency for the poorer quality scores at the beginning and end of reads, which results in subsequent bioinformatic trimming of already relatively short sequences (Clarridge 2004). As less phylogenetic information is available from short sequence reads the reliability of taxonomic classification is reduced (Shendure & Ji 2008). Nonetheless, targeting the hypervariable regions of the 16S rRNA gene can provide informative bacterial identification, despite the shorter read lengths (Huse *et al.* 2008). Further advances in sequencing technology are focused on generating longer reads. Some recent advances in so called ‘third generation’ platforms are capable of generating

reads greater than 400 bp, allowing identification of many OTUs to the species level (Siqueira *et al.* 2012). NGS is also subject to PCR bias with errors in the sequencing reaction and the risk of chimera formation when incomplete PCR products serve as primers amplifying related fragments (Petrosino *et al.* 2009; Schloss 2009). To overcome this, PCR primers targeting multiple variable regions of the 16S rRNA gene have been designed and such issues can be further minimised by the application of stringent quality filtering steps (Schloss 2009). The most significant problem facing NGS is currently the bioinfomatic processing of the raw sequencing reads. As vast amounts of data can be generated in relatively little time, the computing power needed to process the data is providing a major bottleneck in the workflow. Many bioinformatic pipelines for the processing and analysis of data have been developed. The main software programmes available for microbial ecology are Mothur (Schloss *et al.* 2011) and QIIME (quantitative insights into microbial ecology) (Caporaso *et al.* 2010), each of which implements algorithms from various developers into a step by step pipeline.

1.3 Polymicrobial disease in preterm neonates

As research shifts toward the 'next generation' of microbial ecology, led by the advances in NGS technology, it is becoming increasingly clear that the pathophysiology of several diseases do not satisfy Kochs postulates (Nelson *et al.* 2012). These postulates are based on the isolation of one pathogenic organism, which is responsible for the disease. There is now support for the gut microbiota as a 'super organ' involved in a range of clinical conditions including sepsis, obesity, autism, diabetes and particularly inflammatory mediated conditions including Crohn's disease and NEC (Finegold *et al.* 2002; Conte *et al.* 2006; Scanlan *et al.* 2006; Mai *et al.* 2013). These

diseases are polymicrobial, where an element of the pathophysiology of the diseases is a function of the microbial community present.

Exploring how intervention is affecting the development of the neonatal gut microbiota will be vital to guiding clinical practise. There is significant interest in understanding whether clinical interventions could be tailored to engineer a gut microbiota reflective of a healthy term delivered breast fed infant. In this exciting and fast moving field, the application of systems biology (inter-disciplinary approach) to disease states will go beyond exploring 'what is there' and also address 'what it is doing'. This latter functional aspect will be key to exploring polymicrobial infections which will not fulfil Kochs postulates. Where reproducible single causative organisms are not involved, functional aspects of the ecosystem may exist which are independent of the exact microbial ecology.

The combination of poor host defences and invasive life support mechanisms make premature infants who have extended stays in NICUs particularly susceptible to infection (Kaufman & Fairchild 2004). Despite increased survival and improving long-term outcomes among preterm infants, the prevalence of NEC and sepsis remains high. Combined, these major neonatal pathologies cause more late neonatal deaths than any other single cause, accounting for 21% of deaths within the worldwide preterm population (Berrington *et al.* 2012). Infants who develop NEC and/or sepsis are also at higher risk of poor growth, cerebral palsy, and vision and hearing impairment (Stoll *et al.* 2004). Rates vary between neonatal units depending on case-mix and care practices, and correlate tightly with degree of prematurity. These two major morbidities associated with the preterm population are discussed in detail below.

1.3.1 Necrotising Enterocolitis (NEC)

Despite over 3 decades of research, NEC remains the most common cause of mortality and morbidity affecting the gastrointestinal tract of infants in the NICU (Hunter *et al.* 2008). NEC occurs in 5 - 10% of very preterm infants, but is difficult to define robustly unless there is histological confirmation at surgery or at post-mortem. The incidence of NEC is rising, largely due to advances in neonatal care resulting in increased survival of preterm neonates (Fox & Godavitarne 2012). There is an inverse relationship between the risk of NEC and birth weight or gestational age, so that very low birth weight (VLBW; <1500 g) and significantly preterm infants (<28 weeks gestation) carry the greatest burden of disease (Lin *et al.* 2008; Caplan & Frost 2011). NEC is also associated with subsequent sequelae including serious neurodevelopmental delay, poor growth, intestinal obstruction due to scarring, short bowel syndrome, and potential liver failure due to prolonged hyperalimentation (Embleton & Yates 2008). NEC is rare in term neonates, occurring in 1 in 1000 term births, due to the increased maturity and the differential development of the microbiota compared to preterm neonates (Berrington *et al.* 2012).

NEC is a severe inflammatory disorder with the exact mechanism and role of infectious agents in the disease yet to be fully established (Berman & Moss 2011). The disease can arise in any area of the GIT but it most commonly found in the terminal ileum, caecum, and ascending colon (Santulli *et al.* 1975). The pathophysiology of NEC is understood to be multifactorial, conditioned primarily by immaturity of the gastrointestinal tract (Lin *et al.* 2008). More specifically, the triad of key factors in the development of NEC are gut mucosal damage, enteral feeding (especially formula feeding), and abnormal bacterial colonisation and translocation (Neu 2005; Martin & Walker 2006; Caplan & Frost 2011). It is increasingly clear that NEC is not a single 'disease' but is likely to represent the end result of several interacting and modulating factors. The triad of key

factors together cause a cascade of events, involving an exaggerated and uncontrolled pro-inflammatory response, ultimately leading to necrosis of the gut epithelium. The radiological hallmark is the presence of pneumatosis intestinalis, characterised by bacterial production of hydrogen gas in the intestinal wall, and the pathological findings include necrosis, mucosal oedema, intramural air, and haemorrhage (Mannoia *et al.* 2011).

In preterm neonates the mucosal defences are weakened as the GIT is not fully developed and the intestinal villi height and barrier functions are compromised (Santulli *et al.* 1975). In a weakened GIT the bacteria and their products may translocate from the lumen, across the mucosa, into the systemic circulation (Harpavat *et al.* 2012). Mature enterocytes can distinguish between commensal and pathogenic bacteria, removing the latter. However, immature enterocytes lining the preterm GIT may mount an exaggerated immune response to the commensal organisms. This leads to a cycle of destructive intestinal damage allowing bacteria more access, which in turn stimulates an increasing inflammatory response (Nanthakumar *et al.* 2000). Thus, the pathogenesis of NEC focuses on the abnormal immune response to commensal organisms and researchers are currently exploring this immune response mechanistically. One key pathway is increased expression of toll-like receptor 4 (TLR4) which is a pattern recognition receptor (PRR) that responds to lipopolysaccharide (LPS) in the cell walls of Gram-negative bacteria and leads to increased apoptotic and proinflammatory responses (Fusunyan *et al.* 2001; Jilling *et al.* 2006; Lin *et al.* 2008). Commensal bacteria suppress the inflammatory response through inhibition of nuclear factor kappa-B (NF-κB), a transcription factor (Wu *et al.* 2012). TLR4 is increased in formula-fed and hypoxia-stressed rats (simulating NEC development) and it is known that endotoxin binds to and activates TLR4 and immature enterocytes also express high levels of TLR4 (Jilling *et al.* 2006). LPS binds to TLR4 causing the activation of chaperone and signal

transduction molecules, which result in NF-κB translocation from cytoplasm to the nucleus where it promotes the expression of multiple proinflammatory cytokines (Wu *et al.* 2012). One such inflammatory cytokine up-regulated by NF-κB is interleukin-8 (IL-8) which is up-regulated strongly in the presence of LPS by fetal intestinal cells, but not in mature enterocytes (Nanthakumar *et al.* 2000). IL-8 serum levels have also been shown to positively correlate with NEC severity (Nanthakumar *et al.* 2011). Up-regulation of other cytokines including IL-12 and IL-18 has also been implicated in NEC development using rat models (Halpern & Holubec 2002). Greater understanding of the inflammatory cascade in NEC is paramount to developing a prevention strategy which targets critical stages within the cascade. For example, levels of platelet activating factor (PAF) have been found to rise in stool following NEC and administration of PAF to hypoxic-stressed rats induces NEC, but PAF antagonists reduced the incidence and severity (Caplan & Hedlund 1997; Amer *et al.* 2004). Comparably, tumor-necrosis factor alpha (TNF-α) is secreted by activated macrophages to promote inflammation and may trigger the production of PAF in neighbouring cells (Harpavat *et al.* 2012). TNF-α may be implemented in the pathogenesis of NEC and might have potential use as a biomarker in predicting disease as it is found in elevated levels in serum of NEC infants compared to controls (Caplan *et al.* 1990). Interestingly, TNF-α is also important in mediating the inflammation seen in inflammatory bowel disease (IBD), with IBD therapy specifically targeting TNF-α in inflamed tissue (Harpavat *et al.* 2012).

Studies have shown bacterial colonisation is a pre-requisite for NEC by (i) intestinal necrosis being absent in bacterium-free animal models, (ii) outbreaks of organisms that commonly colonise the gastrointestinal tract in NICUs have been associated with epidemics of NEC, and (iii) NEC does not occur *in-utero* when the gut is sterile (Kaufman & Fairchild 2004). Despite previous studies supporting dysbiosis of the gut

microbiota in disease development, no single causative agent has been reproducibly identified by standard culture or molecular approaches. Recent molecular based studies exploring the gut microbiota in NEC have shown that the diversity of the community is substantially reduced prior to diagnosis and this can lead to dysbiosis and the dominance of a single organism. Bacterial organisms reported to be associated with NEC are normally from the phylum Proteobacteria, commonly within the *Enterobacteriaceae* family (Carlisle *et al.* 2011; Mai *et al.* 2011). *Cronobacter sakazakii* (formerly *Enterobacter*) in particular has been responsible for NEC outbreaks in NICUs and was shown to greatly exacerbate the extent of NEC in animal models (Hunter *et al.* 2008). In one human study, standard culture identified a predominance of Staphylococci in NEC, although sample timing in relation to NEC was unknown, but this was not confirmed by molecular methods (Smith *et al.* 2012). In another study, an increase in the abundance of *Enterobacter*, *Klebsiella* and *Pseudomonas* spp. was reported (Björkström *et al.* 2009). Molecular analysis of stool after development of NEC indicated that the gut microbiota post NEC diagnosis was lower in diversity, with a greater relative abundance of Proteobacteria (Wang *et al.* 2009). Similarly, a pyrosequencing study sampling the gut microbiota prior to NEC diagnosis and within 72 h of diagnosis, identified a bloom in Proteobacteria and a decrease in Firmicutes during the interval between sampling (Mai 2011). In other studies, no significant differences in the community profiles of infants with NEC compared to controls was found (Mshvildadze *et al.* 2010).

Due to the inconsistencies in identifying a candidate organism between studies, it is unlikely a single organism is accountable for NEC. The complex multifactorial pathophysiology appears to be influenced by a variety of bacterial genera, acting individually or promoting shifts in communities. Host-bacterial interactions have been characterised, with the activation of a pro-inflammatory, pro-apoptotic, or antiapoptotic

pathway involving the interaction of microbial-associated molecular patterns (MAMPs) with a specific PRR on host cells, as in the NF-κB pathway (Collier-Hyams & Neish 2005; Lin & Stoll 2006). The neonatal intestine first encounters MAMPs following birth with the introduction of gut microbiota from the surrounding environment. As outlined previously, commensal bacteria can suppress inflammatory signalling in intestinal epithelia by inhibition of the NF-κB pathway. Thus, it has been proposed that hyperactive inflammation in preterm infants could be caused by dysbiosis of the normal commensal bacteria, and subsequent lack of bacterially-mediated dampening of inflammatory pathways (Lin & Stoll 2006).

Noteworthy is pathogenic causes of NEC are also not limited to bacteria, with fungal species also being implicated in the disease pathophysiology (Gibbs *et al.* 2007). Fungi represent a large kingdom of eukaryotic organisms that are separate from plants and animals, with the major difference arising due to fungal cell walls containing chitin. A retrospective study found that 15% of all infants diagnosed with NEC had concurrent invasive candidiasis (Coates *et al.* 2005). However, in these cases it is difficult to assess whether the presence of *Candida* spp. was contributing to the disease pathophysiology, or rather presenting as an opportunistic infection as a result of the disease. Importantly, *Candida albicans* has been found in cohorts despite universal prophylaxis with oral nystatin suggesting suppression of *C. albicans* rather than elimination to prevent invasive disease (LaTuga *et al.* 2011). The same authors also showed the presence of *Candida parapsilosis* and *Candida tropicalis* which are less susceptible to nystatin and uniquely *Candida quercitrusa,* typically described in fruit crops, was identified in 6 of the 7 infants (LaTuga *et al.* 2011). Despite their resistance to nystatin, it is unclear if these species play a role in the pathogenesis of NEC.

Archaea, one of the three domains of life (with eubacteria and eukaryote), are highly diverse prokaryotes consisting largely of extremophiles that thrive in extreme

environments including hot water springs, salt brines, and extremely acidic and anoxic niches (Liu & Whitman 2008; Jarrell *et al.* 2011). However, with the advent of NGS it increasingly recognised that archaea are not relegated to existence in extreme environments. Indeed, anaerobic archaea have recently been detected in human stool, vaginal, and oral microbiota demonstrating their ability to colonise the human host (Kulik & Sandmeier 2001; Eckburg *et al.* 2003; Vianna & Conrads 2006). Although work on archaea in the gut microbiota of neonates is limited, particularly for preterm infants, a study by Palmer *et al.* (2007) explored the gut microbiota in full term health infants and showed the prevalence of archaea was considerably lower and more variable than fungi or bacteria. Using qPCR, archaea was not detected in all infants and appeared only transiently over the first few weeks of life. Only one infant was found to still harbour archaea after week 5 of life. However, it is important to note that the DNA extraction was optimised for bacteria (Palmer *et al.* 2007). In healthy human adults, hydrogen consuming methanogens are the predominant archaea in the GIT with *Methanobacter smithii* comprising up to 10% of all anaerobes (Human Microbiome Project Consortium 2012; Miller & Wolin 1986). *Methanosphaera stadmanae* and *Crenarchaeotes* have also been detected but comprise only a minor portion of the gut microbiota (Rieu-Lesme *et al.* 2005). A study exploring the gut microbiota in Korean adults also recovered sequences from halophilic archaea, with *Halorubrum koreense* strain B6 showing the greatest prevalence, reported to be a consequence of diet (Nam *et al.* 2008). Archael diversity in the GIT was found to be more similar in each individual than the bacterial or fungal community (Nam *et al.* 2008). Whether archaeal pathogens exist is still open to debate due to lack of definitive evidence. Characterisation of the structure and function of archaea is challenging due to difficulties in the isolation and cultivation (Eckburg *et al.* 2003). In contrast, archaea in mutualistic relationships may provide health benefits to the host. For example, *Methanobrevibacter smithii*

cocolonization with *Bacteroides thetaiotaomicron* was found to produce a significant increase in host adiposity in a mouse model (Samuel & Gordon 2006). In terms of NEC development, a parallel might be drawn from findings in IBD; the presence of methanogenic archaea in the human gut was indicative of a healthy microbiota, with reduced methanogen presence in individuals with IBD (Scanlan *et al.* 2008).

Studies exploring the role viruses in the pathogenesis of NEC are limited, despite their potential importance. Entero-pathogenic viruses generally infect epithelial cells causing cell destruction which can trigger increased intestinal fluid output (diarrhea) in mature infants (Lodha *et al.* 2005). However, for the preterm neonate who may be incapable of initiating increased intestinal fluid output, the viral infection may lead to epithelial damage and intestinal perforation (Kliegman *et al.* 1993). Due to reduced regenerative capacity of preterm neonates, the epithelial damage caused by entero-pathogenic viruses may lead to the development of NEC. Torovirus is one such virus and was shown to have a much greater prevalence in the stools from patients with NEC (48%) compared to controls (14.5%) (Lodha *et al.* 2005). In a study exploring only infants with NEC, human astrovirus was found in 19% of the patients but had no significant effect on disease severity (Bagci & Eis-Hübinger 2008). In a recent case report, NEC was found to be associated with cytomegalovirus (CMV) and Proteobacteria providing evidence of the potential for viral-bacterial interaction influencing the hosts susceptibility to NEC (Tran *et al.* 2013). Interestingly, in adult IBD, concomitant CMV infection is associated with increased disease severity and it has been suggested that CMV increases vulnerability to bacterial invasion and exaggerates the intestinal immune response (Garrido *et al.* 2013; Tran *et al.* 2013). While conserved marker genes such as 16S can be used for detection of all bacteria, there are no RNA or DNA sequences capable of detecting all viruses (Thurber *et al.* 2009). To overcome this, deep sequencing of all viral RNAs and DNAs has been applied to GIT to accurately characterise the diversity

and complexity of the viral community. It has been shown that a large number of viruses within both the infant and adult GIT are bacteriophages (viruses that infect bacteria), although the adult community had much greater diversity (Breitbart *et al.* 2003; LaTuga *et al.* 2011). In a healthy term neonate the meconium was absent of viral particles, but samples obtained after 1 and 2 weeks of life showed the infant harboured 10^8 viral particles per gram of faeces and the overall viral community composition changed dramatically between each week (Breitbart *et al.* 2008). This suggests the viral community in the neonate is dynamic but, importantly, the abundant viral sequences did not originate from feed suggesting acquisition is non-dietary. Phage may influence the composition of bacterial populations in the intestine by infecting and lysing a specific host, allowing another bacterial strain the opportunity to become abundant (Thingstad & Lignell 1997). This model of predator-prey dynamics is termed "kill the winner" and involves blooms of a specific microbial species leading to blooms of their corresponding phage, followed by decreasing abundances of both (Thingstad 2000). This may relate to the dysbiosis event reported prior to NEC and of subsequent dominance of the gut microbiota by a single organism (Morrow *et al.* 2013). If the observed dysbiosis is causative to the disease then the initial event may be triggered intrinsically, as a result of phages infecting other abundant hosts. Additionally, through lysogenic conversion of the gut microbiota, phages may introduce new phenotypic traits, such as antibiotic resistance and the ability to produce exotoxins (Breitbart *et al.* 2003).

Because the candidate organism varies between studies it is unlikely that NEC is the result of colonisation with a specific organism, rather it is a polymicrobial disease arising as a consequence of community interplay. An important consideration when interpretating any results is whether the observed dysbiosis in the gut microbiota is involved in the pathophysiology of NEC or if it is simply a consequence of disease

progression. For example, in cases of concurrent invasive candidiasis in NEC infants, it is difficult to assess whether the presence of *Candida* spp. is contributing to disease pathophysiology, or rather presenting as an opportunistic infection as a result of the disease.

1.3.2 Sepsis

Neonatal sepsis is the most common cause of neonatal death worldwide with the incidence in the developed world thought to be around 4 cases per 1000 live births. The incidence is higher in the developing world with up to 38 cases per 1000 live births. Analogous to NEC, the level of maturity, measured as birth weight or gestational age, is the most potent indicator of susceptibility to infection (Stoll 1994; Guthrie *et al.* 2003). Thus sepsis is also most common in preterm neonates as a result of the immaturity of the intestinal motility and barrier function. In infants born preterm, around 20 - 30% of all very-low-birth-weight (VLBW; <1500g) infants held in NICU's will suffer from sepsis at some stage; this risk increases to 35% in neonates of extremely-low-birth-weight (ELBW; <1000g) and to nearly 50% in neonates of less than 750g (Stoll *et al.* 2004). Neonatal sepsis can be divided into two types, early onset sepsis (EOS; occurring with 72 hours of birth) which occurs in about 1.5% and late onset sepsis (LOS; occurring after 72 of birth) which occurs in about 21% of VLBW infants. The two types can be grouped under the term 'sepsis' which refers to an infant developing at least one case of proven sepsis irrespective of hours since birth. Due to the increased involvement of the microbial community in LOS, the term sepsis will be used throughout this thesis but will refer, in large, to LOS.

Sepsis is caused by the direct bacterial translocation from the gut into the bloodstream and is typically diagnosed by positive blood culture, along with antibiotic treatment for

a minimum of 5 days and clinical signs suggestive of infection. Sepsis is associated with subsequent sequelae including prolonged ventilation and need for intravascular access, bronchopulmonary dysplasia, NEC, and an increased length of hospital stay. In terms of the gut microbiota in sepsis, there is comparison with the development of NEC. Like NEC, dysbiosis of the gut microbiota is associated with sepsis facilitated by prolonged broad spectrum antibiotic treatment, which profoundly decreases the gut microbial diversity and promotes pathogen predominance (Mai *et al.* 2013). While microbial patterns in the meconium are similar, infants at high risk of sepsis have been shown to possess a less diverse gut microbiota from birth until the onset of sepsis (Madan *et al.* 2012). The gut microbiota in infants who are diagnosed with sepsis is significantly different from that of control infants, with an increased prevalence of Proteobacteria and Firmicutes. Contrary to the inconsistency of a dominant organism in NEC, *Staphylococcus* spp. are regarded as the most frequent bacterial species in the pathogenesis of sepsis worldwide, with increased abundances of these species shown in infants diagnosed with the disease (Venkatesh & Abrams 2010).

Colonisation of the gut microbiota in sepsis by fungi remains relatively unstudied compared to bacteria, despite the incidence of candidemia rising steadily (Benjamin *et al.* 2003). As with bacterially mediated sepsis, blood culture is the 'gold standard' for the diagnosis of neonatal candidemia. Although sensitive for bacterial pathogens, it is a poor diagnostic tool for invasive fungal infections, with as many as 50% of candidemia not being directly detected (Manzoni *et al.* 2008). *C. albicans* is the most prevalent species in the gut and despite antifungal prophylaxis *Candida* spp. remain a leading cause of infectious mortality in the NICU, accounting for around 50% of the cases (Tirodker *et al.* 2003; Zaoutis *et al.* 2007).

As outlined in the previous section, archaea and viruses colonising the gut may also influence the disease status of the human host. As with NEC, no specific research has

been carried out on these domains in the pathogenesis of sepsis. However, it is likely the whole consortium of the gut will be influencing dysbiosis events to some degree. As previously described, viruses may damage the epithelial cell and subsequently host defences allowing the translocation of pathogenic species, such as *Staphylococcus* spp. which may proliferate and ultimately cause sepsis (Venkatesh & Abrams 2010). As with NEC, the question remains; does dysbiosis in the gut microbiota cause sepsis or is it a result of disease progression? Future work will need to employ large cohorts with robust sampling pre and post disease to address this issue of cause or effect.

1.4 Biomarkers for NEC and sepsis

Early identification of developing NEC and sepsis in preterm infants is essential to reduce the devastating effects of the diseases. The current best definition of a biomarker is provided by the National Institute for Health (NIH) working group who define a biomarker as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" (Atkinson *et al.* 2001). The diagnosis of NEC and sepsis remains difficult with advanced stages of disease usually present before symptoms can be detected and in the case of blood culture for sepsis, results may take more than 48 hours and false negatives due to low sample volume are common (Ng *et al.* 2010). Thus culture alone lacks the ability to identify quickly which infants really need antibacterial treatment, and which do not. Concerns around development of antibiotic resistance and adverse effects on the faecal microbiome suggest 'unnecessary' antibiotic use should be minimised. Consequently, much research is currently focused on the detection of an appropriate non-invasive biological biomarker which is capable of predicting which

infants are likely to develop the disease so appropriate timely intervention can be made. As NEC and sepsis are both associated with the gut microbiota, one approach might seek to detect specific organisms related to gut health. It has been shown that *Enterococcus faecalis* can induce anti-inflammatory cytokine IL-10 via regulation of the nuclear receptor Peroxisome proliferator-activated receptor gamma (PPAR-γ) (Are *et al.* 2008). Conceivably, systematic monitoring of an *E. faecalis* in the preterm gut microbiota could be a valuable biomarker in predicting NEC and sepsis (Braniste & Pettersson 2012).

Neonatal sepsis and NEC is associated with hyper-inflammatory host response and much research has focused on a specific panel of acute phase protein biomarkers as potential markers of the inflammatory cascade as well as cell surface antigens. For example, calprotectin, a heterodimeric peptide (36 kDa) which is released following intestinal neutrophil influx is readily detectable in stool and plasma (D'Incà *et al.* 2007). Stool calprotectin, which represents gut wall inflammation, was recently shown to hold promise as a non-invasive diagnostic marker for both NEC and sepsis (Thuijls *et al.* 2010; Dhas *et al.* 2012). Furthermore, stool calprotectin levels have been previously validated as an accurate marker of IBD in both adults and children (Carroccio *et al.* 2003). Although more than 170 different biomarkers have been evaluated for prognosis and diagnosis in sepsis alone, none have sufficient specificity or sensitivity for clinical use (Pierrakos & Vincent 2010). Rather than focus on a single biomarker it has been proposed that combinations of multiple biomarkers may be more effective, thus further acute phase protein biomarkers have also been explored. It has been shown that serum amyloid A (SAA) is equally effective as the commonly used C-reactive protein (CRP) and procalcitonin (PCT) and that serial measurement of all three markers could be used in combination in the diagnosis of NEC (Çetinkaya *et al.* 2011).

More recently the search for a diagnostic test for NEC and sepsis has focused on cytokines and chemokines, particularly IL-6, -8, and -10 and TNF-α (Mussap *et al.* 2012). The serum concentrations of these potential biomarkers change over the course of systemic inflammation with disease progression linked to an exaggerated proinflammatory response without adequate anti-inflammatory compensation (Ng *et al.* 2003). IL-6 is the most investigated cytokine as a potential biomarker for neonatal disease as it is very sensitive; however, it has a very short half-life leading to decreased sensitivity after the start of therapy (Procianoy & Silveira 2004). Cell marker biomarkers have also been explored with neutrophil CD64 perhaps the most commonly investigated. CD64 is the first of three receptors that bind the Fc portion of IgG antibodies and when the immune system encounters an infectious agent the surface expression of CD64 is greatly up-regulated (Standage & Wong 2011). Advances in flow cytometry have enabled rapid and accurate quantification of CD64 and importantly for neonates only minimal blood volume is required (Bhandari *et al.* 2008). CD64 is more sensitive and specific than CRP for detecting systemic infection in adults and children (Gude 2012). It has been shown to be a highly specific indicator of neonatal sepsis but with only moderate sensitivity (Ng *et al.* 2004; Streimish & Bizzarro 2012). In a study by Ng *et al.* (2004), using an optimal cut-off value CD64 was able to correctly identify all sepsis cases and a case of severe NEC. However, 5 infants with clinical pneumonia were missed and 43 infants were incorrectly classified into the infected group. Thus, as alluded to previously, combinations of biomarkers used in parallel may provide the most robust classification of a neonates disease status.

Although candidate markers are targeted based on their crucial roles in inflammation and infection, this methodology is limited to markers of known proteins. To overcome this limitation, Ng *et al.* (2010) used mass spectrometry (MS) based proteomic profiling technology as an unbiased approach to explore host response biomarkers for predicting

NEC and sepsis in preterm infants. From a large consortium of proteins, SAA and apolipoprotein CII (apoC2) were detected as potential biomarkers from plasma capable of early detection of NEC and sepsis with great accuracy. Both of these potential biomarkers are apolipoproteins, which bind lipids to form lipoprotein. Lipoprotein binds to LPS and, as discussed previously, this can trigger a cascade of inflammatory responses (Ng *et al.* 2010).

Another field which is currently emerging for biomarker research is transcriptomics. This methodology evaluates gene-expression patterns and involves the study of all messenger RNA (mRNA) molecules or transcripts and the evaluation in transcription initiation, processing, and degradation of proteins (Young *et al.* 2009). As samples are amplified by PCR only small amounts of biological material are required that can be obtained non-invasively. This approach will require the extraction of high quality RNA which can be difficult and will still be subject to PCR bias. Metabolomics on the other hand does not involve PCR and is receiving increased interest for neonatal biomarker discovery. Metabolomics is the systematic study of the complete set of metabolites in a biological sample in which the metabolic status of the individual is an accurate representation of the disease status (Mussap *et al.* 2012). The technique is able to utilise non-invasive sampling of urine which reflects changes to the metabolic state and extent of absorption and stool which reflects differences in gut flora diversity and any impact on gut permeability and function. Urine is considered the best biological sample to use in metabolomics as it contains the intermediate metabolites which reflect specific metabolic processes related to the current disease status in real time (Fanos *et al.* 2011). While studies utilising transcriptomics and metabolomics for biomarker discovery in NEC and sepsis are currently lacking, it is reasonable to propose that in future a systems biological approach will yield some major advancements in the field.

1.5 Preventative strategies and treatment for NEC and sepsis

As NEC frequently progresses from minor signs of intestinal inflammation to major necrosis within hours, even the most highly specific and sensitive biomarkers may fail to predict the disease. Thus, the development of preventative strategies is paramount (Wu *et al.* 2012). However, as the exact pathogenesis of NEC is yet to be determined, formulating effective preventative and treatment plans remains challenging (Lin *et al.* 2008). Breast milk contains several immune-protective and growth factors, bioactive immune-modulatory cells and other 'immunonutrients' including amino acids, fatty acids, lysozyme, lactoferrin, minerals and metals such as zinc, and prebiotic oligosaccharides (Bhatia 2010). Glutamine and arginine influence gut integrity and sepsis and several vitamins have key roles in antioxidant protection (Embleton & Yates 2008). EBM also contains live bacteria which regulate host-microbe interactions and modify infant gut microbiota development, although to what extent the gut microbiota reflects EBM is currently unclear. For example, passive immunoprotection is provided by maternal secretory Immunoglobulin A (IgA). IgA is thought to promote biofilm formation to facilitate colonisation by protective bacteria, subsequently preventing colonisation by pathogenic species. Consequently, receipt of maternal breast milk is considered a key factor facilitating the development of a healthy microbiota (Martín *et al.* 2009; Roger *et al.* 2010).

The notion that breast milk stimulates a gut microbiota dominated by Bifidobacteria is still an area of active debate with many recent studies finding no association, particularly in preterm populations (Palmer *et al.* 2007). Also provoking controversy is the role and extent to which *Bifidobacterium* sp. are beneficial to human health. This stems from inconclusive research into the effectiveness of probiotic and prebiotic

supplementation at supporting healthy development of the gut microbiota (Szajewska 2010). In preterm birth it is more challenging to adhere to the feeding ideology of exclusive maternal breast milk, thus much interest is currently focused on manipulations of the microflora with prebiotics and probiotics. Prebiotics are non-digestible short chain length carbohydrates that promote the growth and activity of desired beneficial organisms (Lee 2011). They appear to reduce the number of coliforms, stimulate the growth of *Bifidobacterium* and *Lactobacillus* spp., and could potentially stimulate a balanced and effective mucosal immune system in newborns and infants (Nakamura *et al.* 2009). A serendipitous increase in *Lactobacillus* spp. in preterm infants was associated with 1% lactulose added to feeds, potentially attributable to its prebiotic effects (Riskin *et al.* 2010).

Probiotics consist of viable microorganisms, predominantly *Bifidobacterium* and *Lactobacillus* spp., in sufficient quantity to influence the gut microbiota by colonisation. Probiotics have received the greatest interest compared with prebiotics and it is hypothesised that administering probiotics results in a reduction in the growth of potential pathogens including enterobacteria, enterococci and clostridia (Deshpande *et al.* 2007). They may also improve intestinal permeability and modulate the development and persistence of an appropriate mucosal immune response. While current research into probiotic supplementation yields mixed results in terms of efficiency, it should be noted that probiotics are not considered dangerous nor have they been shown to have a detrimental effect on the host (Li *et al.* 2013; Nair & Soraisham 2013). Probiotics reduce the incidence of NEC in many studies, but interestingly do not affect sepsis (Deshpande *et al.* 2010). Few have explored associated changes in the microbiota as a result of probiotic administration. Work that has been reported indicates increased probiotic numbers (*Bifidobacterium* and *Lactobacillus* spp.) at the expense of *Enterobacter*, Enterococci and Clostridia — all organisms previously associated with

NEC (van Acker *et al.* 2001; France *et al.* 2007). Supplementing *Lactobacillus casei* increased the amount in infant stool and stabilised bowel flora (Cox *et al.* 2010). As prebiotics, probiotics, and synbiotics (a combination of prebiotics and probiotics) are increasingly implemented into the practises of neonatal units, future research will be able to explore the age of commencement, optimal dose, and species combinations to maximise and potential benefit. The ability to define and then replicate a 'healthy' gut microbiota would be of practical benefit to infants.

The supplementation of the neonate's diet is an attractive area of research due to possible antimicrobial activity and therapeutic potential of the supplement. Lactoferrin has received significant interest recently with application aimed particularly for preterm infants. Lactoferrin is the major whey protein in human colostrum and breast milk and is a key component of the immune response. It enhances cell proliferation of enterocytes and aids tight endothelial cell junctions (Bäckhed 2011). At lower concentrations, lactoferrin stimulates differentiation of enterocytes and expression of intestinal digestive enzymes. Lactoferrin also suppresses free radical activity when iron is added to milk suggesting further anti-inflammatory actions that could modulate the pathogenesis of diseases linked with free radical generation: NEC, retinopathy of prematurity (ROP) and bronchopulmonary dysplasia (BPD) (ELFIN 2013). Studies on the mechanisms by which lactoferrin exerts its effect in vivo show it is bacteriostatic by inhibiting growth by sequestering iron (Ochoa *et al.* 2006). It has also been shown to bind to LPS on the cell surface of a number of Gram-negative enteropathogens, inhibiting surface expressed virulence factors (Kaur *et al.* 2010). It can also inhibit viral attachment and replication and exerts a fungicidal activity. A recent study showed promise, demonstrating that oral lactoferrin prophylaxis reduced the incidence of sepsis in preterm infants, with the greatest effect occurring amongst ELBW infants (Venkatesh & Abrams 2010). As a positive candidate for regulating a healthy gut microbiota, future

research on the mechanisms and mode of actions, particularly on the composition of the gut microbiota, will be of great importance.

Treatment of NEC and sepsis is carried out on a case by case basis and often relies heavily on the attending clinician. Treatment of NEC is through the administration of broad-spectrum antibiotics following diagnosis for a minimum of 5 days. If the symptoms persist following medical management or in the most severe cases the infant may undergo surgery to remove areas of necrotic bowel segments. Treatment of sepsis is based on antibiotic or antifungal prophylaxis for a minimum of 5 days for the management of bacterial or fungal mediated sepsis, respectively. In cases of negative blood culture but signs suggestive of infection the average days of antibiotic treatment is 5 ±3 days (Cordero & Ayers 2003). Prolonged administration of antimicrobial agents for greater than 5 days is associated with NEC and death (Cotten *et al.* 2009; Alexander *et al.* 2011).

1.6 Aims and objectives

The research conducted in this thesis primarily aimed to follow preterm neonates at risk of NEC and sepsis longitudinally, to explore the temporal progression of the gut microbiota from birth until discharge from the NICU. A range of culture dependent and molecular techniques were used to explore both the bacterial and fungal communities. The role of the gut microbiota as a causative agent in the pathogenesis of NEC and sepsis was the focus of analysis.

The specific aims of each chapter are outlined below…

Chapter 3: The gut microbiota is a highly complex ecosystem increasingly associated with disease pathophysiology. Despite revolutionary advances in traditional culture and molecular based techniques, the coherence of these techniques remains elusive. We aimed to compare the performance of these techniques in assessing the gut microbiota of preterm infants.

Chapter 4: To describe gut colonisation in preterm infants using standard culture and 16S rRNA profiling, exploring differences in healthy infants and those who developed late onset infections.

Chapter 5: Evidence suggests that microbial communities in the preterm gut may influence the development of NEC and sepsis. Existing data often neglect fungi and whether bacteria were metabolically active or not. We sought to characterize the bacterial and fungal stool flora of preterm neonates and organism viability and evaluate any associations with NEC and sepsis

Chapter 6: The preterm gut microbiota is a complex dynamic community influenced by genetic and environmental factors and is implicated in the pathogenesis of NEC and sepsis. We aimed to explore the longitudinal development of the gut microbiota in preterm twins to determine how shared environmental and genetic factors may influence temporal changes and compared this to the expressed breast milk (EBM) microbiota.

Chapter 7: The gut microbiota is significantly associated with the development of NEC and sepsis. Previous studies are limited by cohort size, poor sampling, and methodological restrictions. We aimed to extensively explore the differential community development in patients with NEC and sepsis, matched to controls.

2. Methods

2.1 Patient cohort and samples

Stool samples were collected from preterm infants <32 weeks gestation cared for in the Royal Victoria Infirmary, Newcastle upon Tyne. Stool samples were collected directly from the nappy and when possible were immersed in RNA*later* (Ambion), which is a bacteriostatic agent offering immediate RNA stabilisation and protection. All samples were transferred to the microbiology department at the Freeman Hospital where they were stored at -20 °C.

Clinical information was obtained from notes: Information on discrete (diagnosis of sepsis and NEC, gender, caesarean or vaginal birth, antifungal prophylaxis, mortality, and antibiotic administration) and continuous (day of life (DOL), birth weight, gestation age) variables were given for every patient relating specifically to each sample. NEC was categorised independently by two clinicians (Janet Berrington and Nicholas Embleton) from notes, x-ray and operative findings and classified as surgical (diagnosis confirmed operatively) or medical NEC (un-equivocal pneumatosis). Sepsis was defined by positive blood culture, along with antibiotic treatment for a minimum of 5 days and signs suggestive of infection.

2.1.1 Technique comparison

11 preterm infants had 17 stool samples collected as part of routine surveillance. In 3 patients additional longitudinal samples were collected. 200 mg of fresh sample was

subjected to quantitative aerobic and anaerobic culture; the remainder of the sample was stored at -20 °C for subsequent molecular analysis.

2.1.2 Culture and DGGE: preliminary study

The first and weekly stool samples were collected from preterm infants when available and subjected to quantitative aerobic and anaerobic culture, the remainder of the sample was stored at -20 °C. 38 infants contributed 99 stool samples to the routine microbial culture analysis. Sample volume allowed for 27 patients contributing 44 stool samples to be further analysed using DGGE.

2.1.3 Bacterial and fungal viability

A total of 32 preterm infants had 136 first and weekly stool samples collected and frozen at -20 °C. Of these samples, in 25 infants contributing 65 samples a portion was preserved in RNA*later* (Ambion) and thus available for analysis of the viable community.

2.1.4 Twin comparison

In 12 sets of twins and 1 set of triplets (n = 27) a total of 173 stool samples (n = 173) were collected from birth until 8 weeks. 18 expressed breast milk samples salvaged from residual feeds in 3 sets of twins and the set of triplets were also analysed.

2.1.5 Gut microbiota in NEC and sepsis

A total of 42 preterm infants were enrolled in the study contributing a total of 747 stool samples. Patients were split into two groups; Extremely Preterm (group "EP": gestational age 23 - 26 weeks) and Very Preterm (group "VP": gestation age 27 – 30 weeks). Each group consisted of 21 patients where 7 patients developed proven NEC and/or sepsis matched to 14 patients who acted as controls. 408 samples were included in EP (mean 19 per patient, range 8 - 26) and 339 samples were included in VP (mean 16 per patient, range 8 - 25) owing to the increased stay on the NICU common to the more preterm infants. Sampling aimed for daily collection where possible and the most informative samples, based on diseased status matched to healthy controls, were retrospectively chosen for analysis. Where possible, daily samples were analysed for the first 10 DOL, then every other day up to day 10-20, then every third day up to day 30, then every fourth day up to day 40 and so on until post day 70 where weekly samples were selected until discharge. In cases of NEC and/or sepsis, an increased number of samples preceeding and following diagnosis were also analysed.

2.2 Routine culture of stool samples

Routine culture of stool was carried out at the Freeman hospital. Briefly, suspensions were prepared from 200 mg stool material suspended, then serially diluted to 10^{-5} in quarter-strength Ringers Solution (Oxoid). An aliquot of each dilution (10 μL) was cultured onto Uriselect™ 4 agar (Bio-Rad) incubated aerobically (37 °C; 24 h) and also onto blood agar and blood agar containing 30 mg/L nalidixic acid - both incubated in an anaerobic chamber at 37 °C undisturbed for 72 hours. Lawns of the 10^{-1} suspension were spread onto Sabouraud agar and incubated aerobically at 30 °C for 72 hours.

Bacterial identification of each colony type was achieved using a combination of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) (Bruker Daltonics) and, where necessary, appropriate API kits (bioMérieux). Colony counts were performed on each species and calculated by the relevant dilution factor to give an indication of the relative amounts of each species in the sample. UriSelect 4 agar was employed for the enumeration of *Enterobacteriaceae*, non-fermentative Gram-negative bacteria, staphylococci and enterococci. Streptococci, lactobacilli and anaerobes were enumerated from blood agar plus nalidixic acid whereas Sabouraud agar was used only for enumeration of yeasts and fungi.

2.3 Nucleic acid extraction

All Nucleic acid extraction was carried out in a class II microbiological safety cabinet (Envair Bio 2+). Extractions were performed at room temperature unless otherwise stated.

2.3.1 DNA extraction of cultured isolates (DNeasy Blood and Tissue Kit)

DNA was extracted from cultured isolates using the DNeasy Blood and Tissue Kit (QIAGEN). The initial stages of extraction were dependent on whether the isolate was Gram-negative or Gram-positive. Initially, the bacterial cells were harvested in a microcentrifuge tube by centrifuging for 10 minutes at 5,000 × g. Gram-negative bacterial cell pellets were then resuspended in 180 µL Buffer ATL. 20 µL of proteinase K was added and the suspension vortexed and incubated for 1 hour at 56 °C with occasional vortexing allowing the digestion of proteins. Vortexing after each step, 200

µL of Buffer AL and 200 µL of ethanol were added to the suspension. For Gram-positive bacteria, cell pellets were resuspended in 180 µL enzymatic lysis buffer (Appendix 1) and incubated for 30 minutes at 37 °C allowing lysis of the bacterial cells by damage to the cell walls. 25 µL of proteinase K and 200 µL of Buffer AL were then added and suspensions were incubated for 30 minutes at 56 °C. 200 µL of ethanol was then added and the suspension vortexed. After these differing pre-treatment steps, both Gram-negative and Gram-positive underwent the same process as follows. The mixture was pipetted into a DNeasy Mini spin column placed in a 2 mL collection tube and centrifuged for 1 minute at 6,000 × g. The collection tube containing the flow-through was discarded and the DNeasy Mini spin column was placed in a new 2 mL collection tube and 500 µL Buffer AW1 was added and the column centrifuged for 1 minute at 6,000 × g. The collection tube containing the flow-through was discarded and the DNeasy Mini spin column was placed in a new 2 mL collection tube and 500 µL Buffer AW2 was added and the column centrifuged for 1 minute at 20,000 × g to dry the DNeasy membrane. The collection tube containing the flow-through was discarded and the DNeasy Mini spin column was placed in a new 1.5 mL microcentrifuge tube and 100 µL Buffer AE was pipetted directly onto the DNeasy membrane. Following incubation for 1 minute the column was centrifuged for 1 minute at 6,000 × g to elute the extracted DNA.

2.3.2 DNA extraction of stool (PowerLyzer™ PowerSoil® DNA Isolation Kit)

DNA was extracted from 100 mg of stool sample for analysis of the total community using the PowerLyzer™ PowerSoil® DNA Isolation Kit (MoBio). Briefly, the stool sample was added to a glass bead tube containing 0.1 mm glass beads before 750 µL of Bead Solution was added and the mix vortexed gently for 10-15 seconds. 60 µL of

solution C1 was then added and the mix inverted several times. Solution C1 is a cell lysis buffer which contains SDS to break down fatty acids and lipids associated with the cell membrane. The glass bead tube was secured in a vortex adapter (MoBio) and vortexed at full speed for 15 minutes to mechanically lyse microbial cells. The glass bead tube was centrifuged at 10,000 × g for 3 minutes and the supernatant (400-500 µL) was transferred to a sterile collection tube. 250 µL of Solution C2 was added, the mix vortexed for 5 seconds and incubated at 4 °C for 5 minutes to precipitate humic substances, cell debris, and proteins. The suspension was centrifuged for 1 minute at 10,000 × g to separate the liquid phase containing the DNA from the solid phase containing the non-DNA organic and inorganic material. Avoiding the transfer of any of the pellet, up to 600 µL of supernatant was transferred to a sterile collection tube and 200 µL of solution C3 was added before a brief vortex and incubation at 4 °C for 5 minutes. Like solution C2, this is a second reagent to precipitate additional humic acid, cell debris, and proteins. The suspension was centrifuged for 1 minute at 10,000 × g, again to separate the liquid phase containing the DNA from the solid phase. Avoiding the transfer of any of the pellet, up to 750 µL of supernatant was transferred to a sterile collection tube and 1.2 mL of solution C4 was added to the supernatant and vortexed for 5 seconds. Solution C4 is a high salt solution which facilitates DNA binding to the silica membrane of the spin filter. 675 µL of the supernatant mix was loaded onto a spin filter and centrifuged at 10,000 × g for 1 minute and the flow through discarded. This was repeated until all of the supernatant mix had passed through the spin filter leaving only DNA bound to the membrane. 500 µL of solution C5, ethanol wash solution, was then added to the spin filter and centrifuged for 30 seconds at 10,000 × g to clean the DNA bound to the spin filter membrane by removing residual salt and other contaminants. The flow though was discarded and the spin filter centrifuged again for 1 minute at 10,000 × g to remove all traces of ethanol which can interfere with downstream

applications. The spin filter was then transferred to a sterile collection tube and 100 µL of solution C6 was added to before centrifugation for 30 seconds at 10,000 × g to elute DNA from the spin filter membrane. Solution C6 is a low salt sterile elution buffer which causes DNA bound to the spin filter in the presence of high salt to be selectively released. The spin filter was discarded and the eluted DNA was stored at -20 °C for use in downstream application.

2.3.3 DNA extraction of milk (PowerFood™ Microbial DNA Isolation Kit)

DNA was extracted from milk and syringe samples for analysis of the total community using the PowerFood™ Microbial DNA Isolation Kit (MoBio). Breast milk samples were obtained from frozen stores of expressed breast milk. Syringe samples were obtained by washing the milk residue from fresh syringes using 2 mL of sterile water. Briefly, 1.8 mL of the of the sample was added to a 2 mL collection tube and centrifuged at 13,000 × g for 1 minute. The supernatant was decanted and the tubes centrifuged at 13,000 × g for an additional 1 minute before removing the remaining supernatant. The pellet, which contains the microbial cells, was resuspended in 450 µL of Solution PF1. Solution PF1 is a strong lysing reagent which includes a detergent to break cell walls and help removed non-DNA organic and inorganic material. The suspension was added to a MicroBead tube which was secured in a vortex adapter (MoBio) and vortexed at full speed for 15 minutes to mechanically lyse microbial cells. The MicroBead tube was centrifuged at 13,000 × g for 1 minute and the supernatant (approximately 400 µL) which contains the DNA was transferred to a sterile collection tube. 100 µL of Solution PF2 was added, the mix vortexed for 5 seconds and incubated at 4 °C for 5 minutes to precipitate cell debris, and proteins. The suspension was centrifuged for 1 minute at 13,000 × g to separate the liquid phase containing the DNA

from the solid phase containing the non-DNA organic and inorganic material. Avoiding the transfer of any of the pellet, up to 450 µL of supernatant was transferred to a sterile collection tube and 900 µL of solution PF3 was added to the supernatant and vortexed for 5 seconds. Solution PF3 is a highly concentrated salt solution which facilitates DNA binding to the silica membrane of the spin filter. 650 µL of the supernatant mix was loaded onto a spin filter and centrifuged at 13,000 × g for 1 minute and the flow through discarded. This was repeated until all of the supernatant mix had passed through the spin filter leaving only DNA bound to the membrane. The spin filter was then placed in a sterile collection tube and 650 µL of solution PF4, an ethanol wash solution, was then added to the spin filter and centrifuged for 1 minute at 13,000 × g to clean the DNA bound to the spin filter membrane by removing residual salt and other contaminants. The flow through was discarded and 650 µL of Solution PF5 was added and centrifuged for 1 minute at 13,000 × g to ensure the complete removal of solution PF4 resulting in higher purity and yield. The flow though was discarded and the spin filter centrifuged again for 2 minute at 13,000 × g to remove all traces of PF5, which can interfere with downstream applications. The spin filter was then transferred to a sterile collection tube and 100 µL of solution PF6 was added to before centrifugation for 1 minute at 13,000 × g to elute DNA from the spin filter membrane. Solution PF6 is a low salt sterile elution buffer which causes DNA bound to the spin filter in the presence of high salt to be selectively released. The spin filter was discarded and the eluted DNA was stored at -20 °C for use in downstream application.

2.3.4 RNA extraction of stool (PowerMicrobiome™ RNA Isolation Kit)

RNA was extracted from 100 mg of stool samples stored in RNA*later* (Ambion) for analysis of the viable community using the PowerMicrobiome™ RNA Isolation Kit

(MoBio). Briefly, the stool sample was added to a glass bead tube containing 0.1 mm glass beads before 650 µL of PM1 and 6.5 µL of β- mercaptoethanol (β-ME) (Sigma-Aldrich) was added prior to the sample fully thawing. Solution PM1 is a cell lysis buffer which contains SDS to break down fatty acids and lipids associated with the cell membrane and β-ME is a reducing agent that permanently denatures RNases. The glass bead tube was secured in a vortex adapter (MoBio) and vortexed at 13,000 × g for 10 minutes to mechanically lyse microbial cells. The glass bead tube was centrifuged at 13,000 × g for 1 minute and the supernatant (500-600 µL) was transferred to a sterile collection tube. A 150 µL aliquot of Solution PM2 was added, the mix vortexed for 5 seconds and incubated at 4 °C for 5 minutes to precipitate humic substances, cell debris, and proteins. The suspension was centrifuged for 1 minute at 13,000 × g to separate the liquid phase containing RNA and DNA from the solid phase containing the protein and cellular debris. Avoiding the transfer of any of the pellet, up to 650 µL of supernatant was transferred to a sterile collection tube and 650 µL of solution PM3 and 650 µL of solution PM4 were added before briefly vortexing. Solution PM3 contains the binding salts for total nucleic acid purification and Solution PM4 is 100% ethanol facilitating DNA binding to the silica membrane of the spin filter. 650 µL of the supernatant mix was loaded onto a spin filter and centrifuged at 13,000 × g for 1 minute and the flow through discarded. This was repeated until all of the supernatant mix had passed through the spin filter leaving only total nucleic acids bound to the membrane. 650 µL of solution PM5, an isopropanol containing wash buffer to remove salts from the membrane, was then added to the spin filter and centrifuged for 1 minute at 13,000 × g for optimal performance of the on-column DNase step. The flow though was discarded and the spin filter centrifuged again for 1 minute at 13,000 × g to remove and residual wash. The spin filter was then transferred to a sterile collection tube and 50 µL of DNase I Solution was added, prepared by mixing 45 µL of Solution PM6 with 5 µL of

DNase I stock solution. The reaction was incubated for 15 minutes allowing the DNase in solution PM6 to soak into the membrane and digest the genomic DNA on the column. 400 µL of solution PM7 was added to the spin filter and centrifuged for 1 minute at 13,000 × g and the flow through discarded. This inactivates the DNase enzyme and removes it from the column along with the digested DNA. 650 µL of solution PM5 was added to the spin filter and centrifuged for 1 minute at 13,000 × g and the flow through discarded. 650 µL of solution PM4 was added to the spin filter and centrifuged for 1 minute at 13,000 × g and the flow through discarded. These solutions are wash buffers and are reapplied to desalt the column before the elution step. The column was centrifuged again at 13,000 × g for 2 minutes to remove residual wash solution. The spin filter was then transferred to a sterile collection tube and 50 µL of solution PM8 was added to the centre of the membrane and allowed to sit for 1 minute before centrifugation for 1 minute at 13,000 × g to elute RNA from the spin filter membrane. Solution PM8 is RNase-Free water which causes RNA bound to the spin filter in the presence of high salt to be selectively released. The spin filter was discarded and the eluted RNA was subject to further DNA removal.

2.3.4.1 Removal of contaminating DNA (TURBO DNase)

The extracted RNA underwent additional DNase treatment using TURBO DNase (Ambion) to further remove contaminating DNA. 5 µL of 10X TURBO DNase buffer and 1 µL of TURBO DNase was added to the extracted RNA and the reaction mixed gently. The reaction was incubated at 37 °C for 20-30 minutes. 5 µL of resuspended DNase inactivation reagent was then added and the reaction mixed well. The reaction was incubated for 5 minutes at room temperature with occasional mixing to redisperse the DNase inactivation reagent. The reaction was centrifuged at 10,000 × g for 1.5 minutes before the supernatant containing the RNA was transferred to a fresh tube.

2.3.5 Confirmation of successful extraction

A 1% (w/v) agarose gel was prepared to confirm that nucleic acid extraction had been successful by electrophoresis of the extracted DNA or RNA. The agarose gel was submerged in 1× TAE and a 5 µL aliquot of nucleic acid was added to 1 µL of 6× bromophenol blue (Appendix 3), mixed by pipetting, and then loaded in to the wells of the agarose gel alongside a 5 µL of Hyperladder 1 (Bioline). In RNA gels a lane containing an ssRNA ladder (New England Biolabs) was also included. The electrophoresis tank was run at a constant current of 120 mA for 24 minutes. For DNA staining, 5 µL of SYBR Safe (Invitrogen; 10,000×) was added to the molten agarose prior to it setting. For RNA staining, following electrophoresis the gel was placed in a staining container with 10 µL of SYBR Gold (Invitrogen; 10,000×) added to 100 mL of 1× TAE. The gel was stained in the dark for 30 minutes at room temperate with gentle agitation. Gels were viewed under U.V. light using the Gel Doc 2000 gel documentation system (Bio-Rad) and quantity one™ software (v4.1.1.) to confirm the presence of nucleic acid product of correct size by comparison with fragments from the standard ladder. Print outs were obtained using Mitsubishi Video copy processor (Model P91).

2.4 Reverse transcription polymerase chain reaction (RT-PCR)

A 10 µL aliquot of RNA was pipetted into a PCR tube along with 1 µL of dNTPs (New England Biolabs; 25 mM) and 1 µL of random hexamer primers (QIAGEN; 0.4 µg/µL) and the mixture was heated to 65 ºC for 5 minutes. The mixture was then chilled on ice and 4 µL of 5× first strand buffer, 2 µL of Dithiothreitol (0.1 M) and 1 µL of RNaseOUT™ (Invitrogen; 40 U/µL) was added and mixed by pipette. The mixture was incubated at 25 ºC for 2 minutes and 1µL of Superscript™ II RT (200 U) was added and

mixed by pipette. The mixture was incubated in a c1000 Touch™ thermal cycler (Bio-Rad) at 25 ºC for 10 minutes, 42 ºC for 50 minutes and 70 ºC for 15 minutes. The resulting cDNA can be used as a template in a conventional PCR reaction.

2.5 Polymerase chain reaction (PCR)

All PCR was carried out in a PCR Workstation™ (C.B.S. Scientific) to minimise the risk of contamination and the same C1000 Touch™ thermal cycler (Bio-Rad) was used to avoid potential variation in amplification by different instruments.

2.5.1 Bacterial 16S rRNA amplification

PCR amplification of the V3 region of the bacterial 16S rRNA gene was performed to analyse the bacterial community. The primers used were V3FC (5'- CGC CCG CCG CGC GCG GCG GGC GGG GCG GGG GCA CGG GGG GCC TAC GGG AGG CAG CAG -3') and V3R (5'- ATT ACC GCG GCT GCT GG -3') (Muyzer *et al.* 1993). In order to run the PCR product on DGGE the forward primer contains a 40 bp GC-clamp attached to the 5' end. The reaction was performed with 0.5 mM each primer 1x EX-Taq buffer, 0.3 mM each dNTP, 1 mM of $MgCl_2$, 500 mg BSA, 1.25 U Ex-Taq (Takara) and 1 µL of gDNA or cDNA template made up to 50 µL with sterile 18.2 MΩ H_20. The cycling conditions used were an initial denaturation for 5 min at 95 ˚C followed by 20 cycles of 95 °C for 1 min, 65 °C (-0.5 °C per cycle) for 1 min and 72 °C for 30 s followed by 15 cycles of 95 °C for 1 min, 55 °C for 1 min and 72 °C for 3 min with a final extension at 72 °C for 30 min. The 30 minute final extension step was included in all PCR cycles that were to be analysed by DGGE to prevent visualisation of spurious

double bands which would hinder accurate analysis of the gel images and make excision of bands difficult (Janse *et al.* 2004).

2.5.2 Fungal 28S rRNA amplification

PCR amplification of the fungal community was amplified using PCR primers specific for the 28S rRNA region of the fungal genome. The primers used were U1 (5'- GTG AAA TTG TTG AAA GGG AA -3') and U2C (5'- CGC CCG CCG CGC GCG GCG GGC GGG GCG GGG GCA CGG GGG GGA CTC CTT GGT CCG TGT T -3') (Sandhu *et al.* 1995). The reaction was performed with 0.5 µM each primer 1x Ex Taq buffer, 0.3 mM each dNTP, 1 mM of $MgCl_2$, 500 mg BSA, 1.25 U Ex Taq (Takara) and 1 µL of cDNA template, the reaction was made up to 50 µL with sterile 18.2 MΩ H_20. The cycling conditions used were an initial denaturation for 5 min at 95 ˚C followed by 10 cycles of 95 °C for 1 min, 60 °C (-1 °C per cycle) for 1 min and 72 °C for 30s followed by 25 cycles of 95 °C for 1 min, 50 °C for 1 min and 72 °C for 30 s with a final extension at 72 °C for 30 min (Janse *et al.* 2004).

2.5.3 Confirmation of successful PCR

A 1% (w/v) agarose gel was prepared to confirm that PCR had been successful by electrophoresis of the PCR product. Electrophoresis was carried out as described previously (2.3.4) according to the protocol for DNA. Successful PCR was confirmed by the presence of PCR product of correct size by comparison with fragments from Hyperladder 1 and the absence of a band in the negative control lane.

2.6 Denaturing gradient gel electrophoresis (DGGE)

2.6.1 Culture of common stool isolates for ladder

Common bacterial isolates from preterm infants were cultured in order to construct a "ladder" of known organisms that could be loaded onto a DGGE gel along with samples to allow alignment across multiple gels (Tourlomousis & Kemsley 2010). These species used in the ladder are shown in Table 2.1. Cultured isolates were obtained from fully identified wild type strains from the Freeman hospital. DNA was extracted using the DNeasy Blood and Tissue Kit and amplified as described above.

Table 2.1 - Organisms used to generate DGGE ladder

Bacterial species[a]	Fungal species[b]
Pseudomonas aeruginosa	*Candida parapsilosis*
Serratia marcescens	*Candida dubliniensis*
Citrobacter freundii	*Candida albicans*
Staphylococcus epidermidis	*Candida glabrata*
Clostridium perfringens	*Aspergillus fumigatus*
Bacteroides fragilis	*Exophiala dermatitidis*
Acinetobacter baumannii	*Scedosporium apiospermum*

[a]Ladder used only for alignment of multiple gels. Sequencing of bands was carried out for identification

[b]Ladder used for alignment and identification

2.6.2 Gel preparation

The DCode™ Universal Mutation Detection System (Bio-Rad) was used to cast and run the DGGE gels. Glass plates were cleaned prior to use using 100% ethanol before use to remove any residue and static that may interfere with the gel casting. A thin layer of silicon grease was applied to the 1mm spacers and the sponge in the casting stand to prevent leakage of the denaturant during casting and running (Brinkhoff & Hannen 2001). With the spacers placed on either side of the vertical edges between the large (20 cm x 16 cm) and small (16 cm x 16 cm) glass plates, the alignment card was inserted between the glass plates and the sandwich clamps were tightened to each side of the glass plates. The alignment card was then removed and the gel cast was securely fixed in the casting stand ensuring the base of the glass plates was submerged in the sponge. For optimal results, only the front gel was ran and therefore a balance plate was setup as above without the spacers to prevent the buffer leaking from the upper chamber during electrophoresis.

For the casting of the gel, a 19 gauge needle fitted with a tube and Y-fitting was attached to the centre of the plates. Two plastic 50 mL tubes, labelled high and low, were stood in a rack and the denaturant solutions described below were added to the tubes, the lids secured, and inverted to mix. Bacterial DGGE analysis utilised a denaturant gradient of 34 – 55% denaturant, whereas fungal DGGE employed a denaturant gradient of 40 – 60% denaturant (with 100% denaturant corresponding to 7 mol l^{-1} urea plus 40% v/v formamide). In all cases 12% polyacrylamide gels were used.

Reagent	High	Low
DCODE dye (Appendix 3)	100 µL	0 µL
Denaturing solution (Appendix 4)	25 mL	25 mL
APS (10% w/v)	216 µL	216 µL
TEMED	21.6 µL	21.6 µL

The high and low denaturing solutions were each drawn up into a separate 25 mL syringe with rubber tubing. Each syringe was then attached to the Y-fitting before being secured in a Model 475 gradient former (Bio-Rad) and the cam wheel turned to dispense the solutions into the gel cast. Once the solutions had been dispensed into the cast the needle was removed and a well comb was placed the two glass plates, ensuring no bubbles existed in the gel. The gel was left for a minimum of 1.5 hours to polymerise.

2.6.3 Running conditions

To make up the 1× TAE buffer used in the DGGE, 140 mL of 50× TAE (Appendix 2) in 7 L of distilled water was mixed in the electrophoresis tank. The control module was placed on the electrophoresis tank, the thermostat was set to heat the buffer to 60 ºC and the stirrer was switched on. Following polymerisation, the gel was removed from the casting stand and attached to the core in the front orientation and the balance plate was attached to the core at the rear. Once the buffer had come to temperature, the control module was removed and the core was lowered into the buffer chamber in the correct orientation (red mark on the right hand side) and then the upper chamber was filled with the heated buffer. A 15 µL aliquot of sample was mixed with an equal volume of 2x

DGGE loading buffer (Appendix 3) and loaded onto gel. The control module was replaced and once the temperature had returned to 60 ºC the power pack was set to 200 V for 4.5 hours for bacterial community analysis or 70 V for 17 hours for fungal community analysis.

2.6.4 Staining

The core was removed from the buffer tank and the cast disassembled. A 25 µL aliquot of SYBR gold (Invitrogen, 10,000x) was diluted in 250 mL of 1x TAE in a staining container. The gel was removed from the glass plates and submerged in the staining solution in the dark for 30 minutes at room temperate with gentle agitation. The gel was de-stained in dH_2O to remove any excess stain and viewed under U.V. light using the Gel Doc 2000 gel documentation system (Bio-Rad) and quantity one™ software (v4.1.1.). Print outs were obtained using Mitsubishi Video copy processor (Model P91).

2.7 Sequencing of excised DGGE bands

2.7.1 Gel storage

Following visualisation, gels were stored for retrospective excision of bands to obtain sequence information. Excess moisture was removed before the gel was placed between two A4 sheets of acetate, which was then wrapped in cling film. The gel was then placed in a labelled A4 plastic sleeve and stored in a folder at -80 °C.

2.7.2 Band excision

Bands of interest were excised from either stored or freshly re-ran DGGE gels. The desired stored gel was removed from the -80 °C freezer and carefully transferred from the protective acetate to the Gel Doc platform before it defrosted. Fresh gels were simply placed on the Gel Doc platform following staining. Whilst exposed to U.V. light, bands of interest were excised from the gels using a sterile blue (1 mL) pipette tip and placed in a sterile 1.5 mL microfuge tube. The bands were immersed in 10 µL of 18.2 MΩ H_20 and vortexed for 1 minute. The tubes were stored at 4 ºC overnight to allow DNA to elute from the gel.

2.7.3 Amplification and clean-up of excised band

Following overnight incubation, the tubes containing the excised band were vortexed and centrifuged before the full 10 µL aliquot of the eluate provided a template for PCR. The PCR was carried out as above only the primer containing the 40 bp GC-clamp was replaced by the unclamped primer. Once confirmation of successful PCR was carried out, as above, the PCR product was subject to ExoSAP-IT PCR clean-up (Affymetrix). Briefly, 5 µL of PCR product was mixed with 2 µL of ExoSAP-IT and incubated in a c1000 Touch™ thermal cycler (Bio-Rad) at 37 °C for 15 minutes to degrade unused primers and nucleotides. The reaction was then heated to 80 °C for 15 minutes to inactivate the ExoSAP-IT.

2.7.4 Preparation of *E. coli* chemically competent cells

A sterile 30 mL glass universal containing 10 mL LB media was inoculated with a single fresh colony of TOP10 *E. coli* cells (Invitogen). This culture was incubated for

16 h at 37°C with orbital shaking at 200 rpm for 2 – 3 hours. When the OD600 reached 0.35 – 0.4 nm, the cells were then transferred aseptically into two sterile ice cold universals and incubated on ice for 10 minutes. The cells were then centrifuged for 10 min at 2,700 x g, 4 °C. The supernatant was discarded and the each of the pelleted cells were resuspended in 7.5 mL sterile ice cold MgCl2 by gentle mixing. The cells were pelleted again by centrifugation for 10 min at 2,700 x g, 4°C. The supernatant was discarded and each of the cell pellets were finally resuspended in 0.5 mL of sterilised ice cold CaCl2 and incubated on ice for at least 1.5 h to become competent. For long term storage the appropriate volume of sterile 50% (v/v) glycerol was added to give a final concentration of 15% (v/v). Since transformation required the cells at 50 µL per reaction, 65 µL of cells (in 50% (v/v) glycerol at a final concentration of 15% (v/v)) was aliquoted into individual sterilised 1.5 mL microcentrifuge tubes and stored at -80°C for future use. The full 65 µL aliquot was used per transformation reaction.

2.7.5 Ligation

The ExoSAP-IT treated PCR product was cloned using the pGEM-T® Easy Vector and the 2X Rapid Ligation Buffer (Promega). Tubes were centrifuged to collect the contents at the bottom. Ligation reactions were performed as set out according to Table 2.2. The ligation reaction was mixed by pipetting and incubated overnight at 4 °C to for maximum number of transformants.

Table 2.2 – Ligation reaction

Reaction Component	Standard Reaction	Positive Control
2X Rapid Ligation Buffer,	5 μL	5 μL
pGEM®-T Easy Vector (50 ng)	1 μL	1 μL
PCR product	3 μL	-
Control Insert DNA	-	2 μL
T4 DNA Ligase (3 Weiss units/μL)	1 μL	1 μL
18.2 MΩ H_20	-	1 μL

2.7.6 Transformation

Aliquots of previously prepared TOP10 competent cells were removed from storage (-80 °C) and thawed on ice for approximately 5 minutes. Meanwhile, 2 µL of ligation reaction were added to a sterile 1.5 mL microfuge tube on ice. The 65 µL aliquot of TOP10 competent cells were carefully added to the ligation reaction and mixed by gentle flicking. The tubes were incubated on ice for 20 minutes and then heat-shocked at 42 °C for 50 seconds and immediately returned to ice for a further 2 minutes. 950 µL of room temperature SOC medium (Appendix 5) was then added and the tubes were incubated for 1.5 hours at 37 °C in an orbital incubator with shaking at 150 rpm. Duplicate LB/ampicillin/IPTG/X-Gal plates (Appendix 6) were warmed in an incubator at 37 °C for 30 minutes prior to spreading a lawn from 100 µL of the ligation reaction onto each of the two plates which were incubated at 37 °C overnight (16-24 hours).

2.7.7 Inoculation and confirmation of successful incorporation of insert

White colonies were used as a marker for successful incorporation of the PCR product into the plasmid. The majority of a white colony was inoculated into an LB/ampicillin broth (Appendix 6) which was incubated at 37 °C overnight. PCR amplification was performed to confirm successful incorporation of the insert. The primers used were M13 Forward (5'- CGC CAG GGT TTT CCC AGT CAC GAC -3') and M13 Reverse (5'- TCA CAC AGG AAA CAG CTA TGA C -3'). An initial PCR was set up containing 0.5 mM each primer 1x EX-Taq buffer, 0.3 mM each dNTP, 1 mM of $MgCl_2$, 500 mg BSA, made up to 49.75 µL with sterile 18.2 MΩ H_20 in the absence of template DNA and 1.25 U Ex-Taq (Takara). Using a sterile pipette tip, the remainder of the white colony was transferred to the PCR reaction before the Ex-Taq was added and the PCR was initiated. The cycling conditions used were an initial denaturation for 5 min at 95 ˚C

followed by 35 cycles of 95 °C for 30 s, 65 °C for 30 s and 72 °C for 3 min and 72 °C for 3 min with a final extension at 72 °C for 10 min. An agarose gel was performed as previously described, successful visualisation of a band corresponding to the correct fragment length was used to confirm successful incorporation of the insert.

2.7.8 Plasmid DNA purification (PureYield™ Plasmid Miniprep kit)

Plasmid DNA was purified using the PureYield™ Plasmid Miniprep kit (Promega) alternative protocol for larger culture volumes. Prior to beginning the experiment the cell lysis buffer was warmed to 37 ˚C and inverted to dissolve any precipitate. Briefly, 1.5 mL of the LB/ampicillin overnight bacterial culture was added to a 2 mL microfuge tube and centrifuged at 13,000 × g for 30 seconds and the supernatant discarded, this step was repeated once to process a total volume of 3 mL. The cell pellet was resuspended in 600 µL of sterile 18.2 MΩ H_20 and 100 µL of cell lysis buffer was added before the tube was inverted 6 times until the solution changed from opaque to clear blue. The cell lysis buffer contains SDS to lyse cell membrane and sodium hydroxide which breaks down the cell wall but also causes DNA to linearise by breaking hydrogen bonds. Within 2 minutes, 350 µL of cold neutralization solution was added and the solution was mixed by inversion causing a precipitate to form. The neutralisation solution contains potassium acetate which reduces the alkalinity of the solution causing renaturation of the plasmid DNA but leaving gDNA linearised allowing it to be separated from the plasmid DNA by centrifugation. The solution was centrifuged at 13,000 × g for 3 minutes to pellet the precipitate and the supernatant (~900 µL) was transferred to a PureYield™ Mini-column. The column was centrifuged at 13,000 × g for 15 seconds and the flow through discarded. 200 µL of Endotoxin removal wash was added to the column and centrifuged for 15 seconds, followed by addition of 400 µL of

column wash solution to the column and a 30 second centrifugation at 13,000 × g. The column was transferred to a 1.5 mL microcentrifuge tube and 30 µL of elution buffer was added to the column membrane which was incubated at room temperature for 1 minute. The column was centrifuged at 13,000 × g for 15 seconds and the eluted DNA was stored at -20 ˚C.

2.7.9 DNA quantification (NanoDrop 1000)

For custom DNA sequencing (Eurofins MWG Operon), plasmid DNA is required in a concentration of 50 - 100 ng/µL in a total volume of 15 µL. To quantify the plasmid DNA a NanoDrop 1000 Spectrophotometer (Thermo Scientific) was used. Briefly, 1 µL of dH_2O was applied to the pedestal, followed by 1 µL of the PureYield™ Plasmid Miniprep kit elution buffer to calibrate the instrument. The purified plasmid DNA was mixed and 1 µL was applied to the pedestal and readings were measured in triplicate. An average of the triplicates was calculated and the concentration adjusted as appropriate for sequencing. Samples were either diluted in sterile 18.2 MΩ H_20 or concentrated using a RVC 2-18 rotational vacuum concentrator (Christ) at 60 °C.

2.7.10 Sequencing reaction

Sequencing was carried out commercially by Eurofins MWG Operon using the Value Read service based on Sanger sequencing methods (Sanger *et al.* 1977). Sequencing was performed using the ABI 3730 XL capillary sequencer with BigDye v.3.1 dye-terminator chemistry as per manufacturer's instructions (Applied Biosystems). Briefly, the M13 reverse primer bound to the complementary DNA strand was extended linearly until by chance the corresponding dideoxy terminator nucleotide was incorporated and

thus the extension of the fragment was terminated. Each of the four dideoxy terminators was tagged with a different fluorescent dye which fluoresced upon illumination at specific wavelengths and produced a chromatogram from which sequences were deduced.

2.7.11 Database mining for homologous sequences (NCBI Blastn)

The NCBI Vec Screen tool was employed to identify segments of a nucleic acid sequence that may be of vector origin, these sequences were subsequently removed. The remaining sequence was then input into the Nucleotide Basic Local Alignment Search Tool (BLASTn) in FASTA format and the nucleotide collection was used for comparison.

2.8 Quantitative PCR (qPCR)

2.8.1 Preparation of standard curve

In order to obtain a control strain to prepare the qPCR standard curve, a *Lactobacillus* sp. was isolated by culture from patient 13 (sample 340082K) and DNA was extracted using the DNeasy Blood and Tissue Kit as described above. Different primer sets which both encompass the V3 region of the 16S rRNA gene were used to quantify total bacterial load and genus specific Lactobacilli load (Table 2.3). To generate the standard for the total bacterial analysis, conventional end point PCR was carried out on the extracted DNA using Buffer I (NEB), 0.5 μM each primer, 0.2 mM each dNTP, 1 μL DNA template made up to 49.5 μL with sterile 18.2 MΩ H_2O. After an initial denaturation step of 95 °C for 5 minutes the thermocycler was held at 80 °C while 2.5

units of Taq DNA polymerase (NEB) was added to each reaction giving a total volume of 50 μL per reaction. PCR was resumed with 35 cycles of 95 °C for 30 s, 65 °C for 30 s, and 72 °C for 30 s, followed by a final extension step of 72 °C for 10 min. To generate the standard for the *Lactobacillus* genus specific analysis, conventional end point PCR was carried out on the extracted DNA using Buffer I (NEB), 0.5 μM each primer, 0.2 mM each dNTP, 1.5 mM $MgCl_2$, 1 μL DNA template made up to 49.5 μL with sterile 18.2 MΩ H_2O. After an initial denaturation step of 94 °C for 5 minutes the thermocycler was held at 80 °C while 2.5 units of Taq DNA polymerase (NEB) was added to each reaction giving a total volume of 50 μL per reaction. PCR was resumed with 35 cycles of 95 °C for 15 s, 62 °C for 1 min, and 72 °C for 1 min, followed by a final extension step of 72 °C for 10 min (Tamrakar *et al.* 2007). Successful PCR was confirmed by 1% agarose gel electrophoresis as described above. The PCR products were cleaned to remove unincorporated dNTPs and primers using Exo-SAP-IT and cloned using the p-GEMT easy vector cloning kit before purification of the plasmid using the PureYield™ Plasmid Miniprep kit as described above. The purified plasmids were then pooled into a single microcentrifuge tube and the plasmid DNA was quantified using the NanoDrop 1000.

To achieve absolute quantification of the total or genus specific bacterial load a standard curve with 10 fold dilutions was generated. This requires the calculation of plasmid DNA which is needed to generate the standards. Firstly, the size of the plasmid DNA in base pairs (bp) was calculated based on the combined length of the PCR amplicon and the plasmid (p-GEMT easy vector is 3015 bp). The mass of the plasmid was then calculated by multiplying the size of the plasmid (bp) by the average weight of one bp which is 1.096×10^{-21} (g/bp). The average weight of one bp was calculated by dividing the average molecular weight of a double stranded DNA molecule (660 g) by Avagadro's number (6.023×10^{-23}). The mass of plasmid containing the copy number

needed to achieve a suitable standard curve was then calculated based on a standard curve quantifying between 3×10^6 and 30 copies/g in 10-fold dilutions. Therefore, to calculate what mass of plasmid will contain the copy number of interest, the previously calculated mass of the plasmid was multiplied by the desired copy number for the standard curve. The figure produced by this calculation could then be divided by the required by the volume in the final PCR reaction mix to give the final concentration of the plasmid DNA (g/µL). It was then possible to prepare a standard curve by rearranging the formula $M_1V_1 = M_2V_2$ to $V_1 = M_2V_2 / M_1$. Where V1 is the final volume required to achieve the concentration (unknown), M2 is the concentration of the stock plasmid, V2 is the volume in which the plasmid will be diluted, and M1 is the final concentration of the plasmid which was calculated in the previous step.

Table 2.3 – Primers used in qPCR assay

Primer	Sequence (5'-3')	Target	Reference
Eub 338	ACTCCTACGGGAGGCAGCAG	All bacteria	J Lane, 1991
Eub 518	ATTACCGCGGCTGCTGG	All bacteria	Muyzer *et al.*, 1993
Lacto F	TGGAAACAGRTGCTAATACCG	All lactobacilli	Byun *et al.*, 2004
Lacto R	GTCCATTGTGGAAGATTCCC	All lactobacilli	Byun *et al.*, 2004

2.8.2 Reaction setup

The reaction mix was prepared in a PCR Workstation and pipette filter tips (Fisher) and pipettes designated solely for qPCR setup were used to limit the potential for contamination. No template controls (NTCs) were included in every reaction to check for contamination of all the reagents, plastics and pipettes used in preparing the reaction. Each reaction was performed in triplicate to ensure accuracy. The qPCR protocol was based on the previously described method (Baxter & Cummings 2008). The unknown DNA samples were firstly diluted 1:20. Dilutions were performed in microcentrifuge tubes using sterile 18.2 MΩ PCR grade water as the diluent. These diluted unknown samples and an aliquot of the plasmid DNA stock was heated at 95 °C for 10 minutes to ensure any tertiary structures which may have formed had been eliminated so that efficient amplification was possible. The plasmid DNA was then used to prepare the serial dilutions as calculated previously. The unknown samples (diluted 1:20) were further diluted 1:5 and the equivalent volume of sterile 18.2 MΩ water was used as the NTC. Once the dilutions of the standard curve and unknown samples was complete and the NTC was prepared, the reaction mix was setup ensuring that ABsolute™ QPCR SYBR® mix (Thermo Scientific) had limited exposure to light to prevent degradation. The final reaction mix contained 1x ABsolute™ QPCR SYBR® Green Mix (Thermo-Start™ DNA Polymerase, 3 mM MgCl2) , 0.35 mM each primer, 12.5 µg BSA and 5 µl of DNA in a final volume of 25 µL.

qPCR was carried out on the RotorGene RG-3000 instrumentation (Corbett life sciences). For total bacterial analysis, the cycling conditions used were an initial enzyme activation step at 95 °C for 15 min, then 50 cycles of 95 °C 10 seconds, 65°C for 15 seconds and extension at 72 °C for 20 seconds. For Lactobacilli genera specific analysis, the cycling conditions used were an initial enzyme activation step at 95 °C for 15 min, then 50 cycles of 95 °C 15 seconds, 62°C for 1 minute and extension at 72 °C

for 20 seconds. Target copy numbers for each reaction were calculated from the standard curve and were used to ascertain the number of copies per g of stool then log transformed. Standard deviation was determined (by the Rotor-Gene 6 software) on the replicate threshold cycle (*CT*) value and reactions repeated if the deviation was above 0.4. Samples were considered to be below reasonable limits of detection if the *CT* value was above 30 cycles (Karlen *et al.* 2007).

2.9 Next generation sequencing

Data for the initial technique comparison study (chapter 3) and twin study (Chapter 6) was generated using the 454 pyrosequencing (Roche) platform. Data for the large disease matched control study (chapter 7) was generated using the benchtop MiSeq (Illumina) platform.

2.9.1 454 Pyrosequencing

454 pyrosequencing was carried out commercially by the Research and Testing Laboratory (RTL; USA) using shipped extracted DNA samples. The DNA sequencing was performed on the 454 GS FLX Titanium Sequencing System (Roche) using previously described methods (Dowd *et al.* 2008). Two different primer sets were used in the 454 pyrosequencing. Firstly, for the initial technique comparison (Chapter 3), a portion of 16S RNA (position 341 to 907; *E. coli* numbering) was amplified using the primer set 341F (5'-CCTACGGGAGGCAGCAG-3') (Muyzer *et al.* 1993) and 907R (5'-CCGTCAATTCCTTTGAGTTT-3') (Muyzer *et al.* 1995). For the twin study (Chapter 6), 454 pyrosequencing was carried out using the bifidobacteria-optimised

primer set (position 357 to 926; *E. coli* numbering) 357F (5'-CTCCTACGGGAGGCAGCAGAN-3') and 926Rb (5'-CCGTCAATTYMTTTRAGT-3') (Sim *et al.* 2012).

2.9.2 MiSeq

The 16S sequencing on the MiSeq platform was carried out in house based on the 'Schloss wet-lab MiSeq SOP' (available at - http://www.mothur.org/wiki/MiSeq_SOP). Paired end indexed reads were used to multiplex 192 samples per run (188 samples, 2 positive, and 2 negative controls). Briefly, extracted DNA from the samples was aliquot into 96 well plates and the last two wells left empty for controls. In a new 96 well plate, 17 µL of Accuprime Pfx Supermix (Life Technologies Ltd.) was dispensed into each well, before 1 µL of the DNA template and 2 µL of each paired set of index primers was transferred to the corresponding well. 1 µL of PCR grade dH_2O was added to the negative control well and 1 µL of a mock community was added to the positive control well on each plate. Plates were vortexed briefly and span down before being placed in the thermocycler (BioRad CFX96 Touch). PCR was carried out using the following cycling conditions; initial step at 95 °C for 2 min, then 30 cycles of 95 °C 20 seconds, 55°C for 15 seconds, extension at 72 °C for 5 min, and a final extension at 72 °C for 10 min. Successful PCR was confirmed using an agarose gel (1%) on a subset of 12 samples per plate.

Normalisation was performed by transferring 18 µL of PCR product to the corresponding well on a normalisation plate. 18 µL of binding buffer was then transferred and the contents mixed by pipetting and vortexing before being spun down. Plates were incubated at room temperature for 1 hour. Being careful not to touch the sides of the wells, liquid was removed and 50 µL wash buffer was added and briefly

mixed by pipetting and removed immediately removed leaving no residue. 20 µL of elution buffer was then added and mixed by pipetting and vortex before being spun down. Following incubation at room temperate for 5 minutes 5 µL from each well was pooled and the plates frozen for later use.

Library quality control was carried out using a Bioanalyser (Agilent bioanalyser 2100) and qPCR using KAPA SYBR FAST qPCR Master Mix (Anachem Ltd.). The pooled library underwent serial dilutions to generate the following dilutions; 1:1, 1:10, 1:1000, 1:2000, and 1:4000. For the Bioanalyser, the gel dye mix, ladder, and 1 µL of the 1:1 and 1:10 dilutions were loaded into the necessary wells of a high sensitivity chip. For Kapa qPCR library quantification reactions were carried out in a 10 µL reaction volume with 6 µL of master mix and 4 µL of standards and library dilutions, in triplicate. qPCR was carried out using the following cycling conditions; initial activation step at 95 °C for 5 min, then 35 cycles of denaturation at 95 °C for 30 seconds and annealing at 60°C for 45 seconds. From these results a further dilution was carried out on the median pool dilution amongst the standards, as all pools were normalised to the lowest dilution pool selected.

For the sequencing a 500 cycle reagent cartridge and all reagents and samples were thawed prior to setup. 3.4 µL of read 1 sequencing primers was placed in well 12, 3.4 µL of the index primer was placed into well 13, and 3.4 µL of read 2 sequencing primers was placed in well 14. Sampled were prepared by mixing 10 µL of library and 10 µL of 0.2 NaOH and the PhiX spike was prepared by mixing 2 µL of PhiX, 3 µL PCR grade H_2O, and 5 µL of NaOH. Following 5 min incubation, samples and PhiX were made up to 1 ml with HT1 and then HT1 was used to dilute the library and PhiX to 10 pM. A 5% PhiX run was used so 950 µL of 3.5 pM library and 50 µL PhiX were mixed in a tube and 600 µL of this mixture was loaded into well 17. The flow cell was rinsed with Milli-Q water, wiped with 80% ethanol, and carefully dried prior to

placement within the MiSeq instrument. The cartridge, flow cell, and PR2 bottle were then loaded and on the on screen instructions followed.

2.9.3 Processing the raw sequencing reads

Raw data generated using next generation sequencing technology was bioinformatically processed to reduce the error rate by trimming the primer and barcode sequences and removing reads with a poor quality score. Two pipelines, Mothur and quantitative insights into microbial ecology (QIIME), were utilised in the processing of raw sequence reads.

2.9.2.1 Mothur

For the initial technique comparison study (chapter 3) the raw 454 pyrosequencing .sff files were filtered using Mothur version 1.22 (Schloss 2009). The Schloss lab standard operating procedure (SOP) for 454 data was followed with the following criteria: 1) maximum of 1 mismatch to barcode tags; 2) no ambiguous bases; 3) maximum of 2 mismatches to primer sequence; 4) average quality score of >35 in a sliding window of 100 bp. Detection of potentially chimeric sequences was performed using Chimera.uchime and chimeric sequences were removed from downstream analysis. Following filtering of reads, taxonomic classification information was generated via the Silva database (Schloss *et al.* 2011).

For the large disease control matched study (chapter 7) the fastq files generated were processed using Mothur version 1.31.2 (Schloss 2009). The Schloss MiSeq SOP was followed according to the following criteria: 1) no ambiguous bases; 2) maximum length of 275 bp; 3) maximum of 8 homopolymers; 4) within 2 mismatches of the

sequence being considered. Detection of potentially chimeric sequences was performed using Chimera.uchime and chimeric sequences were removed from downstream analysis. Alignment was generated via the Silva database (Schloss *et al.* 2011). A cutoff of 70 was applied to assign sequences to the trainset9_032012. In total 44,515,418 reads passed processing and were included in the subsequent analysis for Chapter 7.

2.9.2.2 QIIME

For the twin study (Chapter 6), the raw sequencing reads were quality filtered in QIIME (version 1.6.0) (Caporaso *et al.* 2010) using the split-library.py script with the following criteria: 1) exact matches to barcode tags; 2) no ambiguous bases; 3) maximum of 5 mismatches to primer sequence; 4) read-lengths between 200-700 base pairs (bp); 5) average quality score of >25 in a sliding window of 50 bp. Remaining high quality sequences were clustered into operational taxonomic units (OTUs) at 97% similarity using UCLUST (Edgar 2010). Representative sequences for each OTU were aligned using PyNAST (Caporaso *et al.* 2010) and taxonomic identities were assigned using RDP-classifier (version 2.2) (Wang *et al.* 2007) with 50% as confidence value threshold. Detection of potentially chimeric sequences was performed using ChimeraSlayer (Haas *et al.* 2011) and chimeric sequences were removed from downstream analysis prior to tree building using FastTree (Price *et al.* 2010). Sequences were deposited in MG-RAST under the accession numbers 4516545.3 - 4516585.3 (twin study: Chapter 6).

2.10 Analysis of data

2.10.1 DGGE gel processing (TotalLab Phoretix 1D)

Images captured using quantity one™ software (v4.1.1.) were exported in tagged image file format (.TIFF) at 276 dots per inch (DPI) into a new experiment folder. The lanes of the gels were created using the automatic lane creation function and the frames manually adjusted so that all bands were central in the lanes. The background from the lanes was subtracted using a rolling ball method with a radius of 100 pixels. The band detection setting was applied to the gels using a minimum slope of 100, noise reduction of 4, and a %max peak of 2, then bands were added/removed and the band width adjusted manually. As stated previously, a ladder of known organisms was loaded to the outside and middle lanes on each DGGE gel in the experiment so that successful gel alignment could be achieved (Tourlomousis & Kemsley 2010). Bands from the DGGE ladder were assigned standard retention factor (Rf) values based upon the distance they had migrated through the gel, each corresponding band from all gels in the experiment was given the same value. A minimum of five bands per standard has been suggested for accurate interpolation of multiple gels which was exceeded for both the bacterial (6 bands) and fungal (7 bands) analysis (Tourlomousis & Kemsley 2010).

A new database was created in Phoretix 1D Pro based on the analysed gels. The successful alignment of the gels was confirmed by matching all lanes in the database then creating a dendrogram based on the Dice coefficient. Alignment of multiple gels was deemed successful if the ladders clustered together (Appendix 6). An OTU matrix was then created by the software using the Rf values of all bands and the relative intensities, which was exported to Microsoft Excel 2010 to be used for further statistical analysis.

2.10.2 NGS basic data analysis

2.10.2.1 Conversion of biom table

The convert_biom.py command was used to convert the biom formatted OTU table to the classic OTU matix. This command creates a tab delimitated text file which was opened in Microsoft Excel 2010 and saved in the Excel workbook format for downstream data analysis.

2.10.2.2 Rarefaction

The command was used to generate the rarefaction curves based on the 'Observed Species' metric. This is the count of unique OTUs found in each sample. The rarefaction curves were plotted in Microsoft Excel 2010. The number of sequences, sample coverage, and number of OTUs was calculated for each sample using the command in Mothur.

2.10.2.3 Bar chart

The command was used to group the OTU sequences by taxonomic assignment. The html file was used to determine which taxonomic level (from Kingdom to Genus) was the most informative and this was then plotted in Microsoft Excel 2010.

2.10.2.4 Core Microbiome

The compute_core_microbiome.py was used to generate text files ranging from 50% to 100% (at 5% increments), to show which OTUs were present in the particular percentage across all samples. A cut off a less than 85% was implemented (Seekatz *et al.* 2013).

2.10.2.5 Phylogenetic trees (MEGAN)

Phylogenetic trees to explore the distribution of OTUs were generated using the MEtaGenome Analyser (MEGAN) (version 4.70.4). Biom tables generated during the raw sequencing processing were used as the input files.

2.10.3 Diversity indices

Diversity indices were calculated for data generated using all techniques to examine the structure of the microbial communities present in the preterm gut using PAST (Hammer & Harper 2001). Species richness (R) was calculated based on the total number of different species present per sample. For culture this was based on the number of different isolates, DGGE was based on the total number of distinct bands, and NGS was based on the number of unique OTUs. For DGGE and NGS, where abundance information was also obtained, Shannon diversity and evenness were also calculated. The Shannon diversity index (*H'*) was calculated using the formula;

$$H' = -\Sigma\ (p_i \log[p_i])$$

The symbol *pi* is the relative intensity of each species. The log of the relative intensity was multiplied by the relative intensity for every species in all of the samples (*pi*log[*pi*]). The sum of these values for each lane was taken and multiplied by -1 (-Σ) which gives the Shannon diversity of the sample. Species evenness (E) was calculated using by dividing *H'* by the log of R. Dominance (D) was calculated based on $D=sum((n_i/n)^2)$ where n_i is number of individuals of taxon *i*. Dominance scores range from 0 (all taxa are equally present) to 1 (one taxon dominates the community completely).

2.10.4 Similarity and distance indices

The Bray-Curtis indices (Bray & Curtis 1957) was used to compared relatedness between communities using PAST (Hammer & Harper 2001). This indices calculates a relatedness score based on the abundance of OTUs in the samples which ranges from 0 (compleltey different communities) to 1 (identical communities).

2.10.5 Multivariate analysis

OTU matrix files were generated in Microsoft Excel for data generated using all techniques. The OTU matrix for culture data contained the identity of all cultured isolates and either a 1 or 0 to represent presence or absence, respectively. For DGGE and NGS data, where abundance data was obtained, the OTU matrix contained the normalised abundance of each band (for DGGE) or OTU (for NGS). Unless otherwise stated, for the DGGE analysis, all bands were included based on the Rf value and where possible the actual identity of the band replaced the Rf value.

2.10.5.1 Canoco

Canoco version 4.5.1 (Braak 1986) was chiefly used for constrained ordination analysis and principal coordinate analysis (PCoA), with images produced using Canodraw (v4.14). Firstly, the OTU matrices were copied from the Microsoft Excel spreadsheet to the clipboard. The information from the clipboard was input into Canoco using WCanoImp (4.5.2.0) to convert the data from the species matrices in to a format compatible with Canoco.

For constrained ordination analysis, discrete patient variables were assigned positive (1) or negative (0) values and the continuous patient variables were not manipulated before input into the spreadsheet. The OTU matrices were first analysed by detrended correspondence analysis (DCA) using normalised relative intensities followed by either redundancy discriminate analysis (RDA) or canonical correspondence analysis (CCA) in combination with Monte Carlo permutation testing under full model (499 permutations) depending on the DCA axis length (<3.5 RDA; >3.5 CCA). The results from the Monte Carlo permutations were deemed statistically significant if the P-value was < 0.05. Discrete variables were assigned as nominal variables to distinguish between discrete and continuous variables, with discrete variables assigned a centroid and continuous variables represented by an arrow.

PCoA was used to explore the presence/absence relationship between the culture profiles. The Canoco formatted OTU matrix was uploaded into the PrCoord (1.0) program and the Bray-Curtis distance measure was selected. The file produced by PrCoord was then visualised using the principal component analysis (PCA) method with scaling of scores focused on inter-sample distances, no post-transformation of species scores and centring by species only. Plots were manually edited within Canodraw to aid clarity according to important variables.

2.10.5.2 SIMCA

OTU matrices generated from both DGGE and NGS data also underwent multivariate partial least squares discriminant analysis (PLS-DA) using SIMCA 13.0 (Eriksson *et al.* 2006). OUT matrices were uploaded into SIMCA and each sample in the analysis was assigned to a specific group, with the scores of the model visualised in a score plot. The loadings plot was used to determine which OTUs were associated with each of the variables based on the assigned grouping. To check that data was adhering to multivariate normalities, Hotelling's T^2 tolerance limits were calculated and set at 0.95. Plots were edited within Windows picture viewer to aid clarity according to important variables.

2.10.5.3 UniFrac

The sequence reads generated using NGS were analysed using weighted UniFrac (Hamady *et al.* 2009) and visualised using principal coordinate analysis (PCoA). The analysis was carried out within QIIME by providing the OTU table (.biom file which is similar to the OTU matrix), phylogenetic tree file (Newick format phylogenetic tree), and mapping file (text file containing the sample information). Both unweighted and weighted UniFrac plots were produced where the unweighted analysis was based on presence/absence of OTUs and weighted analysis incorporated abundance data of each OTU. Due to the important of OTU abundance, typically these plots were used for visualising the data. The plots were saved as an image file and the sample labels were manually added to the plot using the interactive HTML files within QIIME to show the label for each sample.

To determine significance between groups, weight unifrac was significance was calculated using Mothur based on the phylogenetic tree produced based on the ThetaYC measure of dissimilarity (Yue & Clayton 2005).

2.10.6 Statistics

2.10.6.1 Unpaired *t*-test

Statistical significance of DGGE bands associated with disease were analysed by unpaired *t* test using GrahpPad QuickCalc online software using the formula;

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{X_1X_2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$S_{X_1X_2} = \sqrt{\frac{(n_1 - 1)S_{X_1}^2 + (n_2 - 1)S_{X_2}^2}{n_1 + n_2 - 2}}.$$

Where x bar 1 and x bar 2 are the sample means, S_{X1X2} is the pooled sample variance, and n_1 and n_2 are the sample sizes.

2.10.6.2 Tukey's test

Tukey's test was used to compare the sequenced DGGE bands with pre and post samples from NEC and sepsis patients, compared to controls. This test is a multiple comparisons procedure, used in conjunction with one-way analysis of variance (ANOVA), to find means that are significantly different from each other. This analysis was computed in Minitab 16 (version 16.1.0) with a 95.0 confidence interval and a family error rate of 5.

2.10.6.3 P-test of significance

Martin's P-test of significance (Martin 2002) was used to determine if significant differences were occurring in the NGS data. This analysis compared the profiles of each NGS sample with every other sample in the cohort, indicating potential significance between and within patients. 100 Monte Carlo randomisations were performed.

2.12 Ethical approval

Initial collection was part of routine service and all samples were collected during the course of normal treatment. Ethical approval was obtained from the County Durham and Tees Valley Research Ethics Committee to include molecular techniques in August 2010. For all infant stool samples, prospective parental informed consent was documented at the point of donation to have the samples stored for research purpose. For EBM, informed consent was documented at the point of donation to have the samples stored for research purposes from March 2011.

3. Comparison of culture dependent and molecular techniques in elucidating the gut microbiota of preterm infants

Abstract

Aim: The gut microbiota is a highly complex ecosystem increasingly associated with disease pathophysiology. Despite revolutionary advances in traditional culture and molecular based techniques, the coherence of these techniques remains elusive. We aimed to compare the performance of these techniques in assessing the gut microbiota of preterm infants.

Methods: Faecal samples (n=17) from preterm infants (n=11) were subjected to quantitative aerobic and anaerobic culture and 16S rRNA molecular based analysis using PCR-DGGE and 454 pyrosequencing. qPCR was further used to validate the results of the techniques by exploring both the total bacterial load and the load of the Lactobacilli genus.

Results: Both culture and pyrosequencing identified *Staphylococcus*, *Enterobacter*, and *Klebsiella* as the most prevalent organisms. Despite the increased coverage by pyrosequencing over culture, the techniques elucidated comparable profiles. Although some taxa were only identified by culture, employing genus specific primers in qPCR allowed identification of these taxa.

Conclusion: Specific primers are required to examine specific taxa. Some functionally significant taxa were only identified by culture demonstrating this technique may still offer important insights in the analyses of clinical samples. To fully elucidate complex ecosystems it is paramount to implement an integrative approach to overcome limitations of any single technique.

3.1 Introduction

Culture based techniques have been continuously optimised over the last century, most notably with the implementation of both selective media and the ability to incubate cultures anaerobically (Vaz-Moreira *et al.* 2011). Nevertheless, it is estimated that only 20% of the gut microbiota can be cultivated (Eckburg *et al.* 2005). Therefore, molecular techniques have increasingly been applied to clinical investigations owing to their much greater coverage of the microbial community (Petrosino *et al.* 2009).

Molecular fingerprinting techniques include DGGE and TGGE. In bacterial community analyses, these techniques are based on the 16S rRNA gene that has distinct conserved and variable regions that serve as valuable markers of genetic diversity (Sekirov *et al.* 2010). However, such approaches are subject to PCR bias with amplification efficiency and artifacts arising due to the formation of heteroduplex and chimeric sequences (Thompson *et al.* 2002; Wang & Qian 2009). Amplicons also require isolation and sequencing to identify specific taxa within the community. Recently, high-throughput 454 pyrosequencing has been employed in clinical research. This technique offers a greater depth of coverage of the samples and allows sequences to be classified to the genus level. However, this technique is not absent of PCR bias (Schloss *et al.* 2011).

Understanding how culture and molecular approaches compare is important in characterising the microbial community in the gut of preterm infants and is essential if aetio-pathogenic factors are to be explored. There is currently a lack of studies which asses the coherence of culture dependent and molecular techniques, particularly in a clinical context. Diagnosis and intervention is currently largely directed by the results of culture based analyses in clinical microbiology laboratories. This study utilises clinical

samples to evaluate findings from each technique, while enhancing current information on the composition of the preterm infant gut microbiota.

3.2 Results

3.2.1 Patients and samples

The 11 patients had median gestation of 26wks (range 23-30wks) and birth weight of 915g (range 520g-1370g). Two infants developed confirmed NEC, two infants developed confirmed NEC and sepsis, and four infants developed sepsis (Table 3.1). To allow optimal testing of the techniques, samples between day of life 1 and 145 were analysed.

3.2.2 Species richness

A comparison of the species richness identified by each technique indicated that culture based analyses, with one exception (34a), gave the fewest number of taxa within each sample compared to the two molecular approaches, (Fig 3.1). In the majority of samples (11 of 17), DGGE detected the highest number of taxa.

3.2.3 Prevalent bacterial genera

The most common genera were compared between the three techniques: both culture and pyrosequencing identified *Staphylococcus*, *Enterobacter*, and *Klebsiella* as the most prevalent (Fig 3.2). In DGGE analyses, unlike the other techniques, *Enterococcus* spp. and *Streptococcus* spp. were the most common taxa in the samples followed by *Staphylococcus* spp. and *Enterobacter* spp. The DGGE analyses also failed to detect *Klebsiella* spp. in any sample suggesting that it was not among the most prevalent bands. Pyrosequencing and culture approaches detected this genus in 13 and 5 of the 17 samples respectively. No correlation was observed between the CFU/g from culture and the respective number of reads from pyrosequencing.

Table 3.1 – Demographic data for the clinical cohort

Patient	Sample	Gestation Age (weeks)	Birth Weight (grams)	Delivery Mode	Sex	NEC	Sepsis	DOL
13	13a	23	600	Vaginal	Male	Y	Y	34
	13b							39
	13c							48
	13d							53
	13e							60
2	2a	24	520	Caesarean	Female	N	Y	41
	2b							48
20	20	26	915	Vaginal	Male	Y	N	11
33	33	26	995	Caesarean	Male	N	Y	145
11	11	25	825	Vaginal	Male	Y	Y	6
6	6	30	1370	Caesarean	Male	N	N	6
35	35	26	570	Caesarean	Male	N	Y	36
37	37	28	1090	Vaginal	Female	N	N	1
34	34a	25	800	Vaginal	Male	N	Y	6
	34b							24
28	28	28	1250	Caesarean	Male	N	N	1
29	29	28	1180	Caesarean	Male	Y	N	1

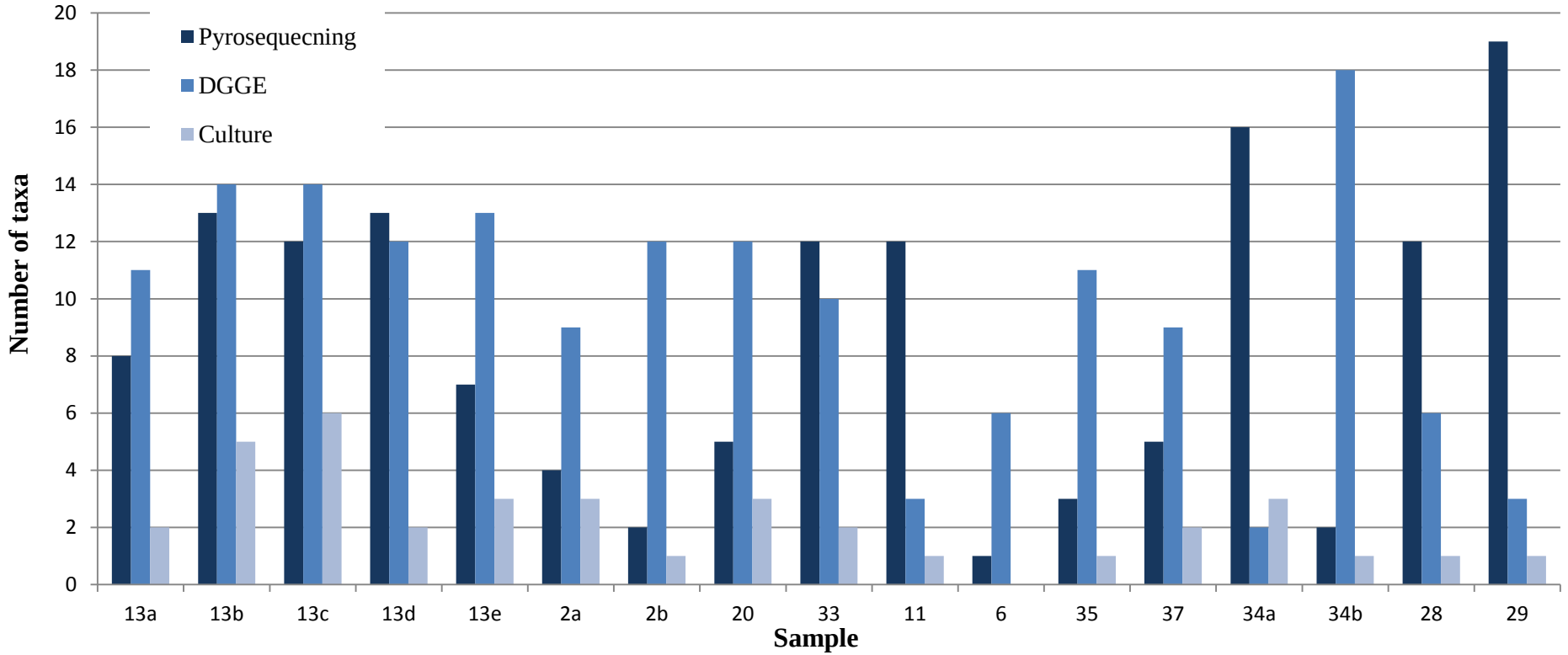


Figure 3.1 – Species richness identified by each technique independent of sequence information

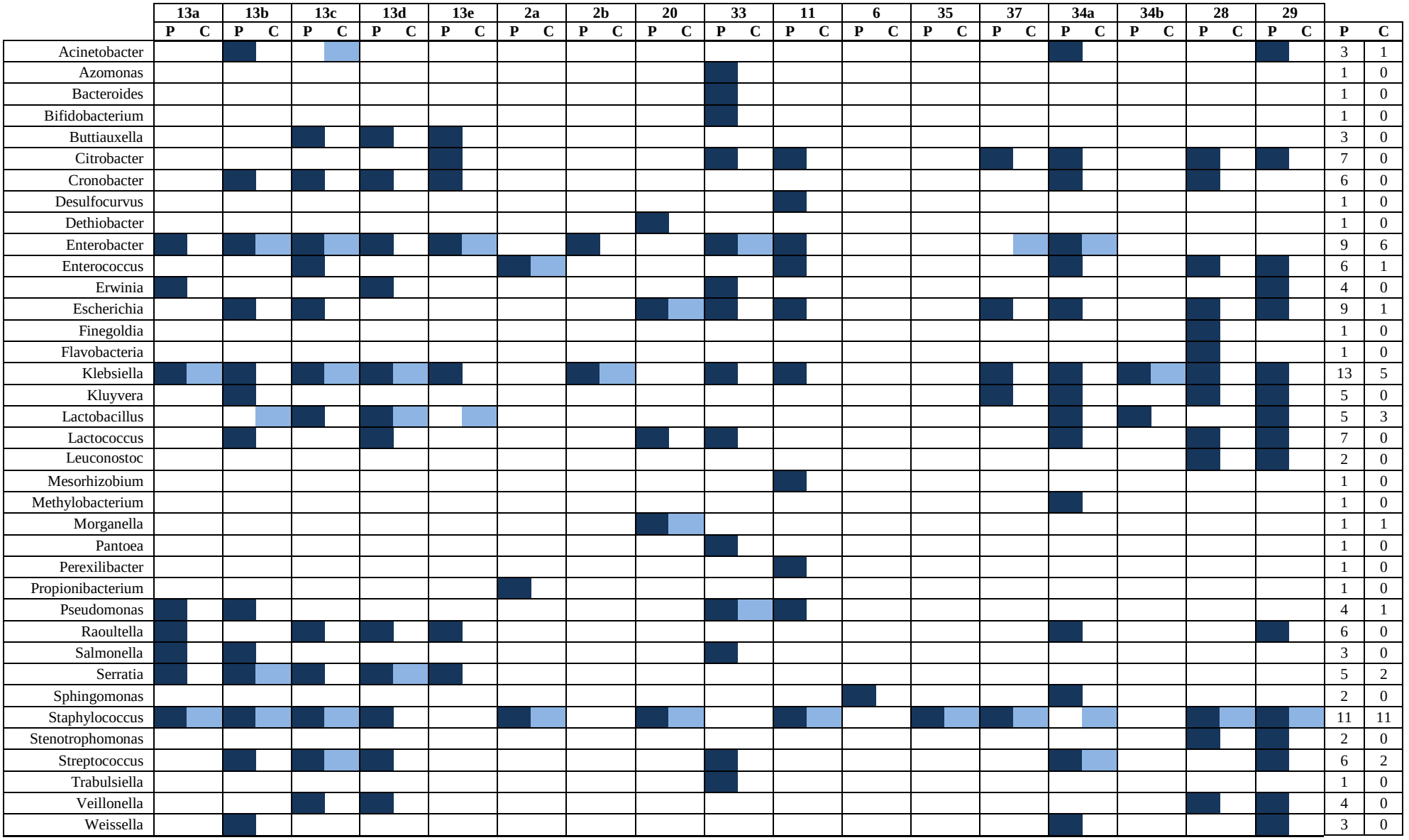


Figure 3.2 – Prevalence of individual genera detected in each sample by pyrosequencing and culture. Schematic representation showing which genus was detected in each sample per technique. Single pyrosequencing reads removed from analysis. P = Pyrosequencing, C = Culture.

3.2.4 Genera mismatches

Culture only identified 23% of the genera detected in pyrosequencing, whereas pyrosequencing identified 97% of the genera found by culture. Those genera identified only by culture within specific individual samples were *Acinteobacter*, *Enterobacter*, *Lactobacillus*, and *Staphylococcus* (Fig 3.2). Noteworthy is these genera were detected by molecular techniques in other samples, but a mismatch occurred in 4 samples (13c, 13e, 37, and 34a). To further explore the discrepancy in the detection of *Lactobacillus* spp. conventional PCR and quantitative-PCR (qPCR) using *Lactobacillus*-specific primers was carried out on longitudinal samples (13a - 13e), where culture identified lactobacillus in multiple samples not detected by molecular techniques (Table 3.2). Both conventional PCR (data not shown) and qPCR (Table 3.2 and Appendix 8) detected *Lactobacillus* in every sample and showed the abundance was greatest in sample 13c. The two samples where *Lactobacillus* was detected by molecular techniques (13c and 13d) had the largest copy number of the genus. Notably, culture was capable of detecting *Lactobacillus* in the sample which contained the smallest copy number of *Lactobacillus.*

Table 3.2 – Counts of bacteria and Lactobacillus in longitudinal stool samples as determined by qPCR

Sample	Total bacterial 16S rRNA copy No. (g)	Lactobacillus 16S rRNA copy No. (g)	% [a]
13a	2.34E+09	5.09E+06	0.22
13b	3.60E+08	1.04E+07	2.87
13c	2.64E+09	6.54E+07	2.47
13d	2.58E+08	1.84E+07	7.14
13e	2.88E+10	1.64E+06	0.01

[a] Lactobacillus load expressed as the percentage of total bacteria

3.2.5 Constrained ordination analysis

Constrained ordination analyses using CCA was applied to pyrosequencing and DGGE data to aid the interpretation of the variation in taxa composition between samples and whether they were associated with environmental or patient demographic data. The CCA revealed similar trends in the data from each technique with the final figures closely representing horizontally flipped versions of each other (Fig 3.3). For both techniques, the principal axis of variation can be interpreted as the disease state and longitudinal samples are shown to cluster together. However, the significance values based on the association between the continuous and discrete variables and community structure differed between techniques. Only gestational age (P=0.046) in the pyrosequencing analyses was a statistically significant driver of community structure. Analysis based on the pyrosequencing data also explained the most variance within the data, with the 2 principle axes explaining 52% of variance (Fig 3.3A) compared to 34% of the variance in the DGGE analysis (Fig 3.3B).

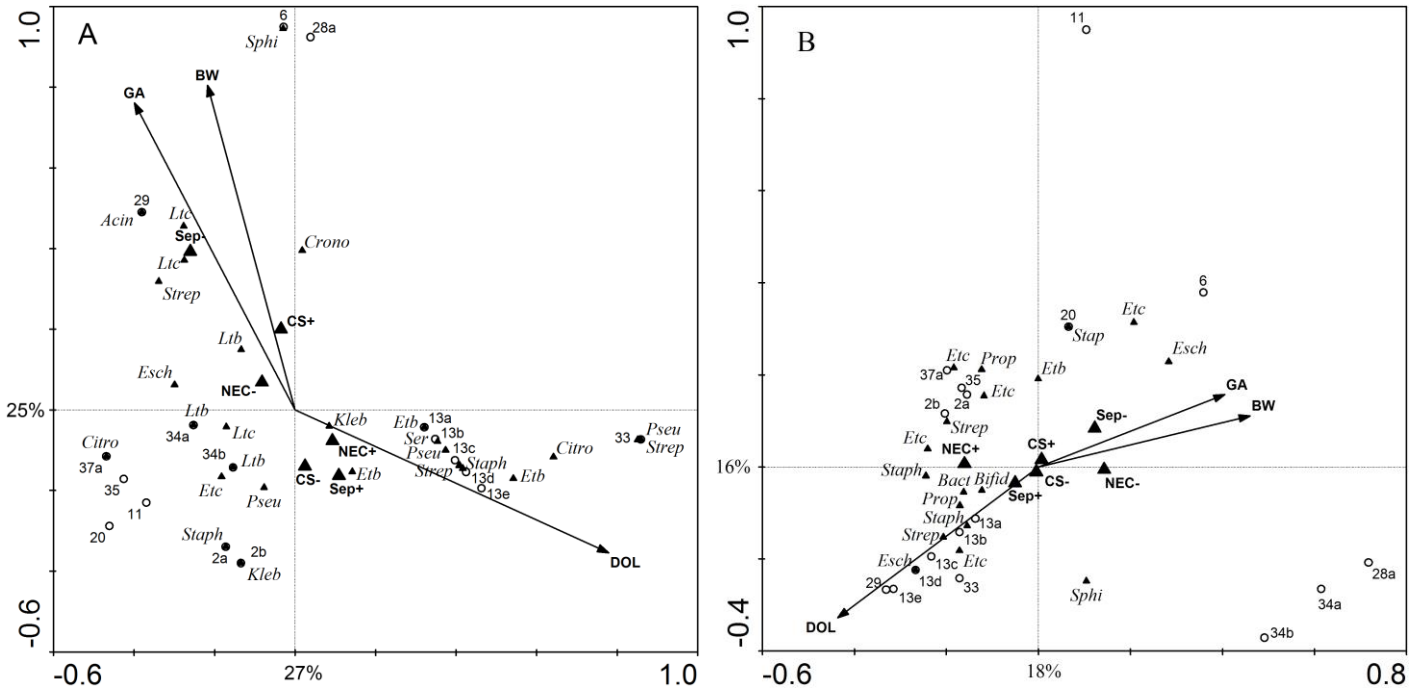


Figure 3.3 - Multivariate analysis on pyrosequencing and DGGE profiles. CCA based on the normalised data from each technique. Clinically insignificant taxa and unsequenced bands were included in the analysis but removed from the final graph for clarity. (A) Pyrosequencing CCA. (B) DGGE CCA. CS+ = caesarean birth, CS- = vaginal birth, GA = gestational age, BW = birth weight. Acin = *Acinetobacter*, Bac = *Bacteroides*, Bifid = *Bifidobacterium*, Citro = *Citrobacter*, Crono = *Cronobacter*, Etb = *Enterobacter*, Etc = *Enterococcus*, Esch = *Escherichia*, Flavo = *Flavobacteria*, Kleb = *Klebsiella*, Ltb = *Lactobacillus*, Ltc = *Lactococcus*, Pseu = *Pseudomonas*, Prop = *Propionibacterium*, Ser = *Serratia*, Sphi = *Sphingomonas*, Staph = *Staphylococcus*, Strep = *Streptococcus*

3.3 Discussion

The techniques utilised in this study are commonly used for studying clinical samples (Björkström *et al.* 2009; Lindberg 2010; Morowitz *et al.* 2010). Here, the focus is not on the role of the gut microbiota in disease, but rather on comparing and contrasting the data derived from molecular techniques with those derived from culture based approaches, in a clinical context.

In accordance with the findings from previous studies, facultative anaerobes including enterobacteria, enterococci and staphylococci were the most frequently detected genera by all techniques (Gewolb *et al.* 1999). However, the two most abundant taxa detected using DGGE (*Enterococcus* spp. and *Streptococcus* spp.) differed from those most commonly detected by culture and pyrosequencing (*Staphylococcus*, *Enterobacter*, and *Klebsiella*). Despite this, the three techniques were comparable in identifying the other predominant members of the bacterial community and, although differences occurred between the bacterial profiles elucidated by the molecular techniques, the overall constrained ordination analysis revealed comparable findings.

This study demonstrated, in agreement with previous work, that molecular techniques allow for greater coverage of the microbial community when compared with traditional culture based approaches (O'Sullivan, 2000; Mshvildadze *et al.*, 2010). It is important to note that culture methods were employed to provide some validation of the data provided by molecular methods. The culture data used in this study reflected the routine approach used in clinical microbiology laboratories, which is routinely used to inform clinical intervention and treatment. Thus, more exhaustive culture techniques involving a wider range of selective media incubated for many days would have been likely to yield additional species. Our aim was to obtain a snapshot of the predominant flora in

the sample, using only 19 culture plates per sample, and to see if the species detected by culture were also detected using molecular techniques. Overall, DGGE detected the greatest number of taxa (11 of 17 samples). This may reflect the fact that a single organism may contain multiple copies of the 16S RNA gene leading to multiple bands and thus potentially distorting the true representation of the microbial community (Kang *et al.* 2010).

There were several samples for which the culture-based approach identified genera that both DGGE and pyrosequencing were unable to detect. This suggests that, despite the improved coverage of samples by molecular techniques, some less abundant taxa may be missed (Vaz-Moreira *et al.* 2011). For example, DGGE did not show *Klebsiella* to be a prevalent genus, a result that could have clinically significant implications as *Klebsiella* has been associated with NEC in both culture and molecular studies (Carlisle *et al.* 2011; Westra-Meijer *et al.* 1983). It is also noteworthy that *Lactobacillus* was missed by both molecular approaches in a number of samples. This genus is considered to be functionally significant in infant gut development (Lin *et al.* 2008) and has been problematical to detect in previous studies that utilised DGGE and pyrosequencing to investigate the preterm gut microbiota (LaTuga *et al.* 2011). We found that pyrosequencing only detected *Lactobacillus* in samples which contained the largest copy number of the genus (Table 3.2). This inability of molecular techniques to detect organisms successfully isolated in culture may be due to large amounts of DNA from non-viable bacteria in the sample or by inefficient DNA extraction from certain taxa (Harris 2003; Zoetendal *et al.* 2006). Furthermore, it has been demonstrated that minor adaptions to primer sequences may facilitate detection of previously problematic sequences, without affecting the ability to amplify other taxa (Sim *et al.* 2012). Increasing the number of sequence reads per sample will allow for the less abundant species to be identified, but this has cost implications.

In contrast, culture based studies have been consistent in identifying the presence of *Lactobacillus* in the gut (Björkström *et al.* 2009; Gewolb *et al.* 1999). This study suggests that although routine clinical culture employed for detection of organisms in stool may fail to identify a substantial number of bacteria from the gut microbiota (O'Sullivan 2000), it has significant utility in identifying viable bacterial populations and in detecting numerically rare but functionally significant bacterial taxa that molecular techniques may miss.

Other issues that need to be addressed in developing molecular approaches as diagnostic tools in clinical microbiology are the application of stringent data pipelines to correct for sources of error which would ultimately affect downstream analyses. Our data was subjected to a filtering pipeline in mothur to minimise the sequencing error rate and incidence of chimeras (Schloss *et al.* 2011). Following filtering, the rarefaction curves for the samples in this study show that the entirety of the preterm infants faecal microbiota was not fully elucidated with the number of reads employed. It is of increasing importance that the reads generated in pyrosequencing are administered to stringent quality-filtering as described in depth elsewhere (Schloss *et al.* 2011).

The limitations of DGGE are widely acknowledged and extensively reported (Green *et al.* 2009). However, due to the relative ease and lack of cost and time associated with generating a community fingerprint it remains a valuable tool in the examination of microbial community structures, particularly in analysing how environmental and demographic variables can impact on the composition and dynamics of the community (Fromin *et al.* 2002). Conversely if identification through sequencing of bands is undertaken then DGGE can become relatively expensive and time consuming. Furthermore, due to the limitations associated with PCR bias and multiple copies of genes, potentially important clinical and functional groups, such as lactobacilli or *Klebsiella* could be missed (Vaz-Moreira *et al.* 2011). More broadly, all molecular

techniques focussing on extracted DNA (rather than RNA) neglect organism viability and will therefore report information on non-viable bacterial populations as well as taxa for which culture conditions were not optimised. This may give a skewed picture of the functionally active bacterial community within the clinical sample (Vaz-Moreira *et al.* 2011).

In summary, molecular techniques allowed for much greater coverage of the premature infant gut microbiota, although there were some functionally significant taxa only identified by traditional culture. Therefore, in order to fully elucidate complex ecosystems such as the gut microbiota, it is important to implement an integrative approach to overcome limitations of any single technique (O'Sullivan 2000). Moreover, when examining specific taxa using a molecular approach, it is paramount to use specific primers as universal primers may fail to detect the desired taxa in the total community. We have shown that there are instances of coherence between culture and molecular based approaches and despite the increased coverage of the latter, the techniques elucidated comparable profiles. Consequently, culture based approaches still offer important insight in the analyses of clinical samples.

4. The preterm gut microbiota: changes associated with necrotising enterocolitis and infection

Abstract

Aim: To describe gut colonisation in preterm infants using standard culture and 16S rRNA profiling, exploring differences in healthy infants and those who developed late onset infections.

Methods: 99 stools from 38 infants (median 27 weeks gestation) underwent routine culture at the hospital microbiology laboratory; a subset of 44 stools from 27 infants (median 26 weeks gestation) underwent molecular analysis using the 16S rRNA gene. Ordination analyses explored effects of patient variables on gut communities.

Results: Standard microbiological culture identified a mean of 2 organisms (range 0-7) and DGGE identified a mean of 12 bands (range 3-18) per patient. *Enterococcus faecalis* and coagulase negative staphylococci (CoNS) were most common by culture (40% and 39% of specimens). Meconium was not sterile and no fungi were cultured. Bacterial community structures in infants with NEC and sepsis differed from healthy infants. Infants who developed NEC carried more CoNS (45% vs 30%) and less *Enterococcus faecalis* (31% vs 57%). 16S identified *Enterobacter* and *Staphylococcus* presence associated with NEC and sepsis, respectively.

Conclusions: Important differences were found in the gut microbiota of preterm infants who develop NEC/sepsis. The relationship of these changes to current practices in neonatal intensive care requires further exploration.

4.1 Introduction

The gut microbiota plays a crucial role in development of immune function, micronutrient production, absorption, mucosal barrier function, and modulating the systemic inflammatory response (Macia *et al.* 2012). However, bacteria have been implicated in the causal pathway for NEC in animal models and by circumstantial evidence in preterms (Morowitz *et al.* 2010). NEC and infection together cause 21% of all deaths in infants born <32 weeks gestation (Berrington *et al.* 2012): both are likely to be affected by gut microbiota. Previous studies have demonstrated associations between the presence of bacterial species and an increased risk of NEC (Westra-Meijer *et al.* 1983; Hoy *et al.* 1990), but candidate organisms differ between studies and it is unlikely a single organism 'causes' NEC. Increasing understanding of microbial community interaction, revealed by molecular technologies, supports the concept that microbial contributions to NEC are mediated by changes in the community interactions and structure (Wang *et al.* 2009; Mai *et al.* 2011). The gut microbiota development in preterm infants is further influenced by policy and practice, reflecting the exposure to breast milk, antibiotics, and probiotics within the NICU (Frank *et al.* 2011).

Molecular techniques overcome the limitation of selective media for culturing organisms by amplifying universally conserved genes within different taxonomic lineages. DGGE is molecular fingerprinting techniques which utilises the highly conserved nature of the 16S rRNA gene to identify bacterial taxa. Molecular analysis exploring the phylogenetic diversity of the gut microbiota have shown that the bacterial diversity increases over time, with shifts in community structure associated with changes in diet and health (Morowitz *et al.* 2011; Koenig *et al.* 2010). Modelling gut microbiota in term infants suggests competition between three phyla *Bacteroides*,

Proteobacteria and *Firmicutes*, explains most of the community dynamics (Palmer *et al.* 2007; Trosvik *et al.* 2010).

In preterm infants, especially those developing NEC or sepsis, data is limited. Many of the studies using culture are more than 20 years old when infants and nursery practices were considerably different (Blakey *et al.* 1982; Stark & Lee 1982; Sakata *et al.* 1985; Hall *et al.* 1990; Hoy *et al.* 1990; Björkström *et al.* 2009). Previous molecular studies are limited by restricted sampling, small numbers and relatively mature gestations studied (Lindberg 2010; France *et al.* 2007; Wang *et al.* 2009; Mshvildadze *et al.* 2010; Mai *et al.* 2011). However, they have demonstrated that bacterial diversity in NEC patients appears different from controls, with fewer *Firmicutes* and a bloom in *Proteobacteria* before NEC onset. More specifically, *Enterobacteriaceae* have been detected more frequently in patients diagnosed with NEC, compared to controls (Mai *et al.* 2011).

We aimed to improve understanding of the gut microbial community in preterm infants and those associated with NEC and sepsis using culture and DGGE, as a first step in evaluating medical or environmental practices that might contribute to these changes.

4.2 Results

4.2.1 Patients and samples

The 38 patients had a median gestation of 27 weeks (range 23-31), birth weight 895g (range 520g - 1850g); the 27 patients contributing molecular data did not differ significantly from the overall population, 35 infants received breast milk, 29 antifungals. NEC was diagnosed in 8 infants (4 surgical) of which 7 (4 surgical)

contributed molecular data; 13 developed sepsis, of which 9 contributed to the molecular data (Table 4.1). Of those with NEC or sepsis, 18 samples were before onset of disease and 14 after.

Table 4.1 - Demographic data and bacterial species detected using standard culture across all 99 samples

Sample	Gestation age (Weeks)	Birth weight (g)	Delivery mode	Sex	Day of Life	Sepsis	Sepsis day onset	NEC	NEC day onset	Live	*Acinetobacter junii*	*Actinomyces naeslundii*	*Bifidobacterium Spp.*	*Citrobacter braakii*	*Citrobacter freundii*	*Citrobacter koseri*	*Clostridium beijerinckii*	*Clostridium perfringens*	*Clostridium sporogenes*	*Enterobacter aerogenes*	*Enterobacter amnigenus*	*Enterobacter cloacae*	*Enterococcus faecalis*	*Enterococcus faecium*	*Escherichia coli*	*Klebsiella oxytoca*	*Klebsiella pneumoniae*	*Lactobacillus spp.*	*Morganella morganii*	*Propionibacterium acnes*	*Propionibacterium granulosum*	*Proteus mirabilis*	*Pseudomonas aeruginosa*	*Serratia liquefaciens*	*CN Staphylococcus*	*Staphylococcus aureus*	*Stenotrophomonas maltophila*	*Group B Streptococcus*	*Streptococcus acidominius*	*Streptococcus oralis*	*Streptococcus mutans*
1a					9																																			
1b [a]	24	670	Vaginal	M	16	N	N/A	N	N/A	Y																															
1c					23																																				
2a					16																																				
2b					24																																				
2c [a]	24	520	Cesarean	F	41	Y	4	N	N/A	Y																															
2d [a]					48																																				
2e					56																																				
3a [a]					14																																				
3b [a]	29	1340	Cesarean	F	20	N	N/A	N	N/A	Y																															
3c [a]					27																																				
4 [a]	27	580	Cesarean	F	24	N	N/A	Y	13	Y																															
5	29	1420	Vaginal	M	1	N	N/A	N	N/A	Y																															
6a [a]					6																																				
6b	30	1370	Cesarean	M	11	N	N/A	N	N/A	Y																															
7a					2																																				
7b [a]	26	485	Cesarean	M	21	N	N/A	N	N/A	Y																															
8a					18																																				
8b [a]	28	1160	Vaginal	M	25	N	N/A	N	N/A	Y																															
8c [a]					34																																				
8d [a]					46																																				
9 [a]	33	1850	Vaginal	M	1	N	N/A	N	N/A	Y																															
10	28	880	Cesarean	M	11	Y	10	N	N/A	Y																															
11a [a]	25	825	Vaginal	M	6	Y	51	Y	17	N																															
11b					38																																				
12a					16																																				
12b [a]					30																																				
12c					36																																				
12d [a]	24	680	Vaginal	F	40	N	N/A	N	N/A	Y																															
12e [a]					44																																				
12f [a]					51																																				
12g					59																																				
13a					25																																				
13b [a]					34																																				
13c [a]	23	600	Vaginal	M	39	Y	38	Y	12	Y																															
13d [a]					48																																				
13e [a]					53																																				
13f [a]					60																																				
14 [a]	28	1040	Cesarean	F	26	N	N/A	N	N/A	Y																															
15a					17																																				
15b					22																																				
15c [a]	25	635	Vaginal	F	29	N	N/A	N	N/A	Y																															
15d [a]					38																																				
16a					1																																				
16b					9																																				
16c	24	645	Vaginal	M	22	N	N/A	N	N/A	Y																															
16d [a]					32																																				
16e [a]					38																																				
17	27	1170	Vaginal	M	13	N	N/A	N	N/A	Y																															
18a	27	1365	Vaginal	M	7	Y	3	N	N/A	Y																															
18b [a]					13																																				
19 [a]	26	870	Cesarean	M	55	Y	13	Y	10	Y																															
20a [a]					7																																				
20b	26	915	Vaginal	M	11	N	N/A	Y	20	N																															
20c					18																																				
20d					32																																				
21a [a]					11																																				
21b					20																																				
21c [a]	28	1100	Cesarean	M	22	N	N/A	N	N/A	Y																															
21d					25																																				
21e					32																																				
21f					39																																				
22a [a]	27	870	Cesarean	M	15	N	N/A	N	N/A	Y																															
22b [a]					23																																				
23a					14																																				
23b	27	790	Cesarean	F	22	N	N/A	N	N/A	Y																															
23c					29																																				
24	29	1020	Cesarean	M	2	N	N/A	N	N/A	Y																															
25a					13																																				
25b	25	810	Vaginal	F	16	N	N/A	N	N/A	Y																															
25c					20																																				
25d [a]					27																																				
26	28	1360	Cesarean	M	13	N	N/A	N	N/A	Y																															
27	25	740	Cesarean	M	37	Y	34	N	N/A	Y																															
28a [a]	28	1250	Cesarean	M	1	N	N/A	N	N/A	Y																															
28b					11																																				
29a [a]					1																																				
29b	28	1180	Cesarean	M	2	N	N/A	Y	17	N																															
29c					9																																				
30a					16																																				
30b					25																																				
30c	24	650	Vaginal	M	56	Y	81	Y	24	N																															
30d					65																																				
30e					93																																				
31	28	1200	Cesarean	F	17	N	N/A	N	N/A	Y																															
32a	27	910	Vaginal	M	17	Y	8	N	N/A	Y																															
32b					25																																				
33 [a]	26	995	Cesarean	M	145	Y	144	N	N/A	Y																															
34a [a]					6																																				
34b	25	800	Vaginal	M	17	Y	8	N	N/A	Y																															
34c [a]					24																																				
34d					31																																				
35a					13																																				
35b	26	570	Cesarean	M	18	Y	23	N	N/A	Y																															
35c					34																																				
35d [a]					36																																				
36 [a]	27	1120	Cesarean	M	60	Y	77	Y	21	Y																															
37 [a]	28	1090	Vaginal	F	1	N	N/A	N	N/A	Y																															
38	24	640	Vaginal	F	27	N	N/A	N	N/A	Y																															

[a] Sample involved in molecular analysis

4.2.2 Standard culture analyses

In total 31 species were identified. Individual stools contained a mean of 2 different species (range 0-7). The most common species were *Enterococcus faecalis* and CoNS (40% and 39% respectively). Infants with NEC were more likely to be colonised with CoNS (45% vs 30%) and less likely colonised with *Enterococcus faecalis* (31% vs 57%). Only five infants were colonised with lactobacilli and one with *Bifidobacterium* sp. All meconium samples (samples collected on day one) were colonised with at least one species (Table 4.1). No fungi were cultured.

In the subset of stools for which molecular analysis was possible, 24 taxa were identified by culture. PCoA was used to study the structure of the community determined by culture, the two major principle components explained 37% of the variance. Along the horizontal axis, profiles from healthy infants distributed to the left (14 samples from a total of 25) and those from infants with NEC and/or sepsis to the right (16 samples from a total of 18) (Fig 4.1A). Samples (11a, 20a, and 29a) collected prior to NEC diagnosis also show a distinct cluster. The most significant taxa associated with this community shift were *Enterococcus faecalis*, which were more frequently isolated from healthy patients (58% of healthy patients vs. 22% of NEC/sepsis infants), and CoNS which were more frequently isolated from NEC/sepsis patients (56% of NEC/sepsis vs. 35% healthy) (Table 4.1).

4.2.3 DGGE analyses

DGGE analyses identified 74 individual taxa (range 3-18, mean 12 per stool). The mean diversity of each sample, (Shannon diversity index (H')), was 1.9 ± 0.45 indicative of relatively low species diversity. The six DGGE samples from infants less than ten days old had a significantly lower H' (1.19) compared to infants ten days or older (H''=2.04).

Community profile variance associated with the categorical variables is shown in Fig 4.1B. These analyses identify variables significantly associated with changes in the community structure. The x axis (presence or absence of NEC/sepsis) explained 19% of the variance of gut bacterial community profiles and was significantly different in sepsis from healthy infants (P=0.016) (Fig 4.1B), before and after onset of disease and again showed the distinct clustering of samples pre NEC. The y axis explained 14% of variation associated with gestational age. Bands on the DGGE gels most associated with changes in bacterial community associated with NEC and sepsis were *Enterobacter*, *Flavobacterium*, *Staphylococcus* and *Propionibacterium* (Fig 4.2).

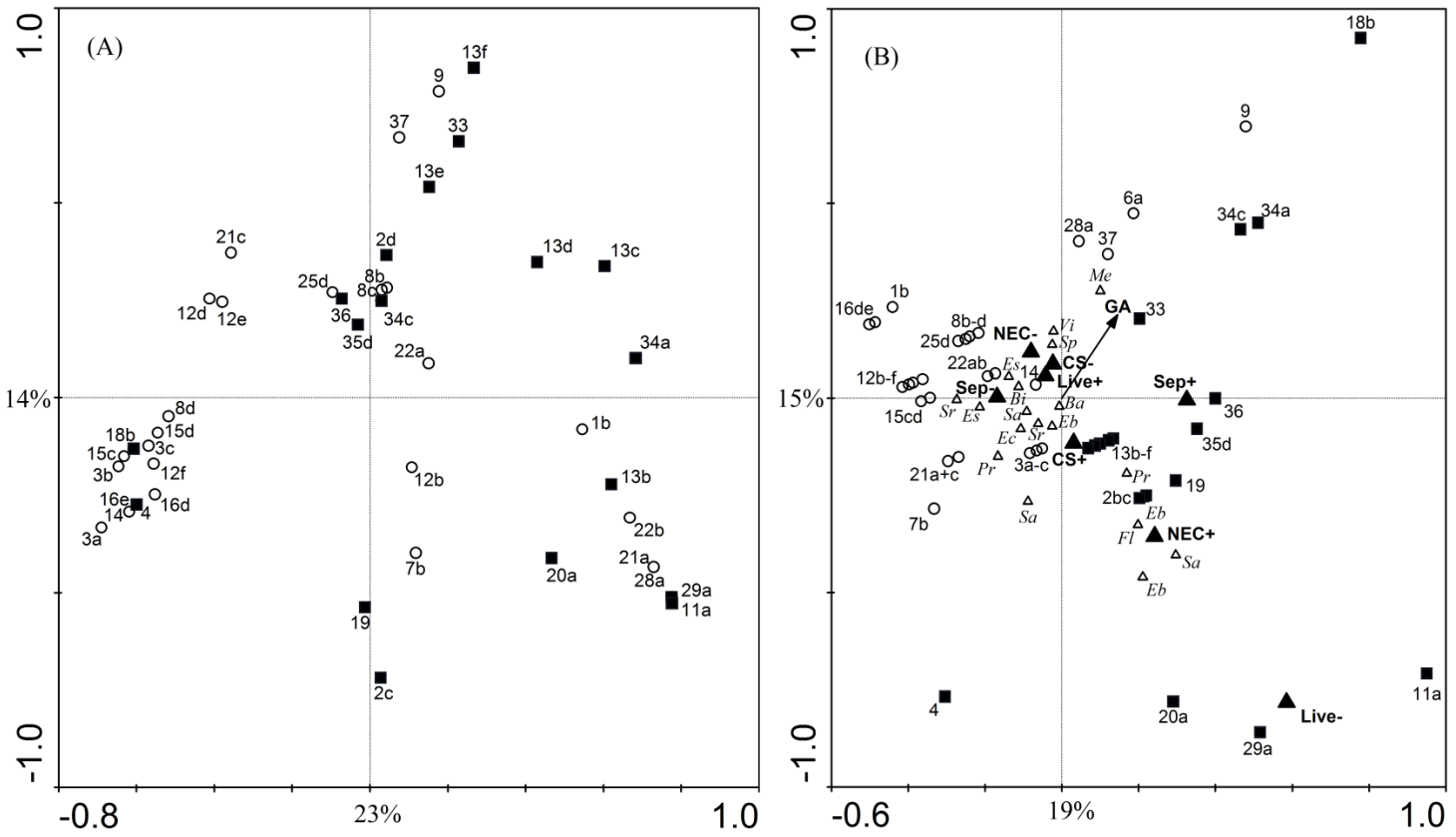


Figure 4.1 – Multivariate analysis on culture and DGGE. Circles denote healthy patients and squares denote patients diagnosed with NEC and/or sepsis. (A) PCoA based on presence and absence of species identified by culture. (B) CCA based on normalised DGGE band matrix. CS+ = caesarean birth, CS- = vaginal birth, GA = gestation age. Sequences obtained are displayed – Ba = *Bacteroides*, Bi = *Bifidobacterium*, Eb = *Enterobacter*, Ec = *Enterococcus*, Fl = *Flavobacteria*, Pr = *Propionibacterium*, Sp = *Sphingomonas*, Sa = *Staphylococcus*, Sr = *Streptococcus*

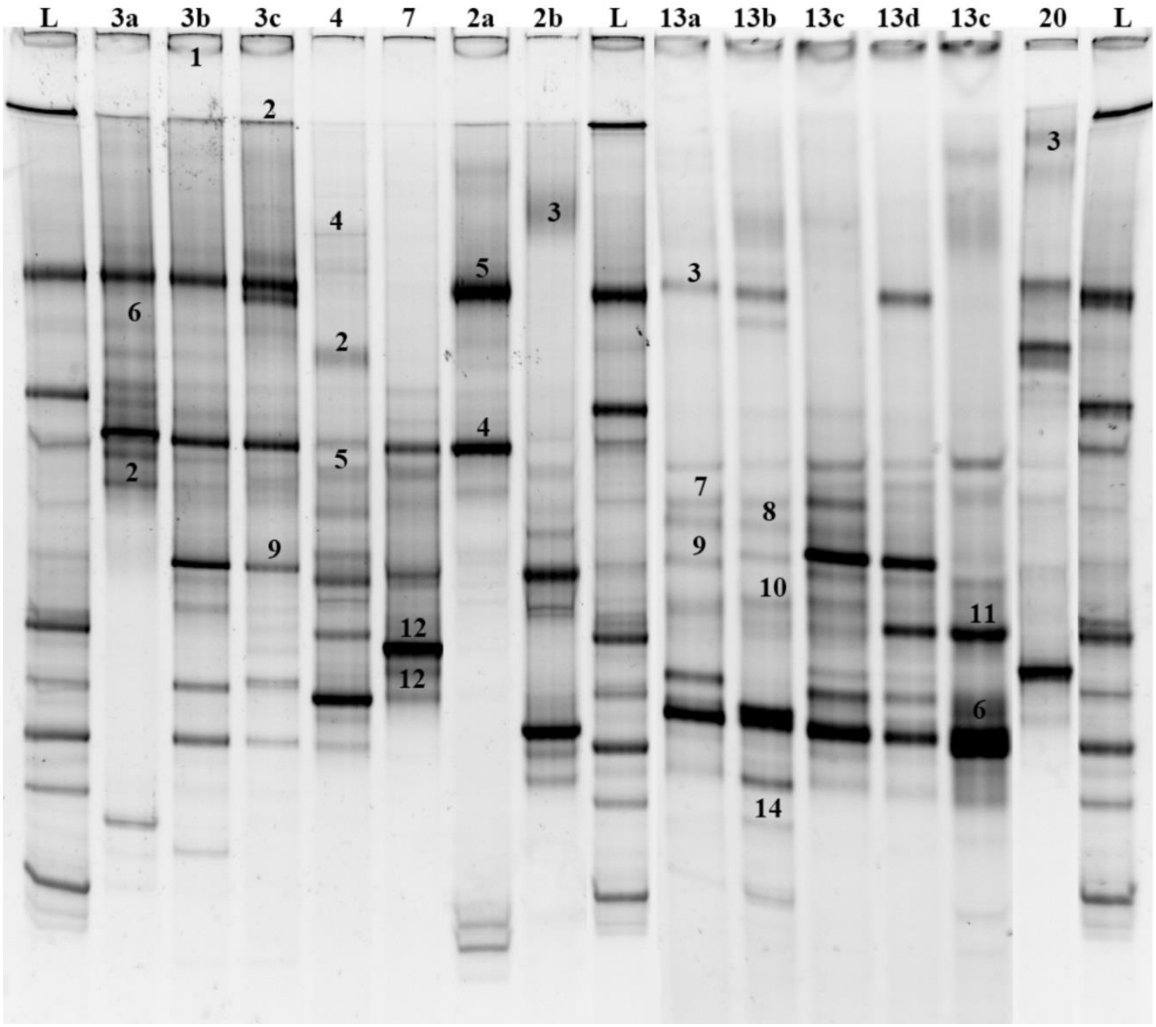


Figure 4.2 – Representation of a DGGE gel showing which bands were excised and sequenced. L - Ladder, 1 - *F.symbiont*, 2 - *E.faecalis*, 3 - *S.epidermidis*, 4 - *S.salivarius*, 5 - *P.acnes*, 6 - *E.cloacae*, 7 - *B.fragilis*, 8 - *B.longum*, 9 - *S.mutans*, 10 - *S.aromaticivorans*, 11 - *E.ludwigii*, 12 - *E.coli*, 13 - *V.atypica*, 14 - *M.populi*

4.3 Discussion

Both standard culture and molecular techniques indicated differences between the gut microbiota of healthy infants and those who develop NEC/sepsis, before and after diagnosis. Not surprisingly, molecular approaches had greater resolution in quantifying these differences. Our molecular analyses are in agreement with earlier work indicating samples of meconium (day 1 stools) were not sterile (Jiménez *et al.* 2008) and bacterial diversity was initially low, increasing with time (Koenig *et al.* 2010).

The results of this study suggest certain taxa are more frequently associated with health and disease states in premature infants: *Enterobacter* spp. and *Staphylococcus* spp. were associated by both methods with NEC and have been previously implicated in the disease state (Acker & Smet 2001; Morowitz *et al.* 2010). Both methodologies also suggest an association between *Enterococcus* spp. and *Streptococcus* spp. and remaining healthy: *Enterococcus* specifically may play a key role in gut development (Are *et al.* 2008). In contrast to Björkström *et al.* (2009), we identified very little colonisation with 'healthy' bacteria: only 6 infants were identified as carriers of lactobacilli or bifidobacteria, compared to 24/44 of their infants. However, while most of our infants received breast milk they were significantly less mature than those in Björkström's study, which is likely to influence the bacterial colonisation. Furthermore, discrepancies between the culture based methodology between studies may lead to further disparities in the ability to detect specific organisms (Björkström *et al.* 2009).

The hypothesis that NEC is due to atypical bacterial community assembly and interactions, requires the bacterial community to be studied as a whole. Molecular methodologies offer the opportunity to explore whether patient related variables are significantly associated with changes in the bacterial community: ordination analyses

show variation in species composition between samples and the underlying environmental factors - samples with similar communities group more closely. We demonstrated that the most significant factor associated with bacterial community structural change was NEC and sepsis: our data highlight changes in the gut microbiota that precede disease. Profiles prior to NEC diagnosis did not cluster with healthy patients, as shown by others (Wang *et al.* 2009). These changes may be modulated by a variety of factors including genetic predisposition, feeding practices, maternal dietary changes in breast fed infants, medical interventions within neonatal intensive care (antibiotics etc.), and may be closely or more distantly temporally linked to disease onset. Gestational age, the single most important risk factor for development of NEC (Martin & Walker 2008; Chauhan *et al.* 2008), was not found to be significant in shaping the bacterial community (P=0.09). This suggests that increased immaturity *per se* is not significantly associated with a bacterial community associated with NEC and sepsis, an important finding. The greater prevalence of Enterobacteriaceae and Staphylococcaceae associated with NEC and sepsis (Fig 4.1B) was more significant. Caesarean section, previously implicated with increased NEC (Marques *et al.* 2010), also showed a weak association with the bacterial community associated with increased incidence of NEC/sepsis.

The ability to define and then replicate a 'healthy' preterm gut microbiota would be of practical benefit to preterm infants, but whether changes in the gut microbiota observed are a consequence of changes that cause NEC/sepsis, or *vice versa* remains unclear. This study demonstrates how some clinical factors can help explain some variance, but not all. Some changes observed predate disease diagnosis, suggesting that microbiomic changes may be part of a causal pathway. However, the number of informative samples in this study is relatively small as sampling was convenience rather than targeted, thus infants contribute variably to the overall analysis. Much larger datasets will be required

to more precisely elucidate how these factors operate independently and when combined. As some of the factors that influence the gut microbiota (antimicrobial choices and duration) are within clinical control, further exploration of the detailed effects of these is needed. Other potentially important gut microbiota manipulations, such as the delivery of probiotics (Embleton & Yates 2008) and lactoferin (Venkatesh & Abrams 2010) to preterm infants will also require careful evaluation in the future and may further enhance our understanding. Current evidence suggests that probiotics might be protective against NEC but not sepsis (Deshpande *et al.* 2007). However, the mechanism of action in relation to the gut microbiota is poorly studied and there are currently no studies exploring whether long term detrimental effects might occur (Neu 2011).

Clinical effects of changes in the gut microbiota may be delayed by days or weeks, and detailed sampling over long time periods is required to fully understand these. Manipulations of the gut microbiota made by early medical decisions may also have very long lasting (metabolic/allergic) effects which require further evaluation, and tracking of these infants into later (adult) life may be illuminating. Given the significant mortality and morbidity of these disease states to preterm infants, further exploration of the gut microbiota should be considered an important research priority.

5. Bacterial and fungal viability in the preterm gut: necrotising enterocolitis and sepsis

Abstract

Aim: Evidence suggests that microbial communities in the preterm gut may influence the development of NEC and sepsis. Existing data often neglect fungi and whether bacteria were metabolically active or not. We sought to characterize the bacterial and fungal stool flora of preterm neonates and organism viability and evaluate any associations with NEC and sepsis.

Methods: Stool samples (*n*=136) from 32 patients (<32 weeks gestation) were collected between birth and day of life 95. Seven infants developed NEC and 13 developed sepsis. Stools were analysed by PCR-DGGE for assessment of the total bacterial and fungal communities by analysis of 16S rRNA and 28S rRNA respectively. In a subset of the cohort consisting of 65 samples (25 infants) the viable (RNA) bacterial and fungal communities were analysed. Analyses were performed to examine the possible effects of demographic or treatment related factors and the development of NEC or sepsis.

Results: In total 80 (66 viable) bacterial species were identified overall, and 12 fungal taxa (none viable). Total bacterial communities significantly differed between healthy infants and those with NEC or sepsis, with *Sphingomonas* spp. significantly associated with NEC. Significant drivers of community structure differed based on total or viable analysis. Antifungal prophylaxis was associated with altered bacterial community and reduced bacterial richness observed in week 4 correlated with high antibiotic exposure.

Conclusions: Total and viable communities differ in preterm infants, and non-viable fungal species are present in infants on prophylaxis. Exploration of viability and non-bacterial contributors to the total community may increase understanding of NEC and sepsis.

5.1 Introduction

Prematurity remains the leading predisposition to neonatal death and long term disability (LaTuga *et al.* 2011), with infection and NEC increasing in preterm infants (Berrington *et al.* 2012). Associated mortality and long term consequences for survivors underpin the need for improved understanding and prevention of both prematurity and the associated morbidities (Stoll *et al.* 2004). Recent interest has focused on the role of the gut microbiota in the pathogenesis of NEC and sepsis.

The bacterial community within the gut microbiota plays a crucial role in the development of immune function, micronutrient production and absorption, mucosal barrier function, and modulation of the systemic inflammatory response (Macia *et al.* 2012). Although bacterial colonisation is considered a pre-requisite for NEC, no single causative bacterial agent has been identified (Kaufman & Fairchild 2004). Studies reporting specific bacterial associations with NEC suggest a role for Proteobacteria taxa, commonly *Enterobacteriaceae* (Carlisle *et al.* 2011; Mai *et al.* 2011). Preterm neonates also show delayed colonisation by 'healthy commensal' organisms, especially bifidobacteria and lactobacilli, potentially attributable to medical management of prematurity (Schwiertz *et al.* 2003; Jacquot *et al.* 2011).

The diversity of the total gut microbiota in term neonates increases over time with abrupt community shifts associated with diet changes or antibiotic treatments (Koenig *et al.* 2010). Total community analysis is based on DNA and will include both metabolically active (live) and dead microorganisms. Analyses based on the total gut microbiota provide a phylogenetic picture of the community, but do not reflect the viable community (Tannock *et al.* 2004), which differs from the total community in adults (Peris-Bondia *et al.* 2011). Viable community analysis is based on RNA and

includes only metabolically active microorganisms but RNA imposes increased problems due to the potential for degradation, compared with DNA. RNA is unstable at room temperature and requires stringent sample handling procedures to ensure the sample is transported to the laboratory without degradation or a shift in microbial profiles. To prevent this degradation during transit it is paramount the sample is stored in an RNA stabilisation reagent (Hernandez *et al.* 2009).

The fungal community within the gut microbiota of preterm infants remains relatively unexplored, despite the importance of candidal infection in neonates and increasing anti-fungal prophylaxis within NICUs (Manzoni *et al.* 2009; Manzoni *et al.* 2011; Vergnano *et al.* 2011). A study which relied on cultivation identified 16 of 30 patients had no evidence of fungal colonisation, but found fungal sepsis to be a significant lethal factor in the surgical mortality of NEC (Smith *et al.* 1990). In necrotising pancreatitis, although the initial severity was comparable, patients with fungal infection tended to have a more complicated course and worse outcomes compared with those with bacterial infection (Grewe *et al.* 1999). In this study, a low dose of antifungal prophylaxis was advised in the management of necrotising pancreatitis, however, how antifungal prophylaxis affects the neonatal gut microbiota remains largely unknown.

We aimed to examine total (DNA) and viable (RNA) bacterial and fungal communities in the gut of preterm infants exposed to current NICU practices, exploring potential associations of the gut microbiota with a diagnosis of sepsis or NEC, antifungal prophylaxis, gender, birth mode, gestational age, birth weight, and postnatal age.

5.2 Results

5.2.1 Clinical Characteristics

Demographic information is in Table 5.1 and details of all individual infant antibiotic exposure are presented in Appendix 9. In total, 32 patients contributed to the analysis of the total community and 25 patients contributed to viable analysis. The demographics of the subset used in the viable analysis did not differ significantly from the complete cohort. 30 infants (23 from viable subset) received some breast milk and 30 (23 from viable subset) received antifungal prophylaxis (fluconazole). None received probiotics or prebiotics. Seven developed NEC (3 surgical): four contributed to the viability analysis. 6 samples predated and 12 post-dated NEC diagnosis. Thirteen infants developed sepsis with 5 infants having more than one episode: 10 contributed to the viability analysis. 22 samples predated and 32 post-dated sepsis diagnosis. Organisms causing sepsis were detected by blood culture and identified as: 10 coagulase negative staphylococci (CONS), 1 *Staphylococcus aureus*, 2 *Enterococcus faecalis*, 2 *Escherichia coli*, 1 *Klebsiella pneumoniae*, 1 *Pseudomonas aeruginosa*, 1 *Micrococcus luteus*, and 1 *Candida parapsilosis*.

5.2.2 Total communities

DGGE identified 80 individual species (mean 14 per stool, range 2-26). Interestingly, mean numbers of total bacterial species did not differ between healthy, NEC, and sepsis patients (Table 5.1). The most prevalent bacteria, identified by sequencing DGGE bands (see Methods 2.7), belonged to the genera *Enterococcus*, *Streptococcus*, and *Escherichia* (Appendix 9). DGGE analyses identified 12 fungal species (mean 2 per

stool, range 0-6). Half of infants in the cohort carried at least one fungal species, half showed no fungal colonisation. For most individual infants where fungi were identified they were in every sample with high intra-patient concordance. No fungal species were observed in any stool from any infant who developed NEC (all on fluconazole) (Table 5.1). Based on the DGGE ladder, *Candida* spp. were the most abundant fungi (61%) with *C. albicans* and *C. glabrata* responsible for 30% and 29% respectively.

Constrained ordination using CCA was carried out on the total bacterial community that explained 35% of the total variance (Fig 5.1). The principal axis, explained 19% of the variance separating patients according to disease state with healthy infants clustered separately from infants with NEC or sepsis. Patients with NEC (P=0.002) or sepsis (P=0.002) had significantly different profiles compared to healthy infants. Only colonisation with *Sphingomonas* spp. was significantly (P=0.0001) associated with NEC. Delivery mode (P=0.01) and gender (P=0.012) also influenced the bacterial community. Only 2 patients did not receive antifungal prophylaxis resulting in significantly different (P=0.03) bacterial profiles. Interestingly, the bacterial community was not significantly influenced by gestational age and birth weight.

The role of the bacterial community in NEC and sepsis was further assessed using PLS-DA to explore the gut microbiota pre disease diagnosis and post disease diagnosis when the patients will be subject to practises within the NICU (Fig 5.2). Healthy controls were found to cluster together. Samples predating and following disease diagnosis clustered distinctly from healthy controls.

Overall, the species richness in the total bacterial community increased with increasing postnatal age in all infants. However, the species richness of the fungal community remained relatively constant with a low richness observed throughout the initial weeks of life (Fig 5.3).

Table 5.1 – Demographic data and species richness from whole patient cohort

	Healthy (Patients = 17 / Samples = 80)	NEC[a] (Patients = 7 / Samples = 18)	Sepsis (Patients = 8 / Samples = 38)
Median gestational age (weeks)	26 (range 23 - 29)	27 (range 24 - 31)	25 (range 24 - 28)
Median birth weight (grams)	780 (range 495 - 1435)	920 (range 705 - 1660)	810 (range 615 - 1030)
Gender ratio: male/female	8 / 9	5 / 2	7 / 1
Birth mode ratio: Vaginal/Caesarean	10 / 7	3 / 4	5 / 3
Deceased	0	0	1
Mean taxa (per sample)			
Total Bacterial	14	14	14
Viable Bacterial	6	5	5
Total Fungal	2	0	3
Viable fungal	0	0	0

[a] Patients diagnosed with both NEC and sepsis are presented as NEC (n=6)

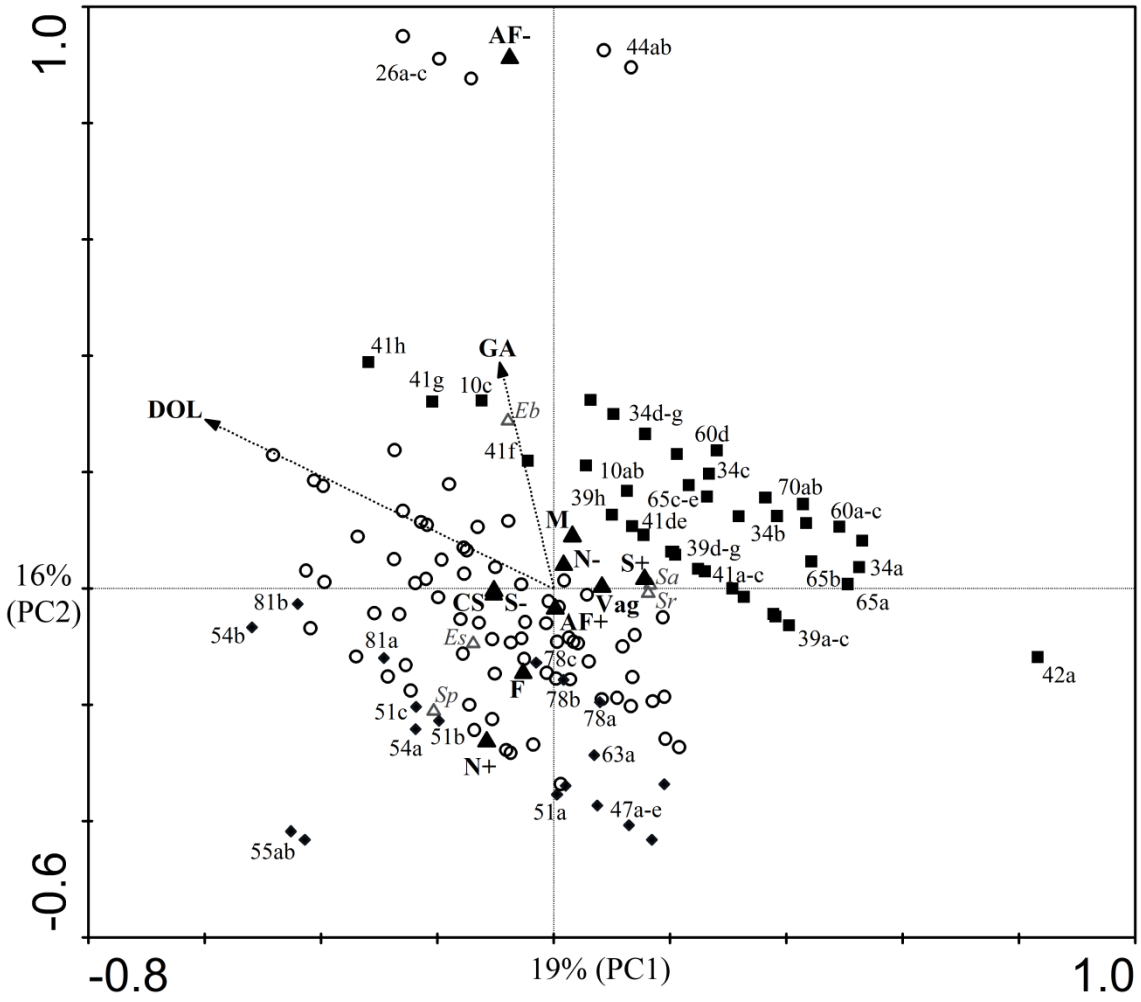


Figure 5.1 – CCA based on the total bacterial community. Comparison of patient profiles (Healthy patients (○), patients diagnosed with NEC (♦), and patients diagnosed with sepsis (■)) with discrete variables (▲) including; N = NEC (*P*=0.002), S = sepsis (*P*=0.002), AF = antifungal treatment (*P*=0.03), CS = caesarean / Vag = vaginal birth (*P*=0.01), M = male / F = female (*P*=0.012) and continuous explanatory variables (--►) including; GA = Gestation age (*P*=0.148), DOL = day of life (*P*=0.002). Sequenced bands (∆) include – Eb = *Enterobacter*, Es = *Escherichia*, Sp = *Sphingomonas*, Sa = *Staphylococcus*, Sr = *Streptococcus*

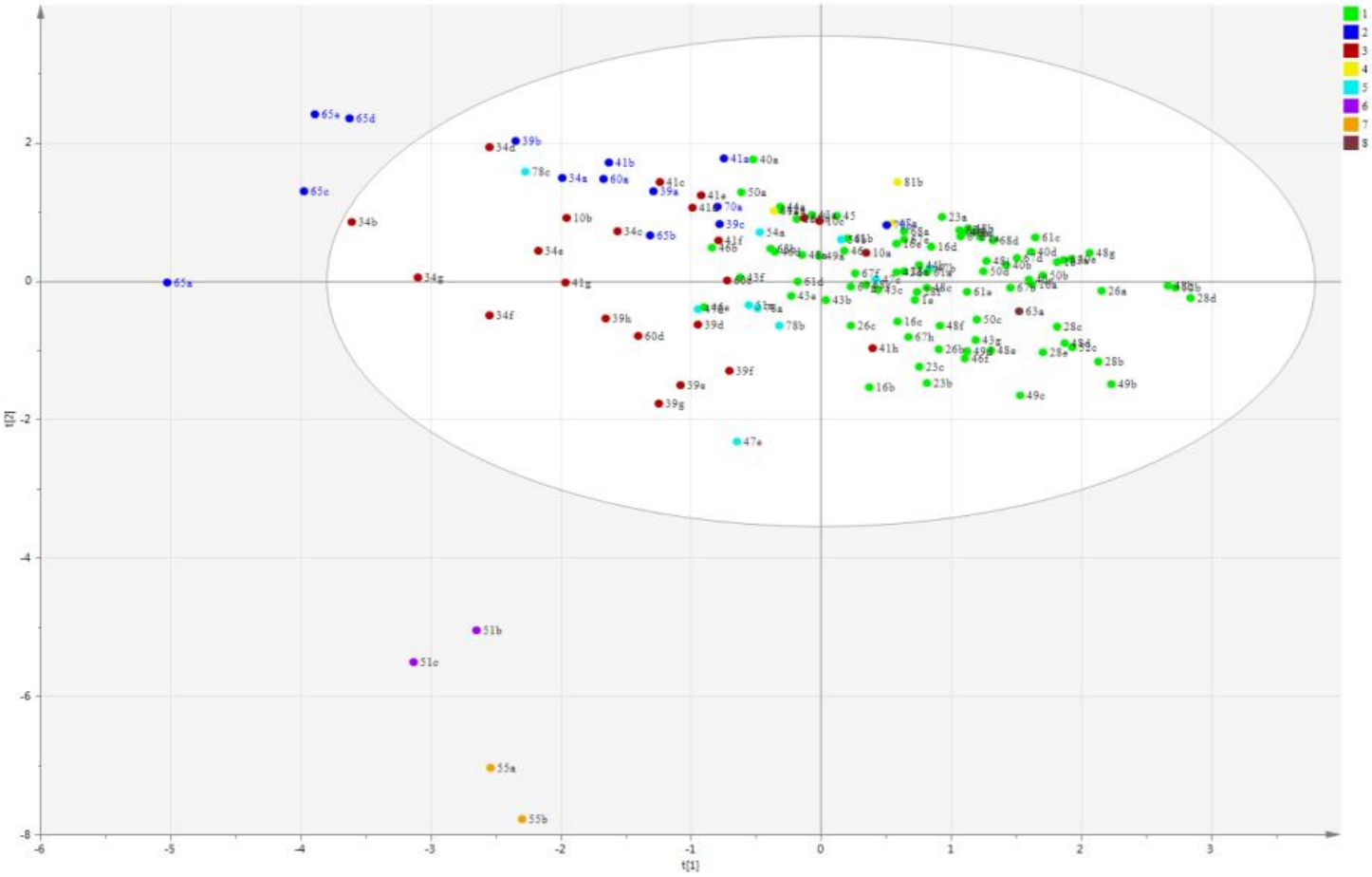


Figure 5.2 - PLS-DA exploring disease status on the gut microbiota. Grouped according to disease state. Groups are as follows – 1 = healthy, 2 = pre sepsis, 3 = post sepsis, 4 = pre NEC / pre sepsis, 5 = pre NEC / post sepsis, 6 = post NEC / pre sepsis, 7 = post NEC / no sepsis, 8 = post NEC / post sepsis. Numbers adjacent to each point refer to patients (Appendix 9). Lowercase letters refer to longitudinal samples from each patient.

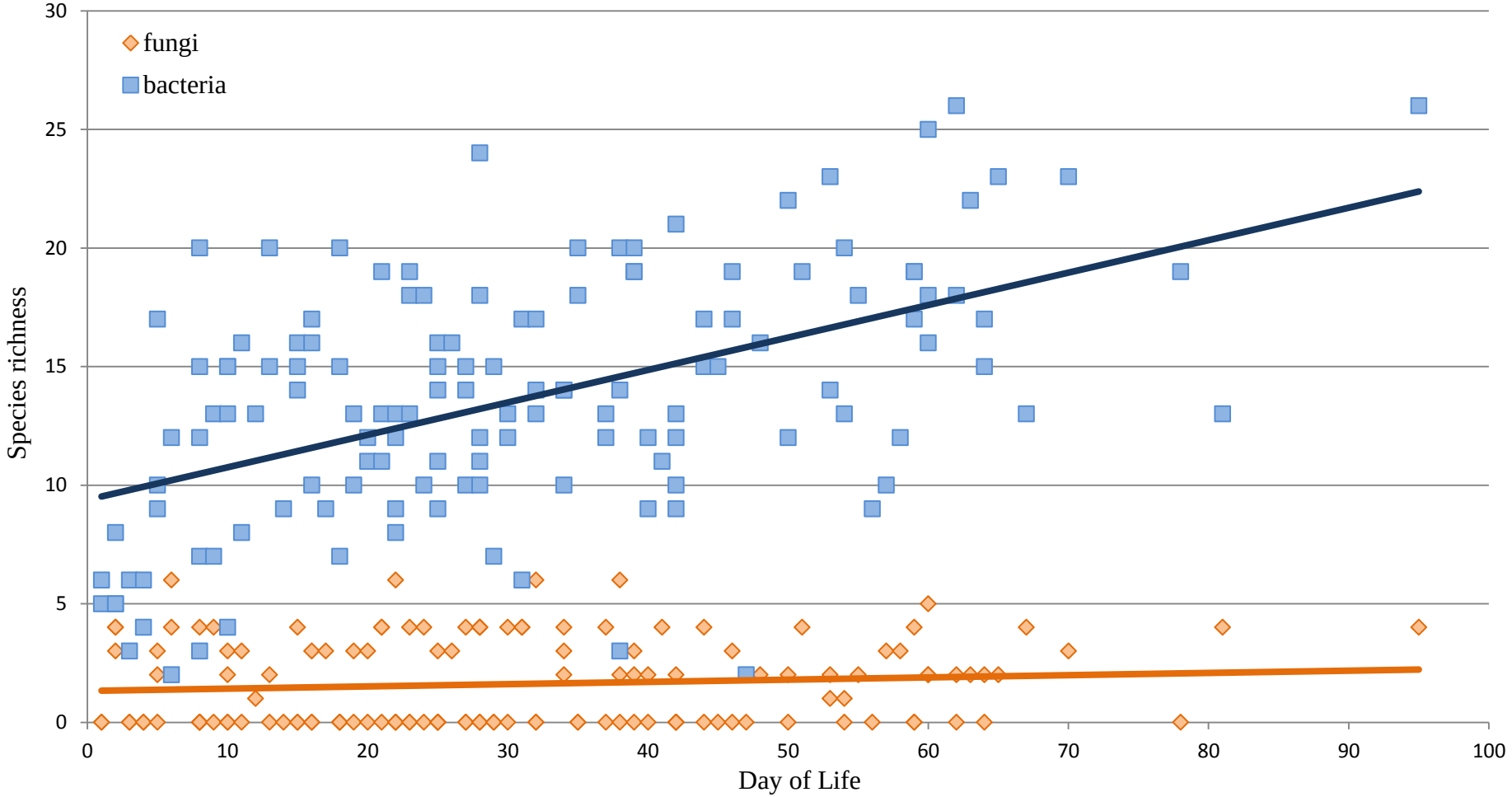


Figure 5.3 – Species richness. Total bacterial and fungal communities colonising the preterm neonate.

5.2.3 Viable communities

DGGE analyses on the viable bacterial community demonstrated the presence of 66 individual species (mean 6 per stool, range 1-14). There was no statistical difference between the numbers of viable bacterial species detected in infants with NEC or sepsis compared with healthy infants (Table 5.1). The most abundant viable bacteria matched the total community (*Enterococcus*, *Streptococcus*, and *Escherichia* spp.) (Appendix 9), but the constrained ordination differed (Fig 5.4). While both CCAs explained the same variance in the first 2 axes (35%), only gestational age (P=0.002), day of life (P=0.004), and sepsis (P=0.004) had a significant effect on the viable community (Table 5.2). Colonisation with *Sphingomonas* spp. remained significantly associated with NEC (P=0.0001).

To ensure the differences between the total and viable analysis was not a result of differences between the cohort size, an additional CCA using DNA matched to the viable cohort was also performed. This confirmed the bacterial analysis on the full cohort, showing gender (P=0.002) and birth mode (P=0.012) to be significant and gestation age to be insignificant in shaping the gut microbiota (Appendix 9). This is further supported by cluster analysis where a dendrogram based on the Bray-Curtis coefficient showed samples largely cluster based on sample, rather than grouping based on DNA or RNA analysis (Fig 5.5). This can be visualised on the DGGE gels where the DNA and RNA samples were ran out alongside one another; typically the most abundant bands from the DNA profile are observed in the RNA profile (Appendix 9).

No viable fungi were detected in any sample (Table 5.1). To ensure this was not a methodological issue, cultured isolates of *C. albicans* were spiked into stool and the RNA methodology followed. The *C. albicans* species were successfully detected by PCR-DGGE.

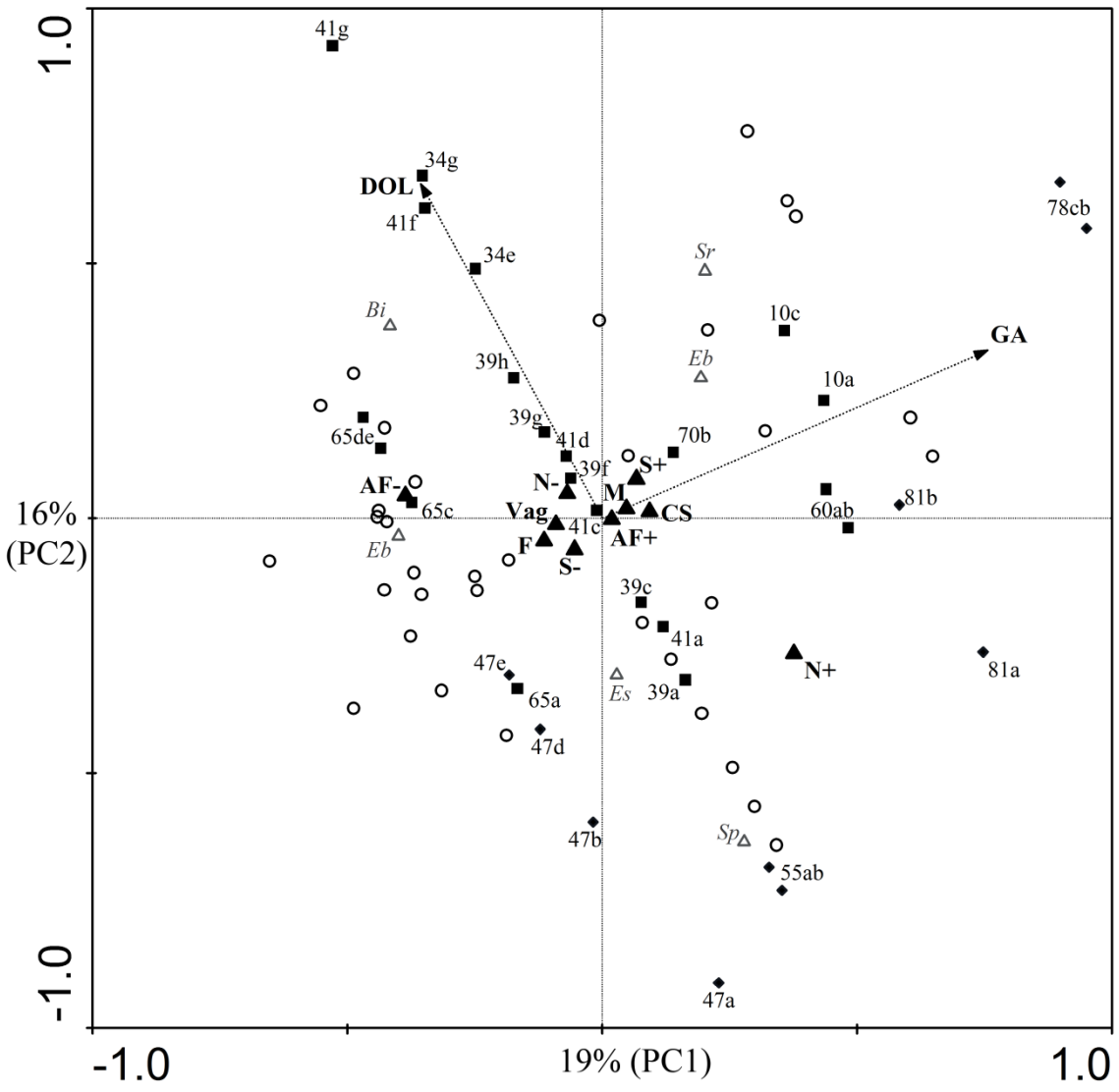


Figure 5.4 – CCA based on the viable bacterial community. Comparison of patient profiles (Healthy patients (○), patients diagnosed with NEC (♦), and patients diagnosed with sepsis (■)) with discrete variables (▲) including; N = NEC (*P*=0.188), S = sepsis (*P*=0.004), AF = antifungal treatment (*P*=0.144), CS = caesarean / Vag = vaginal birth (*P*=0.366), M = male / F = female (*P*=0.166) and continuous explanatory variables (--►) including; GA = gestation age (*P*=0.002), DOL = day of life (*P*=0.004). Sequenced bands (∆) include - Bi = *Bifidobacteria*, Eb = *Enterobacter*, Es = *Escherichia*, Sp = *Sphingomonas*, Sr = *Streptococcus*

Table 5.2 – Comparison of the *P* Values from the total and viable CCA

Variable	Total CCA (*P* Value)	Viable CCA (*P* Value)
NEC	0.002	0.188
Sepsis	0.002	0.004
Antifungal treatment	0.03	0.144
Delivery mode	0.01	0.366
Gender	0.012	0.166
Gestation age	0.148	0.002
DOL[a]	0.002	0.004

[a]DOL – Day of life

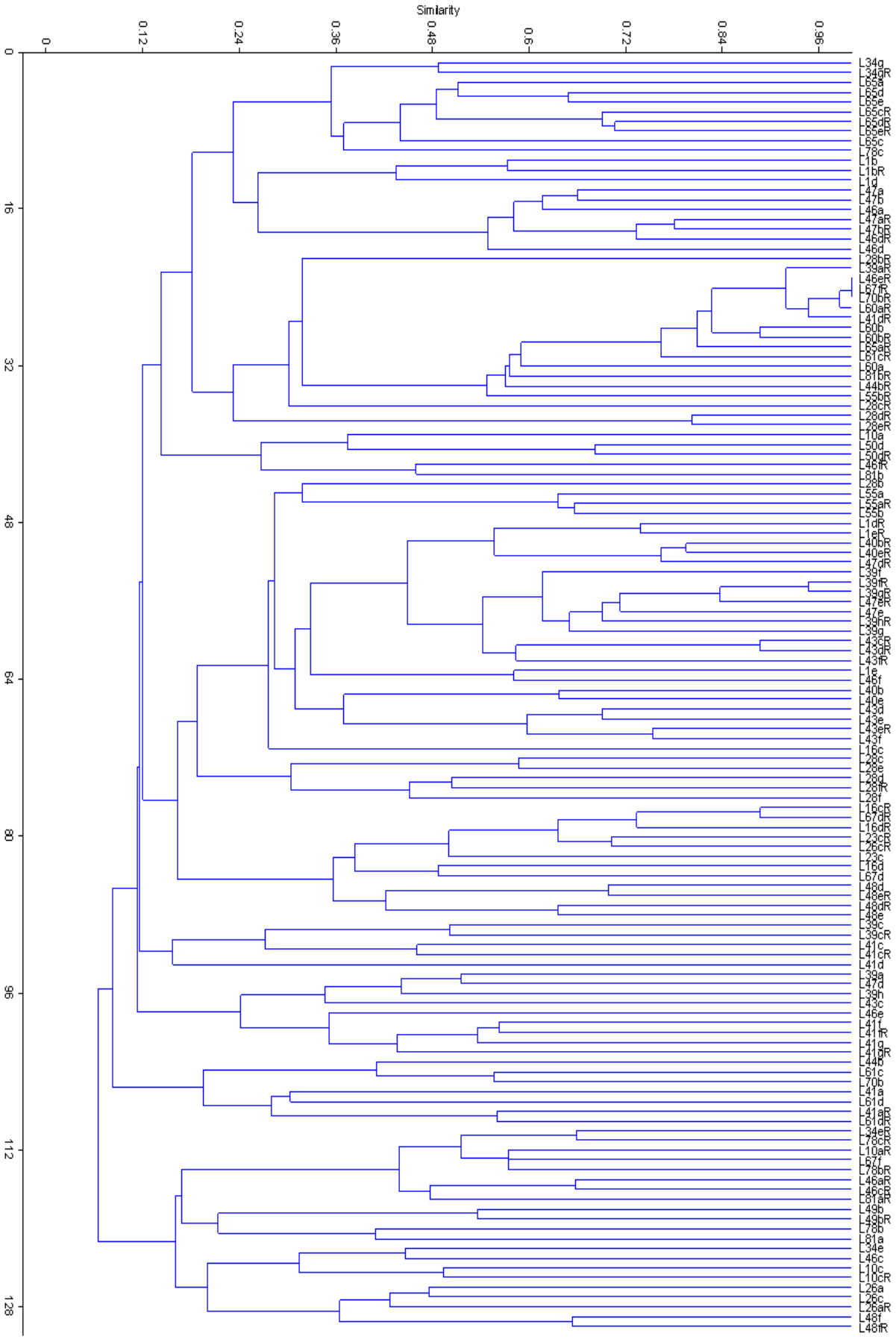


Figure 5.5 – Cluster analysis based of both DNA matched to the RNA samples from the subset of the cohort. Dendrogram based on the Bray-Curtis coefficient.

5.2.4 Effects of increasing age on the total and viable bacterial communities

Total bacterial samples were matched with the corresponding viable sample to compare changes with increasing age. Diversity increased from week 1 of life, although this increase was not continuous, with fluctuations in the bacterial community structure occurring throughout the first 9 weeks (Fig 5.6). Trends were similar for total and viable communities but numbers of bacteria deemed viable were lower than that of the total community. Overall the diversity and richness of the samples increased over the first 9 weeks correlating with reduced antibiotic exposure as well as increasing age, but a reduction was noted in week 4 when the diversity was more established and antibiotic administration was still relatively high. Numbers of samples available each week were variable preventing further statistical analysis of this current cohort.

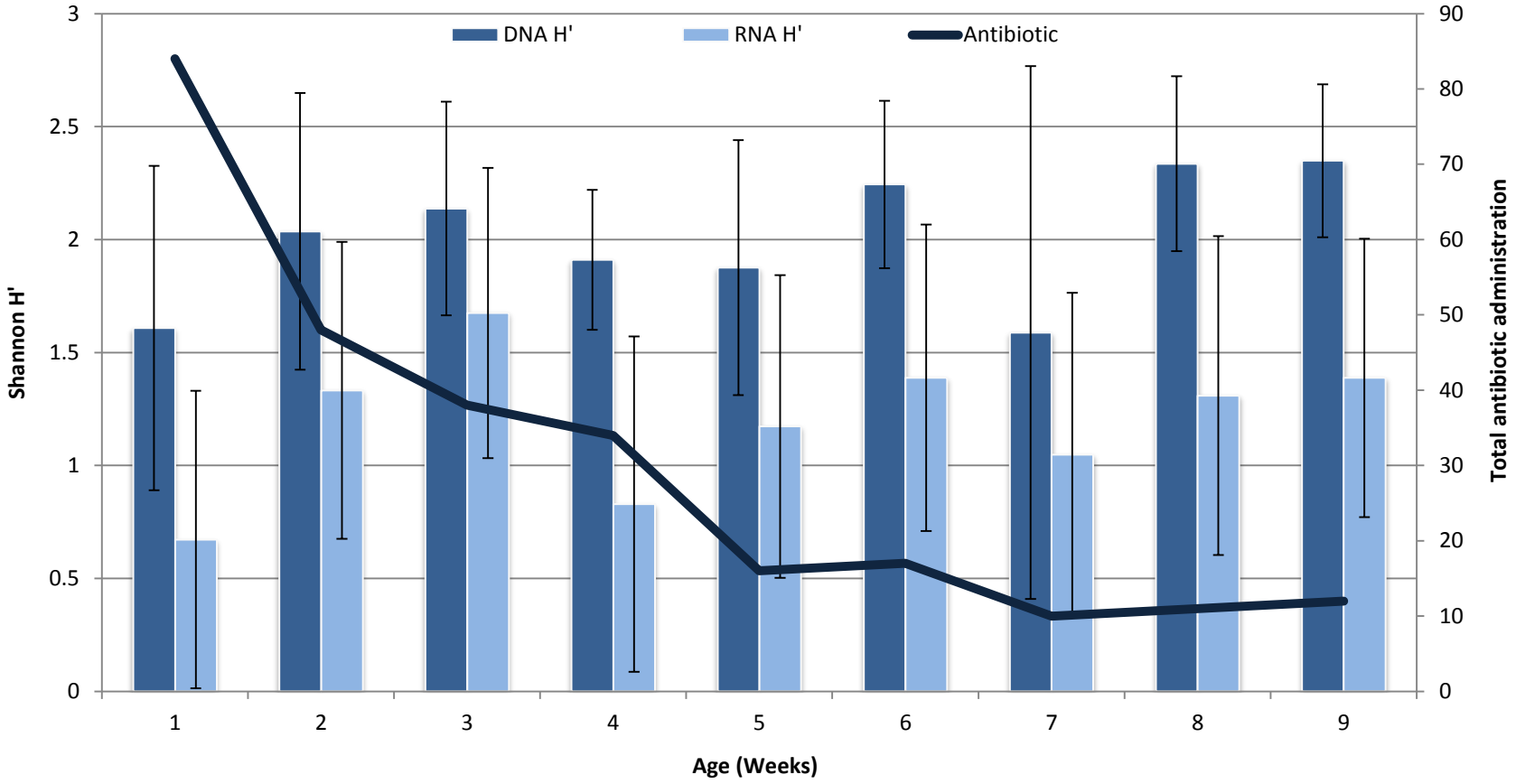


Figure 5.6 - A comparison of the total and viable bacterial Shannon diversity (H') with the total number of antibiotics administered each week. Error bars represent the standard error of mean.

5.3 Discussion

We have explored two novel aspects of the preterm faecal microbiota: fungal colonisation and the role of organism viability. Total bacterial profiles of infants with NEC and total and viable profiles of infants with sepsis were significantly different from healthy infants, supporting a role for bacterial colonisation in the pathophysiology of these diseases (Stenger *et al.* 2011). This is a descriptive study not a case-control study, which aimed to describe variables related to the development of the gut microbiota in preterm neonates over a period of time. Subsequent analyses will be subject to influence by practices within the NICU that confound associations, making the direction of effect unclear and the apparent effect of the disease states may not be causal. However, when examining samples pre and post disease diagnosis, using PLS-DA we observed that samples before disease onset still group separately from healthy patients (Wang *et al.* 2009). This suggests differences in the gut microbiota predate the onset of disease.

Importantly and to our knowledge uniquely, *Sphingomonas spp*. colonisation was significantly associated with NEC in both the total and viable analysis. This genera belongs to the Proteobacteria phylum, recently reported to increase before NEC onset (Carlisle *et al.* 2011; Mai *et al.* 2011). Sphingomanads have previously been identified within biofilms formed in water supply systems although this would appear to be an unlikely route of colonisation within NICU where most water to which infants are exposed is sterilised (Koskinen *et al.* 2000). Further work is needed to determine if the association of *Sphingomonas* spp. with NEC remains significant in a larger cohort and elucidate mechanisms of pathogenicity, which will include sampling the neonatal intensive care environment.

Unlike other studies we have shown no significant difference between the number of total bacterial species, whether total or viable, in populations of infants with NEC or sepsis compared with healthy infants. Differences in our findings between other studies may depend on the timing of sampling in relation to disease onset (Wang *et al.* 2009; Smith *et al.* 2012). Surprisingly gestation only influenced viable data, and birth weight was never found to have a significant effect on the gut microbiota, despite the well-recognised importance of these factors for the development of NEC (Stoll 1994; Guthrie *et al.* 2003; Lin *et al.* 2008; Caplan & Frost 2011). The most prevalent genus of bacteria identified in our total and viable analyses (*Enterococcus*, *Streptococcus*, and *Escherichia*) agreed with those previously identified (Schwiertz *et al.* 2003; Magne *et al.* 2005; Mshvildadze *et al.* 2010; LaTuga *et al.* 2011). Interestingly, *Staphylococcus* spp. were not found amongst the most common organisms despite these taxa contributing to the majority of positive blood cultures and being previously reported as a prevalent genus in the gut (Jacquot *et al.* 2011; Madan *et al.* 2012). This may reflect specific antibiotic preferences used within individual units.

Fungal colonisation was identified in half the infants but no viable fungi were detected in this cohort with almost universal fluconazole use. The stability of fungal profiles within individuals suggests that fungal DNA persists in the gut long after fluconazole prophylaxis due to its fungistatic properties successfully inhibiting replication. Although small in number (n = 2), the infants who did not receive fluconazole had significantly different total bacterial profiles, although again this could be confounded by their clinical differences that resulted in them not requiring fluconazole. Interestingly, fluconazole has direct antibacterial properties especially against Gram positives (Sud & Feingold 1982). There may also be community effects from liberating niches that would otherwise be occupied by fungi. For example, *Pseudomonas aeruginosa* colonisation increases in the murine lung following antifungal drug

administration, which has important implications in lung injury (Ader *et al.* 2011). Despite fluconazole use identified fungal species were dominated by *Candida* spp (Fairchild *et al.* 2002; Cahan & Deville 2011). Consistent with previous reports, we also did not find a correlation between fungal colonisation and mortality or NEC (Cahan & Deville 2011).

A major benefit of assessing the viable community rather than just the total community is that it potentially gives insight into taxa that are driving major metabolic activities and participating in microorganism/host interactions, and thus may give insight into strategies to alter outcomes (Tannock *et al.* 2004; Peris-Bondia *et al.* 2011). However, the importance of separately assessing the viable community within preterm infants is currently unclear as no previous studies have explored viability. We found important differences in the variables deemed to be significantly driving bacterial community structure based on either total or viable analysis. While sepsis was found to be significant in both analyses, gestational age was only significant for the viable community and gender and birth mode were only significant for the total community.

Analyses incorporating the viable portion of the gut microbiota may gain increasing importance when assessing potentially important gut microbiota manipulations. Of current interest are the use of probiotics (Embleton & Yates 2008), prebiotics (Szajewska 2010) and lactoferrin (Venkatesh & Abrams 2010), however the mechanism of action and effect of such treatments on the gut microbiota require further research (Shanahan 2010).

LaTuga *et al.* recently showed a low to moderate total bacterial diversity with a mean H' of 1.02 from a cohort that was younger with a lower gestation compared to ours (LaTuga *et al.* 2011). We report relatively high total bacterial diversity (mean H' of 2.04) but our associated viable bacterial diversity was lower with a mean H' of 1.18. It has also been shown that antibiotic administration decreases the numbers of anaerobic

bacteria in the gut microbiota, with counts of bifidobacteria particularly reduced (Penders *et al.* 2006). Our results support this with bifidobacteria being detected in 8.8% of the total community and only 3.1% of the viable community. From week 5 onwards both the total and viable community generally show an increasing diversity and richness, by which stage the majority of patients were no longer receiving antibiotic treatment.

Molecular based research into the association of the total bacterial community with NEC has increased our understanding of the disease: the complex multifactorial pathophysiology appears to be influenced by a variety of bacterial genera, individually, or promoting shifts in communities. Stool samples allow for non-invasive elucidation of the gut microbiota; however the stool microbiota may not precisely represent the gut microbiota (Durbán *et al.* 2011). The stool microbiota is thought to represent a combination of shed mucosal bacteria and a separate nonadherent luminal population (Eckburg *et al.* 2005). Employing molecular techniques allows greater coverage of the microbial community with only 20% of the gut microbiota reported to be cultivable (Eckburg *et al.* 2005). This may allow detection of important species not readily cultivated such as *Spingomonas* spp. here associated with NEC. High throughput next generation sequencing platforms, such as 454 pyrosequencing, are becoming increasing employed in clinical research due to the detection of low abundance taxa (Sekirov *et al.* 2010). However, as in DGGE, pyrosequencing is also subject to PCR bias (Petrosino *et al.* 2009).

This novel study employing a relatively large cohort helps to further elucidate total as well as viable organisms of the gut microbiota in association with NEC and sepsis. We show abnormal bacterial colonisation in association with the development of NEC and sepsis, with colonisation by *Sphingomonas* spp. significantly associated with NEC. While antifungal prophylaxis significantly affects the total bacterial community, the

presence of fungal species in the gut was not demonstrated to affect bacterial richness. Further work is needed to investigate the role of community microbial dynamics in the pathophysiology of NEC and infection, while additional exploration of the total and viable communities may add further to our understanding.

6. Development of the preterm gut microbiota in twins at risk of necrotising enterocolitis and sepsis

Abstract

Aim: The preterm gut microbiota is a complex dynamic community influenced by genetic and environmental factors and is implicated in the pathogenesis of NEC and sepsis. We aimed to explore the longitudinal development of the gut microbiota in preterm twins to determine how shared environmental and genetic factors may influence temporal changes and compared this to the expressed breast milk (EBM) microbiota.

Methods: Stool samples (n = 173) from 27 infants (12 twin pairs and 1 triplet set) and EBM (n = 18) from 4 mothers were collected longitudinally. All samples underwent PCR-DGGE (denaturing gradient gel electrophoresis) analysis and a selected subset underwent 454 pyrosequencing.

Results: Stool and EBM shared a core microbiota dominated by *Enterobacteriaceae*, Enterococcaceae, and Staphylococcaceae. The gut microbiota showed greater similarity between siblings compared to unrelated individuals. Pyrosequencing revealed a reduction in diversity and increasing dominance of *Escherichia* sp. preceding NEC that was not observed in the healthy twin. Antibiotic treatment had a substantial effect on the gut microbiota, reducing *Escherichia* sp. and increasing other *Enterobacteriaceae*.

Conclusions: This study demonstrates related preterm twins share similar gut microbiota development, even within the complex environment of neonatal intensive care. This is likely a result of shared genetic and immunomodulatory factors as well as exposure to the same maternal microbiota during birth, skin contact and exposure to EBM. Environmental factors including antibiotic exposure and feeding are additional significant determinants of community structure, regardless of host genetics.

6.1 Introduction

The gut microbiota is crucial to both health (immunomodulation, protection, nutrition and metabolism) and disease (inflammation, diabetes, obesity and allergy) (Sekirov *et al.* 2010). Due to the complexities of both the microbial community and factors that affect it, exploring individual variables (including diet, medical interventions/exposures and genetic components) is challenging. Studying twins or higher order multiples may provide unique insights, with previous studies suggesting the gut microbiota may be subject to host genetics. Healthy twins have been shown to develop a comparable gut microbiota after term birth (Palmer *et al.* 2007), in childhood (Stewart *et al.* 2005) and into adulthood (Zoetendal *et al.* 2001), suggesting genetic or shared environmental factors shape the gut community. Existing evidence is confounded by the genetic diversity of humans as well as strong environmental effects (Benson *et al.* 2010). Twin studies, therefore, offer important insights into the significance of the host genetic background in affecting GIT microbiota development. Nevertheless, there is currently a lack of research exploring the temporal changes of the gut microbiota in preterm twins.

Preterm neonates provide a unique cohort to study the dynamics of the gut microbiota due to intensive care practises and the susceptibility of these infants to complex disease. For example, NEC and sepsis are complex diseases which together affect over 20% of all preterm infants and are associated with differential microbiota development compared to term infants (Berrington *et al.* 2013). Understanding how heritable traits affect the gut microbiota may help elucidate the interactions influencing disease states (Benson *et al.* 2010). Studying twins with NEC or sepsis may help elucidate the role of specific exposures that may be key to reducing incidence of these diseases. Indeed, Benson *et al.* (2010) hypothesised that genetic heritability to complex diseases might involve the predisposition to particular patterns of microbial colonisation.

For term infants vaginal delivery and receipt of maternal breast milk are key factors that facilitate the development of a 'healthy' microbiota. Breast milk contains many immunomodulatory factors that support growth and prevent infection including lysozyme, lactoferrin, and oligosaccharides as well as live bacteria which regulate host-microbe interactions (Cabrera-Rubio & Collado 2012) and modify infant gut microbiota development (Martín *et al.* 2009). Preterm infants are less likely to experience vaginal birth or only receive breast milk feeds and are more likely to experience many medical interventions that affect the microbiota. To what extent the infant gut microbiota reflects the maternal breast milk microbiota is currently unclear. However, due to a lack of similarity with siblings from a different birth and the high similarity observed in both monozygotic and dizygotic twins, it is conceivable that the environmental exposures including diet is influential in shaping the neonatal gut microbiota (Palmer *et al.* 2007).

In a twin cohort we aimed to explore the longitudinal development of the gut bacterial community after preterm birth by analysis of stool and EBM. In addition we focused on the development of NEC in one set of twins with regular longitudinal sampling and where only one infant was diagnosed with NEC.

6.2 Results

6.2.1 Patients and samples

Demographic information from each patient is summarised in Table 6.1 and further details based on each individual sample can be found in the appendix (Appendix 10 - disc). In total, 27 patients (12 twin pairs and 1 triplet set) contributed to the study. No patients received probiotics or prebiotics. Five patients developed NEC with one case

causing fatality (patient 29); all NEC cases were treated by medical intervention with non-requiring surgery. Five patients developed sepsis (two of whom also had NEC) with two infants having multiple episodes and one case causing fatality (patient 92). Cultured organisms included CoNS, *Klebsiella pneumonia*, *Staphylococcus aureus*, and *Pseudomonas.* Specific genotyping information was not available and therefore the classification of twin zygosity was based on chorionicity data: different sex twins are dizygotic, same-sex twins are either monochorionic (monoaminotic (1 sac) or diamniotic (2 sacs)) which are always monozygotic, or dichorionic diaminotic which could be monozygotic or dizygotic. For the triplets, two (patients 145/148) were monochroionic monoaminotic and the other (patient 147) was born in a different sac.

Table 6.1 Summary of patient demographics

Pat. No.	Del. Mode	GA (week)	Birth Wt (g)	Sex	Chorionicity	NEC	NEC Diag. (DOL[a])	BC[b] +	BC[b] (DOL[a])	Organism
22	CS	27	870	M	Dichorionic Diamniotic	N		N		
23		27	790	F		N		N		
28	CS	28	1250	M	Monochorionic Diamniotic	N		N		
29		28	1180	M		Y	17	N		
39	CS	25	780	M	Dichorionic Diamniotic	N		Y	15	CoNS[d]
41		25	820	M		N		Y	15	*S.aureus*
46	CS	26	830	M	Dichorionic Diamniotic	N		N		
47	Vag	26	760	F		Y	45	Y	3 / 8	CoNS[d]/ *Pseudomonas* sp.
51	CS	27	1060	M	Dichorionic Diamniotic	Y	16	Y	17 / 56	CoNS[d] / CoNS[d]
55		27	1100	M		N		N		
68	Vag	26	760	M	Dichorionic Diamniotic	N		N		
70		26	860	M		N		Y	40	CoNS[d]
92	CS	25	740	M	Dichorionic Diamniotic	N		Y	8	*K.pneumonia*
93		25	670	M		N		N		
100	CS	27	1050	M	Monochorionic monoamniotic	N		N		
101		27	910	M		N		N		
112	Vag	25	700	F	Dichorionic Diamniotic	N		N		
113		25	680	M		N		Y	57	CoNS[d]
135	CS	29	910	F	Monochorionic Monoamniotic (TTTS)[c]	N		Y	54	*S.aureus*+ *K.pneumoniae*
136		29	1275	F		N		N		
139	CS	30	1470	M	Dichorionic Diamniotic	Y	28	N		
140		30	1455	F		N		N		
145	CS	31	990	M	Monochorionic Monoamniotic	N		N		
148		31	1455	M		N		N		
147		31	1540	M	Dichorionic Diamniotic	N		N		
151	Vag	27	1020	F	Dichorionic Diamniotic	N		N		
154		27	1060	M		Y	21	N		

[a]Day of life. [b]Blood culture. [c]TTTS – Twin to Twin Transfusion Syndrome. [d]CoNS – Coagulase negative *Staphylococcus*.

6.2.2 Stool profiles

PLS-DA of all samples from all patients based on DGGE data demonstrated twin pairs had comparable profiles which were distinct from unrelated individuals (Fig 6.1A). Samples which fell outside the ellipse (indicating Hotellings T^2 range, at 95% confidence) belonged to the triplets or were associated with late onset infection. Specifically, this comprised sample 139t collected one day post NEC diagnosis as well as samples 39c and 41b collected three days prior and on the day of sepsis diagnosis, respectively. Based on DGGE analysis of the cohort (Fig 6.1A), a subset of longitudinal samples from the triplets (represented by red squares) and twin pair discordant for NEC (represented by orange circles), which showed significant variation in community development, were selected for pyrosequencing. The pyrosequencing data was analysed at the genus level using weighted UniFrac.

The pyrosequencing data was based on a subset of stool samples with focus on the triplets and twin pair discordant for NEC (Appendix 10). Rarefaction curves generated from the pyrosequencing data show for the majority of samples the curves surpass exponential phase and are plateauing out (Fig 6.2). Therefore, the majority of diversity within these samples has been captured, but the rarer lower abundant OTUs may not have been detected. The pyrosequencing profiles were analysed using weighted UniFrac. In accordance with the DGGE data, samples grouped with their related twin, showing high intra-sample similarities in the development of the gut microbiota (Fig 6.1B). This is reflected in the bar plots which show each set of twins and the triplets developed a distinct gut microbiota (Fig 6.3). *Proteobacteria* and *Firmicutes* dominated samples in both DGGE and pyrosequencing. From pyrosequencing, 5 OTUs were found in the core microbiota, that is genera present in over 85% of in stools (Seekatz *et al.* 2013), from the families *Enterobacteriaceae*, *Enterococcaceae*, and *Staphylococcaceae*.

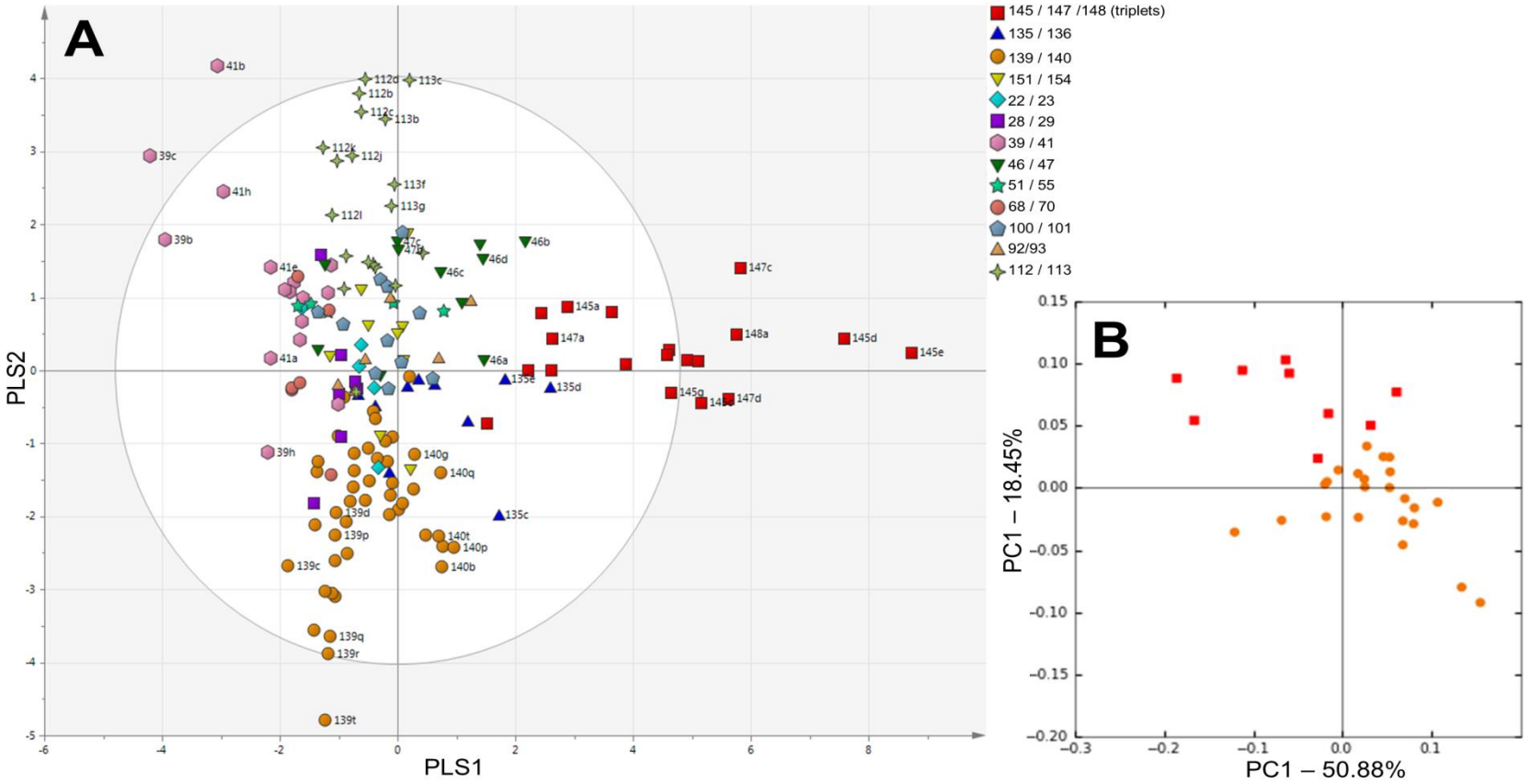


Figure 6.1 - Community profiles of gut (stool) microbiota from preterm multiples. Subjects are symbolised based on related multiples. A) Partial least squares discriminant analysis (PLS-DA) based on DGGE data of all stool samples. The ellipse indicates Hotellings T^2 range, at 95% confidence. Selected sample labels removed for clarity. B) Weighted UniFrac PCoA based on pyrosequencing data at genus level, generated from a subset of stool samples. Triplet samples (patients 145, 147, and 148; red squares) and discordant twins (patients 139 (NEC) and 140; orange circles).

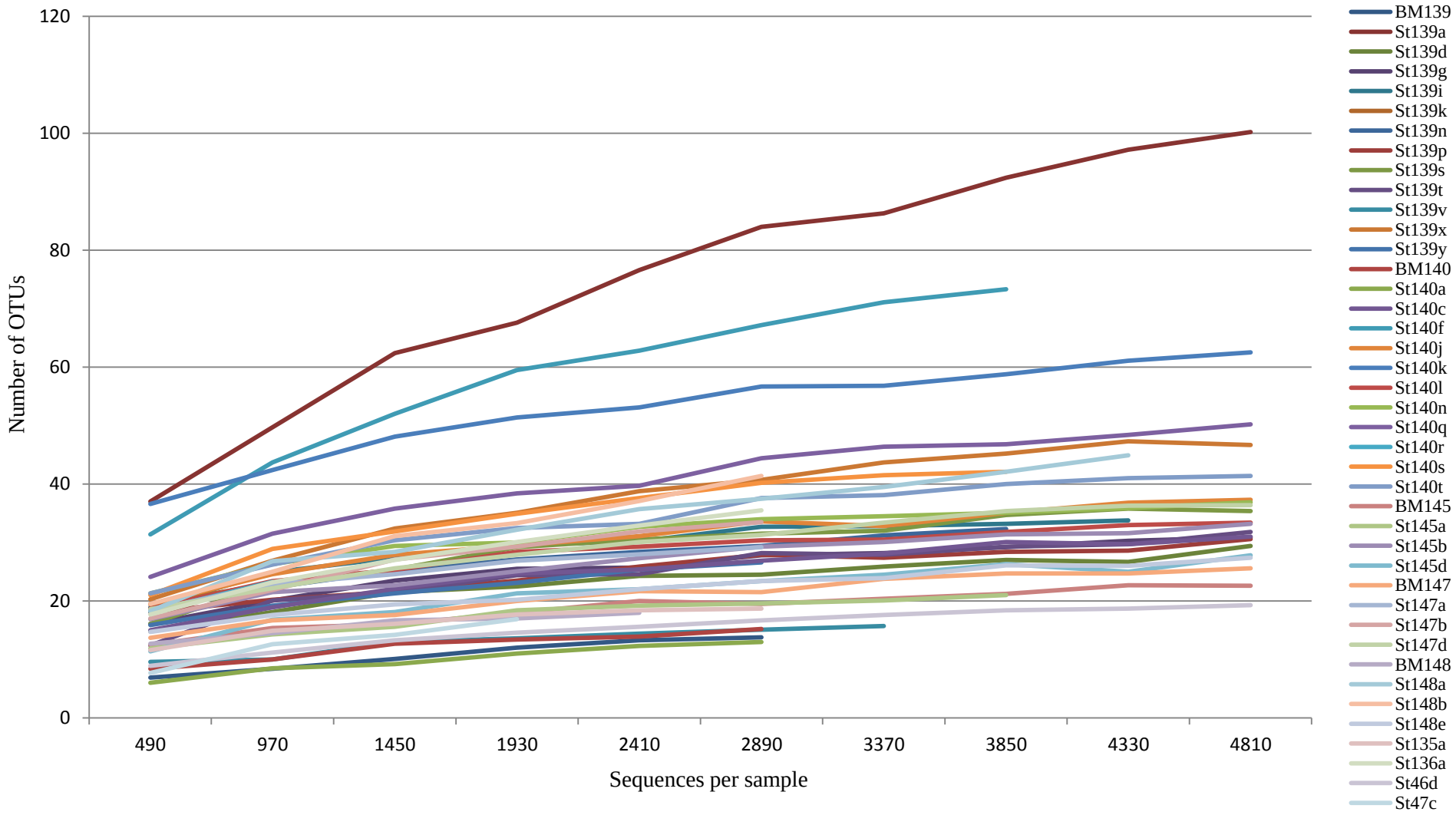


Figure 6.2 - Rarefactions curves. Produced in QIIME to 5000 sequences comparing all samples (stool and expressed breast milk).

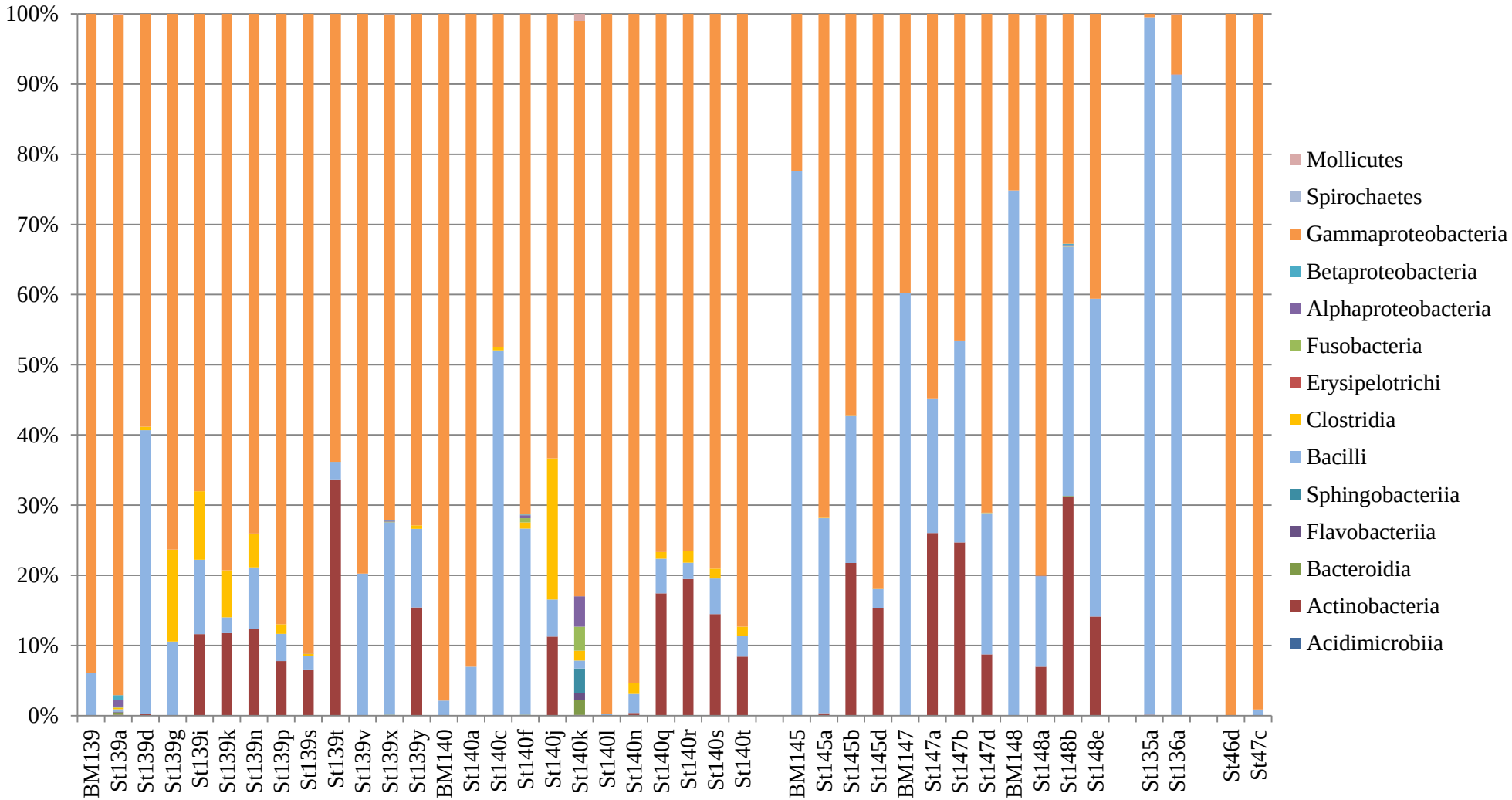


Figure 6.3 - Order level bar plot of all samples (stool and expressed breast milk) which underwent 454 pyrosequencing.

6.2.3 Comparison of breast milk with respective stool samples

The DGGE profiles of all extracted EBM samples were compared with the respective stool sample. EBM samples showed a relative lack of diversity making them cluster near the origin: despite this PLS-DA did reveal EBM samples clustered with the stool samples of the respective set of multiples (Fig 6.3A). There are cases where different EBM samples from the same mother cluster separately showing the EBM microbiome was not stable (e.g. BM148e / BM147b / BM145a / BM145i or BM139b / BM139f / BM139q).

EBM from the triplet set and twin set (patients 139/140) also underwent pyrosequencing. EBM samples from each mother showed high intra-sample similarity (Fig 6.3B). For twins 139/140 the EBM profiles (BM139/BM140) clustered closely with the respective stool (139s/140s). This clustering was less robust in the triplet set although triplet EBM (BM145/BM147/BM148) was still more similar to triplet stool (145a/147b/148b) than stool of other infants. Noteworthy, due to the inclusion of only two sets of multiples, principal component 1 (74%) represents a large amount of the variance. The similarity of EBM samples with stool can be visualised in the bar plot (Fig 6.2). Three OTUs were found in the core microbiota in EBM and, like in the stool core microbiota, were from families *Enterobacteriaceae*, *Enterococcaceae*, and *Staphylococcaceae*.

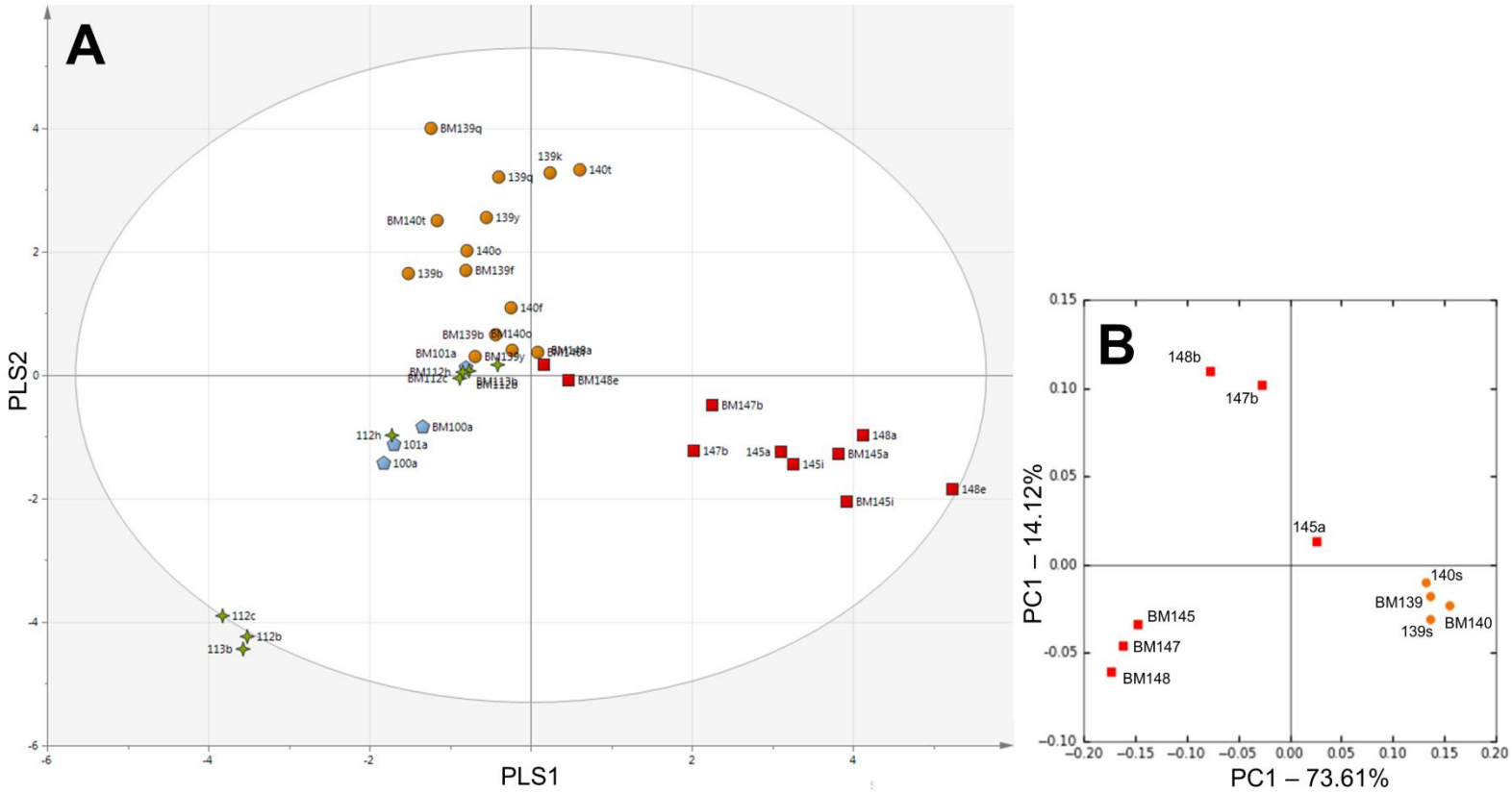


Figure 6.4 - Comparison of breast milk with respective stool profiles. Subjects are symbolised based on related multiples. A) Partial least squares discriminant analysis (PLS-DA) based on DGGE data of all breast milk (EBM) samples matched to respective stools. The ellipse indicates Hotellings T^2 range, at 95% confidence. B) Weighted UniFrac PCoA based on pyrosequencing data at genus level, generated from a subset of EBM and respective stool samples.

6.2.4 Comparison of sequenced DGGE bands

All the DGGE bands which were excised and sequenced (n = 17) corresponded to 6 different bacterial genera; *Actinomyces*, *Corynebacterium*, *Enterobacter*, *Enterococcus*, *Klebsiella*, and *Propionibacterium* (Appendix 10). The richness of these sequenced OTUs were compared at genus level with development of NEC or sepsis. *Corynebacterium* and *Enterococcus* were significantly (P = 0.001) more abundant prior to NEC diagnosis, with the latter being present in much greater abundance across all groups (Fig 6.5A). Levels of *Actinomyces* was shown to be significantly (P = 0.001) higher following NEC diagnosis. Only *Propionibacterium* was found to be significantly affected by sepsis with the richness increasing significantly (P = 0.001) following diagnosis (Fig 6.5B).

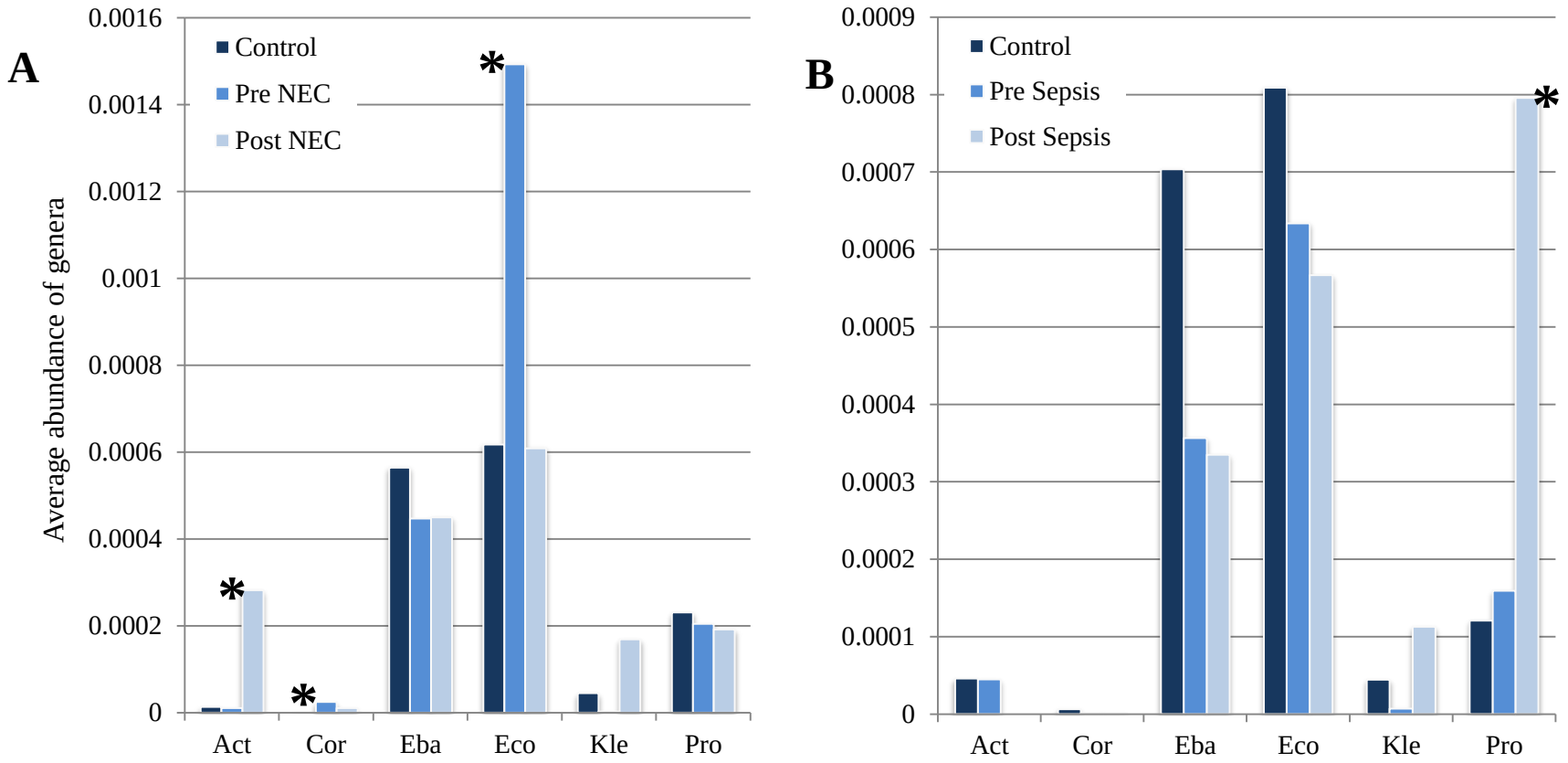


Figure 6.5 – Bar plot showing the average abundance of sequenced DGGE bands. Abundance based on DGGE band intensity. Act - *Actinomyces*, Cor - *Corynebacterium*, Eba - *Enterobacter*, Eco - *Enterococcus*, Kle - *Klebsiella*, Pro - *Propionibacterium*. A) Abundance of each genus in NEC. B) Abundance of each genus in sepsis. * Denotes significance.

6.2.5 Gut microbiota in NEC (twin set 139/140)

A reduction in diversity was observed in patient 139 at least 5 days before NEC was apparent clinically (Fig 6.6A), which was not shared by the twin. A less pronounced reduction in diversity did occur in the control twin (140) earlier in development (day 18) that coincided with antibiotic administration and diversity was quickly re-established when antibiotics ceased, a recovery not seen in the infant with NEC over the same time scale.

To gain insight into the taxa involved in this dysbiosis, matched samples were selected for 454 pyrosequencing based on prominent changes in the DGGE profiles. Pyrosequencing of matched twin samples (full profiles in the bar plot; Fig 6.3) incorporated the 12 most abundant OTUs into the analysis (Fig 6.6). These results are in accordance with the DGGE data and allow greater detail about specific contributors to the overall diversity to be observed. While community structure in the twins was initially comparable, twin 139 showed reducing diversity with an increasing abundance of *Escherichia* sp., before NEC diagnosis on day 28 (Fig 6.6B). Conversely, in the sibling there is an increase in diversity and *Escherichia* sp. was present in much lower abundance (Fig 6.6C). After antibiotic receipt (day 16 in twin 140 and day 28 in twin 139) both twins demonstrate reduced *Escherichia* sp. abundance and an increased abundance of other *Enterobacteriaceae*, rapidly reversing in twin 140 on antibiotic cessation. In twin 139 NEC and subsequent antibiotic treatment significantly (P = 0.028) altered the bacterial community in comparison to its sibling, with a new notable bloom in *Klebsiella* sp. and a smaller increase in *Parascardovia* sp. (family – *Bifidobacteriaceae*).

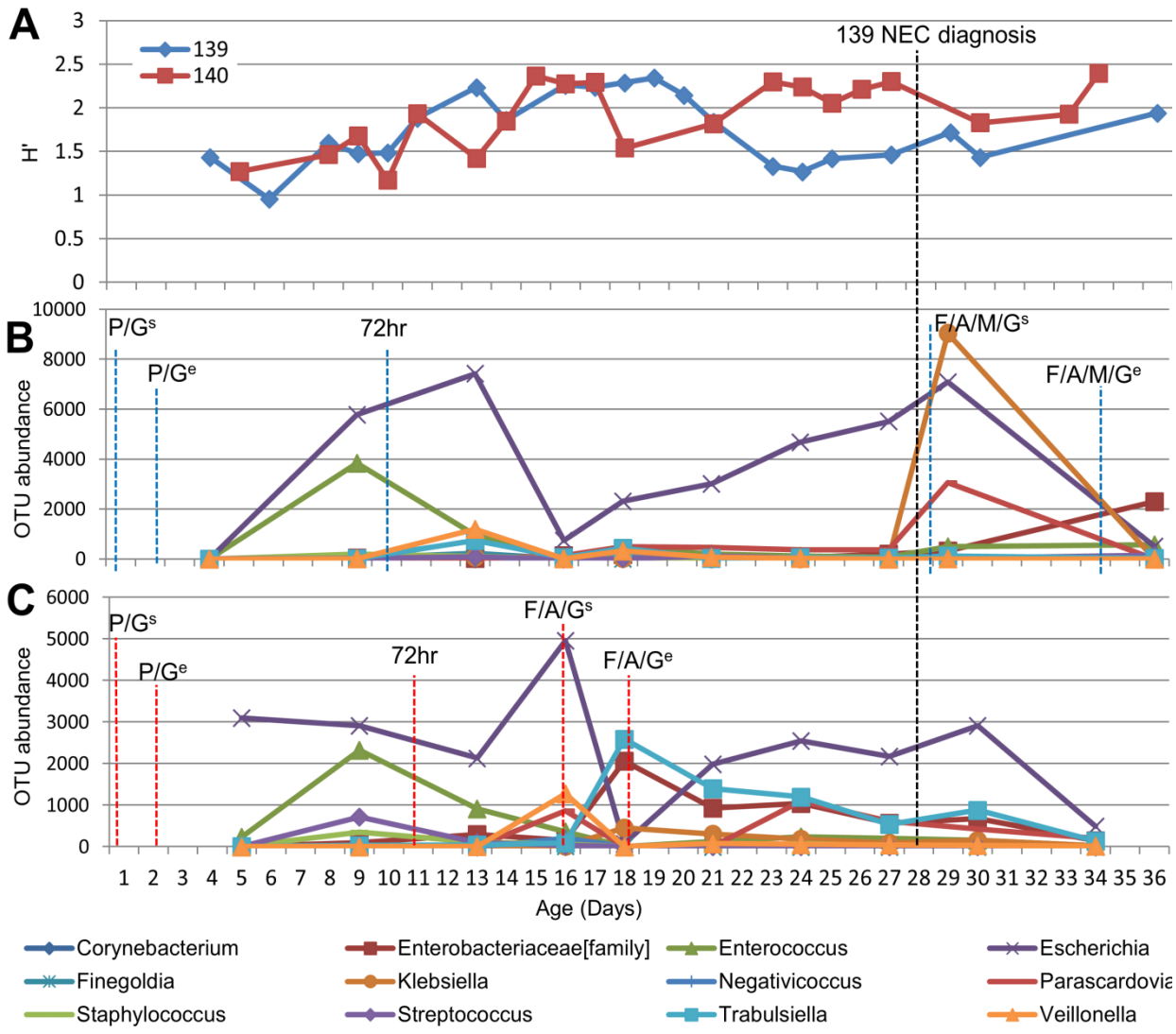


Figure 6.6 - Development of gut microbiota in twin pair 139/140 mapped to life events.

P – Penicillin, G – Gentamicin, F- Fluctoxacillin, A - Amoxycillin, M – Metronidazole, ^{s} – Start of antibiotics, ^{e} – End of antibiotics, 72hr – full enteral feed (at least 150 ml/kg/day) sustained for 72 hours. a) Shannon Diversity indices (*H'*) of twin pair based on DGGE data. b) Turnover of the most prevalent bacterial OTUs throughout the first 36 days of life in twin 139 where antibiotics were prescribed for NEC. c) Turnover of most prevalent bacterial OTUs throughout the first 34 days of life in twin 140 where antibiotics were prescribed due to pyrexia (fever).

6.3 Discussion

In this study, we hypothesised that the bacterial community in related twins and a set of triplets would be comparable and reflect maternal EBM community. In cases of twin pairs discordant for NEC, we hypothesized that they would show differences in microbial community development prior to NEC onset.

We have demonstrated in a preterm population with multiple clinical exposures that the development of the gut microbiota is more similar between genetically related individuals than between other preterm infants. However, due to similar environmental exposures encompassed by related individuals, this may not be a direct result of host genetics. Interestingly, community structure was similar for all triplets even though two (145/148) were monochorionic monoamniotic and the other (147) was dichorionic. This suggests that shared factors (genetic or environmental) are important in determining the gut microbiota even in an environment with many complex variable factors that also affect community development (Ley *et al.* 2005; Palmer *et al.* 2007).

Proteobacteria and *Firmicutes* dominated the gut microbiota over the initial weeks of life as previously reported (Mshvildadze *et al.* 2010). Stool and EBM shared a core microbiota of the families *Enterobacteriaceae*, *Enterococcaceae*, and *Staphylococcaceae*. Others describing the EBM microbiota noted the presence of *Streptococcaceae* which in was present in low abundance in our maternal cohort (Collado *et al.* 2009; Hunt *et al.* 2011). We also detected a low abundance of bifidobacteria and lactobacilli in stool and EBM despite their reported prevalence by others (Penders *et al.* 2006; Collado *et al.* 2009; Martín *et al.* 2009). This may be attributable to differences among subjects, unit flora, and detection by differing techniques (Hunt *et al.* 2011), but occurred despite the use of optimised universal

primers in pyrosequencing designed to facilitate the detection of Bifidobacteria (Sim *et al.* 2012). This might be attributable to differences between cohorts, with significantly preterm infants having delayed colonisation with bifidobacteria; geographical or demographic differences may also account for the low prevalence of this organism (Palmer *et al.* 2007). In accordance with the results of this study, previous studies have shown that the milk microbiota is not stable throughout lactation and EBM appeared to be an ongoing source of new flora contributing to the dynamic nature of the bowel microbiota (Cabrera-Rubio & Collado 2012). Furthermore, genera which typically reside on adult skin including *Staphylococcus*, *Corynebacterium*, and *Propionibacterium* were found in high abundance in the gut microbiota suggesting skin contact may be an important source of bacterial acquisition, even within the nursery environment (Cogen *et al.* 2008).

The importance of the gut microbiota in disease is increasingly recognised, despite a lack of consistent causative agent between studies. DGGE analysis of the whole cohort showed levels of *Corynebacterium* and *Enterococcus* were significantly elevated prior to NEC diagnosis. This is contrary to our routine culture based analysis where *Enterococcus faecalis* was more commonly isolated from control infants (Chapter 4, section 4.2.2). This suggests that molecular based approaches, which take species abundance into account, may offer important insights otherwise missed by traditional approaches. Interestingly, *Enterococcus faecalis* has recently been suggested as a potential biomarker in predicting NEC (Braniste & Pettersson 2012). Following diagnosis of NEC and subsequent antibiotic administration, the community showed elevated levels of *Actinomyces*.

While there was no significant dominance of a genus prior to sepsis, *Propionibacterium* was significantly more abundant following diagnosis and antibiotic treatment. This is in accordance with recent work, where dysbiosis of the normal microbiota led to sepsis

and not enrichment of a potential pathogen (Mai *et al.* 2013). Interestingly, following antibiotic treatment for either NEC or sepsis, members of the Actinobacteria phylum were significantly more abundant; correspondingly, this phylum is usually associated with control infants (Mai *et al.* 2011).

In a focused temporal exploration from a twin pair (139/140) discordant for NEC, we showed clear changes attributable to antibiotic exposure and NEC development, with effects on the dominance of *Escherichia* sp. and the abundance of other *Enterobacteriaceae* sp. (Madan *et al.* 2012). The significantly different community observed in sample 139t is likely attributable to a temporary bloom in *Klebsiella* sp. following NEC diagnosis, which was reduced in the subsequent sample following broad spectrum antibiotic administration. While there are few comparable studies in NEC, a twin study methodology has been utilised in other IBD, such as ulcerative colitis and Crohn's disease. Interestingly, analogous to NEC, dysbiosis is a major factor in the pathogenesis of these diseases, consistent with a lack of a single causative agent (Lepage *et al.* 2011). Specifically, a decreased diversity in the gut microbiota of ulcerative colitis and Crohn's disease patients compared to healthy controls has been noted, (Gophna *et al.* 2006; Wang *et al.* 2009) and multiple studies report increased abundances of *Proteobacteria*, particularly *Escherichia* sp. (Darfeuille-Michaud *et al.* 2004; Sasaki *et al.* 2007). This increase in *Proteobacteria* is an emerging theme in NEC pathogenesis (Mai *et al.* 2011), but is probably one of several factors needed for NEC development. *Escherichia* spp. are reported pathogens (Kaper *et al.* 2004) and the association of this genus with inflammatory mediated disease warrants further investigation.

Our use of molecular approaches for community profiling circumvents the known limitations of culturing human gut species (Eckburg *et al.* 2005). We utilised cost effective DGGE to educate sample selection for 454 pyrosequencing. Studies on

preterm infants, especially of multiple births, are difficult due to the exclusivity of the cohort and thus only 27 patients could be included. With the exception of the twin set 139/140, twins discordant for disease often lacked informative longitudinal samples, in part due to feeding being suspended resulting in reduced excrement. Although the number and timing of samples collected from each set of twins was generally comparable, varying numbers of samples were collected between twin sets which may bias some analysis. The data generated using DGGE and pyrosequencing were in agreement, perhaps due to primer sets encompassing the V3 hypervariable region of the 16S rRNA genome in both analyses. Pyrosequencing allows larger fragments to be amplified allowing the use of bifidobacteria-optimised universal primers. Despite this, bifidobacteria was not found to be a prevalent genus (Palmer *et al.* 2007) which requires further study due to the potential use of this taxa as probiotic in therapeutic intervention (Embleton & Yates 2008).

In summary, this study represents a unique temporal analysis of the gut microbiota in preterm twins, cared for within the complex environment of neonatal intensive care. Although twins discordant for late onset infection showed differences in gut microbiota development, overall, related infants harboured bacterial communities more similar to each other than nonrelated infants. As well as shared genetic and immunomodulatory factors, this is likely a result of exposure to the same maternal microflora during birth, skin contact and exposure to EBM. We have also shown that other environmental factors, particularly antibiotic exposure, have additional significant effects on the gut microbiota in genetically related infants. These findings reflect other studies in which antibiotics have been shown to alter the gut microbiota in term (Palmer *et al.* 2007) and adult popultions (Simões & Maukonen 2013) and the exact role of individual antibiotics in altering the preterm gut microbiota warrants further investigation. We have further noted potential concurrence between community changes associated with other

inflammatory mediated diseases, such as ulcerative colitis and Crohn's disease, and those increasingly reported to occur in NEC.

7. Next generation sequencing of the gut microbiota in preterm neonates: A case-control study

Abstract

Aim: The gut microbiota is significantly associated with the development of NEC and sepsis. Previous studies are limited by cohort size, poor sampling, and methodological restrictions. We aimed to extensively explore the differential community development in patients with NEC and sepsis, matched to controls.

Methods: In total, 42 preterm infants were enrolled contributing a total of 747 stool samples. Patients were split into two groups consisting of 21 patients where 7 patients developed proven NEC and/or sepsis matched to 14 controls. All samples were analysed on the MiSeq and paired end reads underwent contig assembly and processing in Mothur.

Results: The gut microbiota was relatively unstable in the initial weeks of life. The core microbiome consisted of *Klebsiella Oxytoca*, *Escherichia coli*, *Staphylococcus*, *Enterococcus*, and *Veillonella.* The diversity of the gut microbiota generally increased over time and was not consistently reduced prior to disease diagnosis. However, increases in the abundance of certain OTUs were observed prior to NEC diagnosis, particularly with *Escherichia coli*. Organisms isolated in blood culture for the diagnosis of sepsis were typically abundant in the gut. Caesarean delivery resulted in increased colonisation by *Staphylococcus*, but after 3 weeks of life the effect of birth mode was lost. Gestational age had a significant ($P = 0.001$) influence of the bacterial community.

Conclusions: The preterm gut microbiota is a complex and dynamic community with a multitude of factors influencing its development. Gestational age had important influences on the community. While no consistent associations between reduced diversity or increased dominance prior to disease diagnosis were observed, *Escherichia*

coli was prevalent prior to diagnosis of NEC. A diverse community seems to be important to the health of a neonate supporting the notion of probiotics to stabilise the gut microbiota.

7.1 Introduction

The gut microbiota has received significant interest over the past decade, with the use of NGS technologies helping to reveal the true diversity of this complex ecosystem (Fukatsu 2012). Pyrosequencing provided the majority of early NGS data. More recently the Illumina SBS platform has been favoured, owing to its ability to accurately identify homopolymer runs and with the introduction of the bench top MiSeq the technology has become more accessible and increasingly affordable (Luo *et al.* 2012). Indeed, the MiSeq was recently reported to be the best bench top sequencing platform currently available (Loman *et al.* 2012). However, despite constantly improved chemistry, NGS techniques are still subject to PCR bias, including amplification efficiency and artifacts arising due to the formation of heteroduplex and chimeric sequences (Thompson *et al.* 2002; Wang & Qian 2009).

While the preterm gut microbiota is considerably less diverse than that of healthy term neonates, previous studies have lacked the coverage to accurately determine its true diversity (Arboleya *et al.* 2012). Deep sequencing has been implemented in a small cohort of 11 preterm infants and revealed an extremely diverse community consisting of bacteria, fungi, viruses, bacteriophage, and surprisingly roundworm (LaTuga *et al.* 2011). Data from the human genome project suggests a huge number of reads is required to fully elucidate complete diversity (Schatz *et al.* 2010). This is due to the presence of low abundance, often transient taxa in the community which require deep coverage in order to identify them. In contrast, the dominant taxa tend to show greater temporal stability and typically represent the core microbiome, which are all the OTUs that are present in all samples over a certain percentage cut-off (Jalanka-Tuovinen *et al.* 2011). In a preterm population we have previously shown a core community of 5 OTUs from the *Enterobacteriaceae*, *Enterococcaceae*, and *Staphylococcaceae* familes at 85%

cut-off (Chapter 6; Stewart *et al.* 2013). However, the importance of the non-core organisms are increasingly recognised as important as improved sequencing technology has facilitated the coverage of these taxa (Matsuda *et al.* 2009; van der Gast *et al.* 2011). Previous preterm gut microbiome studies have lacked the sequencing coverage to accurately explore this satellite community, which may have important implications in health and disease (Sobhani *et al.* 2011).

It has been reported that by one year of age an infant's gut microbiota is reflective of an adult community with increased *Bacteroides* and a reduction in *Proteobacteria* as well as greater stability, in the absence of medical intervention (Palmer *et al.* 2007). In comparison, the neonatal gut microbiota is less stable, with antibiotics known to significantly delay and alter its development (Mai *et al.* 2013). This may have important consequences for the preterm infant where delayed establishment of a diverse community can predispose the infant to NEC and sepsis. NEC and sepsis are both significant diseases which primarily affect the preterm population, with gut microbiota development significantly different compared to healthy controls (Mai et al. 2013; Stewart et al. 2013). Reports that the gut microbiota undergoes shifts and reduction in its diversity prior to disease diagnosis suggests that it is, in part, involved in the pathogenesis of NEC (Wang *et al.* 2009). Therefore, further information regarding the microbial involvement in the pathogenesis of these diseases may offer important information to improve the ability for both early detection and the subsequent clincal management.

NEC and sepsis are difficult diseases to diagnose with current staging criteria deemed outdated. Typically the patient will not present with the symptoms of NEC (abdominal distension) or sepsis (pyrexia) until the advanced stages of disease progression. Thus, subsequent management of the disease often needs to be more invasive leading to increased morbidity (Berrington *et al.* 2012). Medical management of NEC and sepsis

is based on antibiotic treatment for a minimum of 5 days. Where medical management is insufficient and in the more severe cases of NEC, the patient may undergo surgery to remove the necrotic tissue.

In an ideal study one would recruit a cohort of diseased patients matched to healthy controls. Regular longitudinal sampling to monitor the development of the gut microbiota as well as robust sampling preceding and following disease diagnosis would be essential if the aetiology of disease was to be explored. However, current data on the preterm gut microbiota is often limited by scarce and irregular sampling making resulting conclusions superficial (Mai *et al.* 2013; Mai *et al.* 2011; Wang *et al.* 2009). In clinical microbial ecology, comparison between diseased and healthy patients is challenging due to the number of confounding variables which exist between patients due to differing host genetics and environmental/clinical exposures. Preterm infants are subject to intensive care practise which involve being housed in sterile incubators with limited environmental microbial exposure (Claud *et al.* 2013).

To overcome the limitations of previous studies, we have investigated a large cohort of patients with NEC and/or sepsis, matched to healthy controls. Robust longitudinal sampling allowed the most informative samples to be selected retrospectively. All samples underwent NGS on the MiSeq platform to achieve a high depth of coverage.

7.2 Results

7.2.1 Patients and samples

Demographic information from each patient is summarised in Table 7.1 and further details based on each individual sample can be found in Appendix 11 (disc). A total of 42 preterm infants were enrolled in the study contributing a total of 747 stool samples. Patients were split into two groups; Extremely Preterm (group "EP": gestational age 23 - 26 weeks) and Very Preterm (group "VP": gestation age 27 – 30 weeks). Each group consisted of 21 patients where 7 patients developed proven NEC and/or sepsis matched to 14 patients who acted as controls. Of the diseased patients in the EP group, 4 patients developed NEC only with 1 fatality, 1 patient developed sepsis with *Staphylococcus hominis* and *Staphylococcus epidermidis* and was later diagnosed with NEC, 1 patient developed sepsis with CoNS and was deemed to have died of NEC which was not diagnosed until death, and 1 patient developed sepsis with *Escherichia coli*. Of the diseased patients in the VP group, 2 patients developed NEC and 5 patients developed sepsis with 2 cases of *Staphylococcus aureus* and single cases of CoNS, *Enterococcus faecalis*, and *Streptococcus agalactiae*. No fatalities occurred in the VP group. No patient in the cohort received probiotics or prebiotics.

Table 7.1 – Demographic summary of each group

Group Averages	Extremely Preterm (GA[a] 23 – 26 wks)		Very Preterm (GA 27 – 30 wks)	
	Diseased (n=7)	**Control (n=14)**	**Diseased (n=7)**	**Control (n=14)**
GA[a] (wks)	25	25	28	28
Birth Weight (g)	653	845	1165	1224
Birth Mode (CS[b] / Vaginal)	2 / 5	5 / 9	4 / 3	8 / 6
Gender (Male / Female)	1 / 6	8 / 6	6 / 1	12 / 2
Fatalities	2	0	0	0
NEC	5	0	2	0
NEC + Sepsis	1	0	0	0
Sepsis	1	0	5	0

[a]Gestational age;

[b]Caesarean

7.2.2 Overview of preterm gut microbiota development in whole cohort

Sequencing by synthesis (SBS) resulted in sample coverage of over 99% for all samples (Appendix 11 - disc). This facilitated subsequent analysis as the majority of low abundance OTUs were detected. Culturing and subsequent identification of isolates by MALDI-TOF and full length 16S sequencing identified an unclassified *Enterobacteriaceae* OTU as *Klebsiella oxytoca* and an important *Escherichia* OTU as *Escherichia coli*. Overall, a core microbiome present in over 85% of samples consisted of 5 OTUs corresponding to *K. oxytoca*, *E. coli*, *Staphylococcus*, *Enterococcus*, and *Veillonella*. The OTUs with presence in every sample corresponded to *Escherichia*, *Staphylococcus*, and *Enterococcus* (Appendix 11). Although not present in the core microbiome (85% cut-off), *Bifidobacterium* was found to be an abundant taxon (Table 7.2).

To investigate the role of delivery mode on the development of the gut microbiota patients who contributed week 1 stool (n = 26) were analysed. The first stool collected in each week, up to week 7, were included to visualise the average profile for both caesarean and vaginal delivery (Fig 7.1). The Bray-Curtis index, which takes into account abundance, was used to compare the relatedness of the communities each week. Profiles over the first 3 weeks of life show the least similarity, with an increased abundance of *Staphylococcus* in caesarean infants. Greatest variation occurred in week 2 with notable dominance by *Staphylococcus* in caesarean infants. From week 4 the *Staphylococcus* dominance in caesarean samples is lost and profiles show greater similarity. Profiles from week 7 showed the greatest similarity.

Table 7.2 – Abundance of bacterial OTUs from whole cohort

Average per sample (%)	Core cut-off (%)	Taxonomy
29.97	95	Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; unclassified
21.48	100	Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia
15.17	100	Firmicutes; Bacilli; Bacillales; Staphylococcaceae; Staphylococcus
14.22	100	Firmicutes; Bacilli; Lactobacillales; Enterococcaceae; Enterococcus
3.40	75	Actinobacteria; Actinobacteria; Bifidobacteriales; Bifidobacteriaceae; Bifidobacterium
2.08	85	Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae; Veillonella
1.49	<50	Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides
1.31	60	Firmicutes; Bacilli; Lactobacillales; Streptococcaceae; Streptococcus
1.10	60	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas
0.99	<50	Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Morganella
0.88	<50	Actinobacteria; Actinobacteria; Actinomycetales; Actinomycetaceae; Actinomyces
0.64	55	Firmicutes; Clostridia; Clostridiales; Clostridiaceae_1; Clostridium_sensu_stricto
0.50	<50	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Moraxellaceae; Acinetobacter
0.48	<50	Proteobacteria; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; unclassified
0.46	<50	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas
0.45	<50	Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Yersinia
0.41	55	Firmicutes; Clostridia; Clostridiales; Clostridiales_Incertae_Sedis_XI; Finegoldia
0.38	<50	Actinobacteria; Actinobacteria; Actinomycetales; unclassified; unclassified
0.35	70	Proteobacteria; Betaproteobacteria; Burkholderiales; Oxalobacteraceae; Herbaspirillum
0.33	<50	Firmicutes; Clostridia; Clostridiales; Clostridiales_Incertae_Sedis_XI; Peptoniphilus
0.29	50	Actinobacteria; Actinobacteria; Actinomycetales; Corynebacteriaceae; Corynebacterium
0.28	<50	Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Proteus
0.24	<50	Actinobacteria; Actinobacteria; Actinomycetales; Corynebacteriaceae; Corynebacterium
0.22	<50	Firmicutes; Bacilli; Lactobacillales; Streptococcaceae; Streptococcus
0.21	<50	Firmicutes; Bacilli; Lactobacillales; Lactobacillaceae; Lactobacillus
0.19	<50	Actinobacteria; Actinobacteria; Actinomycetales; Actinomycetaceae; Actinomyces
0.18	<50	Firmicutes; Bacilli; Lactobacillales; Streptococcaceae; Streptococcus
0.16	<50	Firmicutes; Bacilli; Lactobacillales; Lactobacillaceae; Lactobacillus
0.14	<50	Firmicutes; Clostridia; Clostridiales; Clostridiales_Incertae_Sedis_XI; Anaerococcus
0.13	<50	Proteobacteria;Gammaproteobacteria;Xanthomonadales;Xanthomonadaceae;Stenotrophomonas
0.12	<50	Proteobacteria; Alphaproteobacteria; Rhizobiales; Brucellaceae; Brucella
0.12	<50	Fusobacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae; Fusobacterium
0.11	<50	Actinobacteria; Actinobacteria; Actinomycetales; Dermabacteraceae; Dermabacter

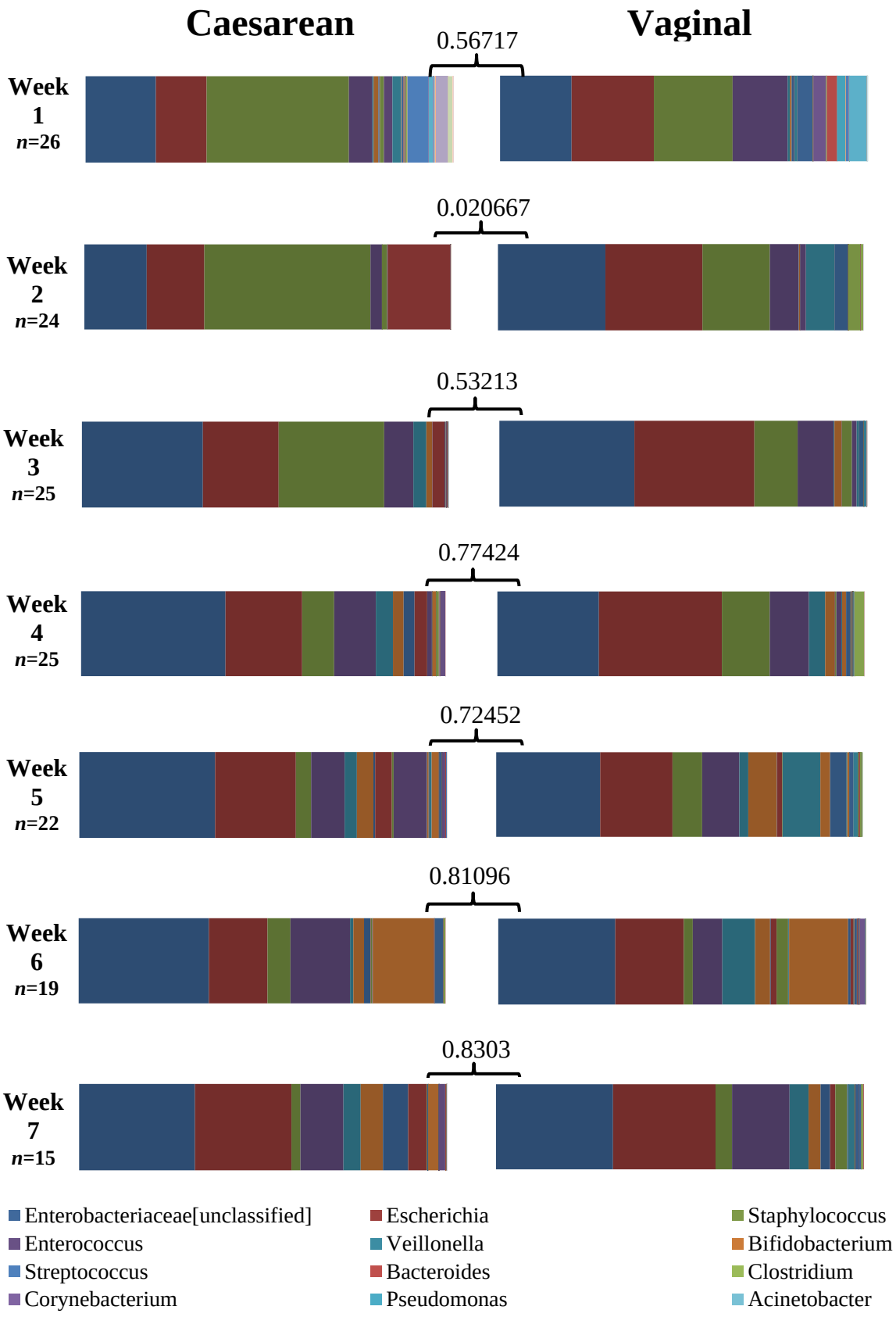


Figure 7.1 – Effect of birth mode on bacterial community development over the first 7 weeks of life in whole cohort. Legend shows 12 most abundant taxa. Similarity scores based on Bray-Curtis index where 1 represents identical communities.

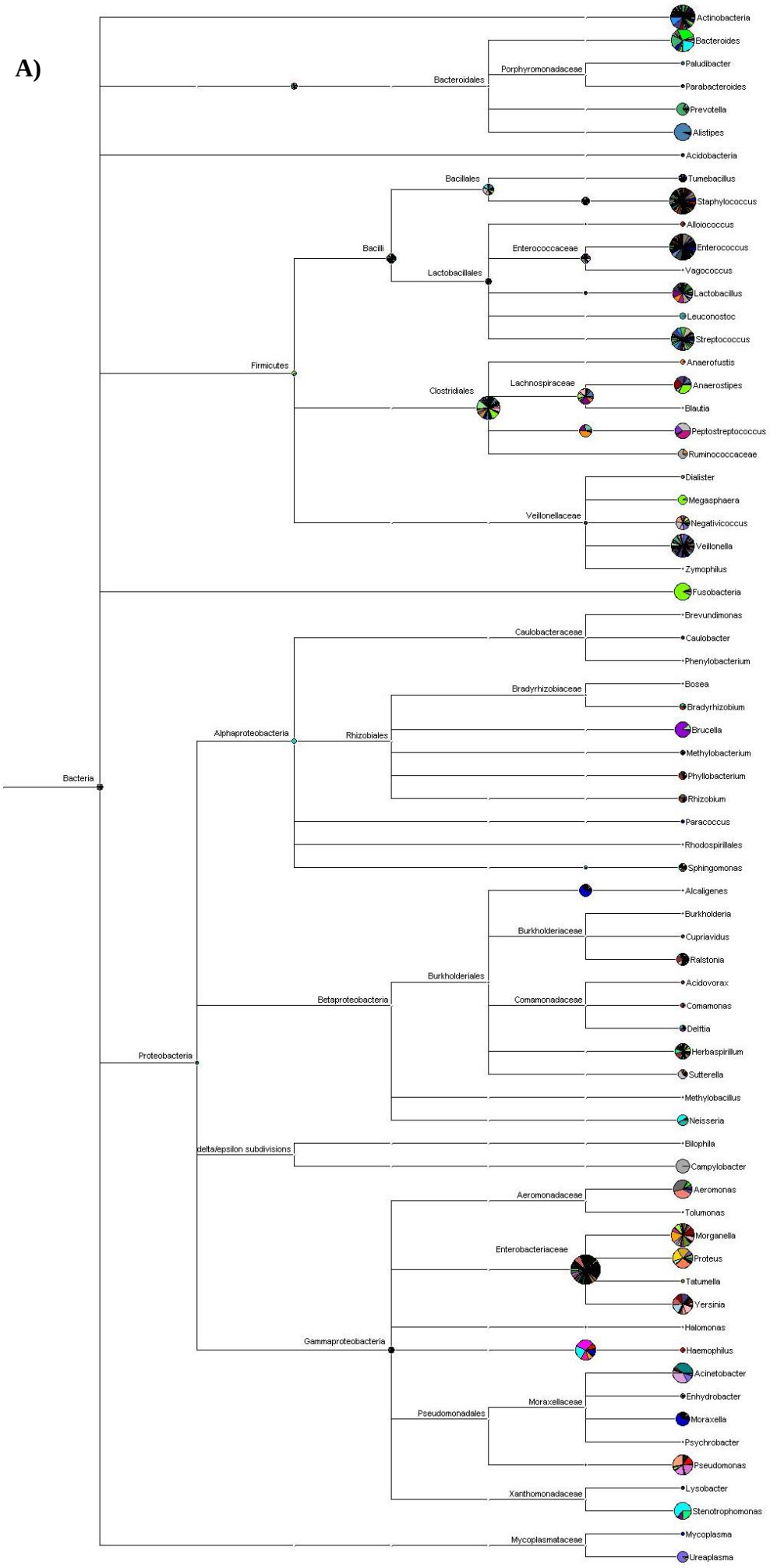
7.2.3 Comparison of the extremely preterm vs very preterm gut microbiota

The study design allowed for the comparison of EP and VP infants. Applying weighted UniFrac statistical analysis, which considers branch length of the phylogenetic tree, revealed gestational age had a significant effect (P = 0.001) on the gut microbiota development. This significant difference can be visualised by generation of phylogenetic trees (Fig 7.2). The diameter of the circle at each node is proportional to abundance of this OTU. The segments within each circle consist of all the samples where the OTU was detected, relative to the abundance within each sample. Clinically important OTUs that differed significantly include *Lactobacillus* (P = <0.001), and *Ureaplasma* (P = <0.001) which were greater in EP and *Pseudomonas* (P = <0.001) which was greater in EP (full list in Appendix 11).

The difference between the EP and VP was further explored by plotting the normalised abundance of both the dominant and satellite OTUs for healthy infants over the initial weeks of life (Fig 7.3). Only the first sample per patient in each week was included. This revealed the individual nature of each infants gut microbiota. The 5 OTUs from the core microbiome contributed a significant proportion to the communites throughout, although the dominant organism varied within and between patients. However, these taxa became less abundant in the later weeks of life. In contrast, the satellite taxa became more abundant. The abundance of *Bifidobacterium*, satellite taxa, generally increased in the later weeks, with particular presence around week 10 resulting in reduced dominance by the core microbiome. Patient 234 (VP group) had a very distinct gut microbiota that was almost entirely dominated by *K. oxytoca*.

Figure 7.2 – Phylogenetic tree of all samples based on gestational age. The diameter of the circle at each node is proportional to abundance of this OTU. The segments within each circle consist of all the samples where the OTU was detected, relative to the abundance within each sample. A) Extremely preterm. B) Very preterm.

A)


Actinobacteria
Bacteroides
Paludibacter
Porphyromonadaceae
Parabacteroides
Bacteroidales
Prevotella
Alistipes
Acidobacteria
Tumebacillus
Bacillales
Staphylococcus
Alloiococcus
Enterococcus
Enterococcaceae
Bacilli
Vagococcus
Lactobacillales
Lactobacillus
Leuconostoc
Streptococcus
Anaerofustis
Firmicutes
Lachnospiraceae
Anaerostipes
Clostridiales
Blautia
Peptostreptococcus
Ruminococcaceae
Dialister
Megasphaera
Veillonellaceae
Negativicoccus
Veillonella
Zymophilus
Fusobacteria
Brevundimonas
Caulobacteraceae
Caulobacter
Phenylobacterium
Bradyrhizobiaceae
Bosea
Bradyrhizobium
Alphaproteobacteria
Brucella
Rhizobiales
Methylobacterium
Bacteria
Phyllobacterium
Rhizobium
Paracoccus
Rhodospirillales
Sphingomonas
Alcaligenes
Burkholderia
Burkholderiaceae
Cupriavidus
Ralstonia
Burkholderiales
Acidovorax
Betaproteobacteria
Comamonadaceae
Comamonas
Delftia
Proteobacteria
Herbaspirillum
Sutterella
Methylobacillus
Neisseria
Bilophila
delta/epsilon subdivisions
Campylobacter
Aeromonas
Aeromonadaceae
Tolumonas
Morganella
Enterobacteriaceae
Proteus
Tatumella
Yersinia
Halomonas
Gammaproteobacteria
Haemophilus
Acinetobacter
Enhydrobacter
Moraxellaceae
Pseudomonadales
Moraxella
Psychrobacter
Pseudomonas
Lysobacter
Xanthomonadaceae
Stenotrophomonas
Mycoplasma
Mycoplasmataceae
Ureaplasma

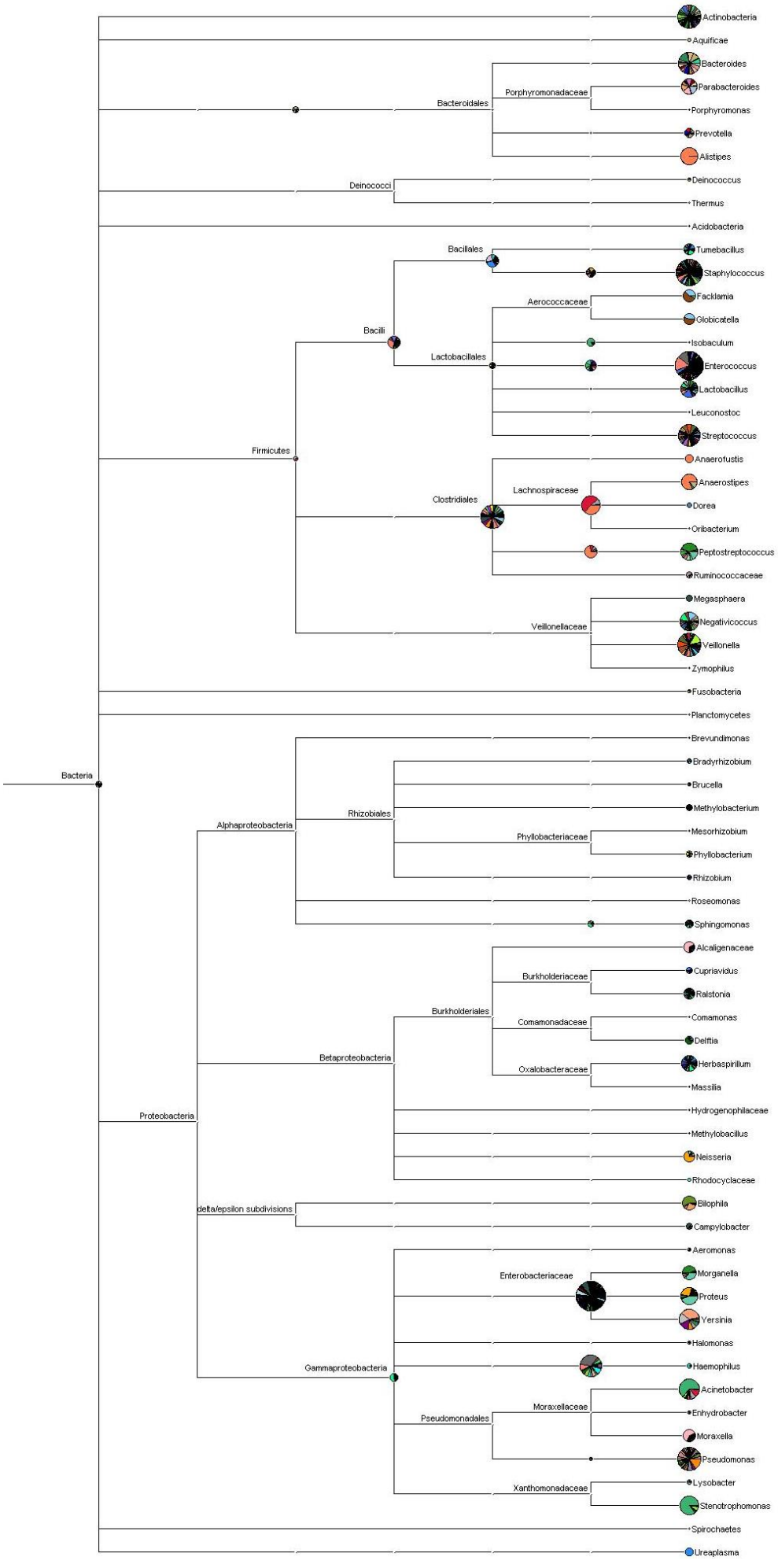
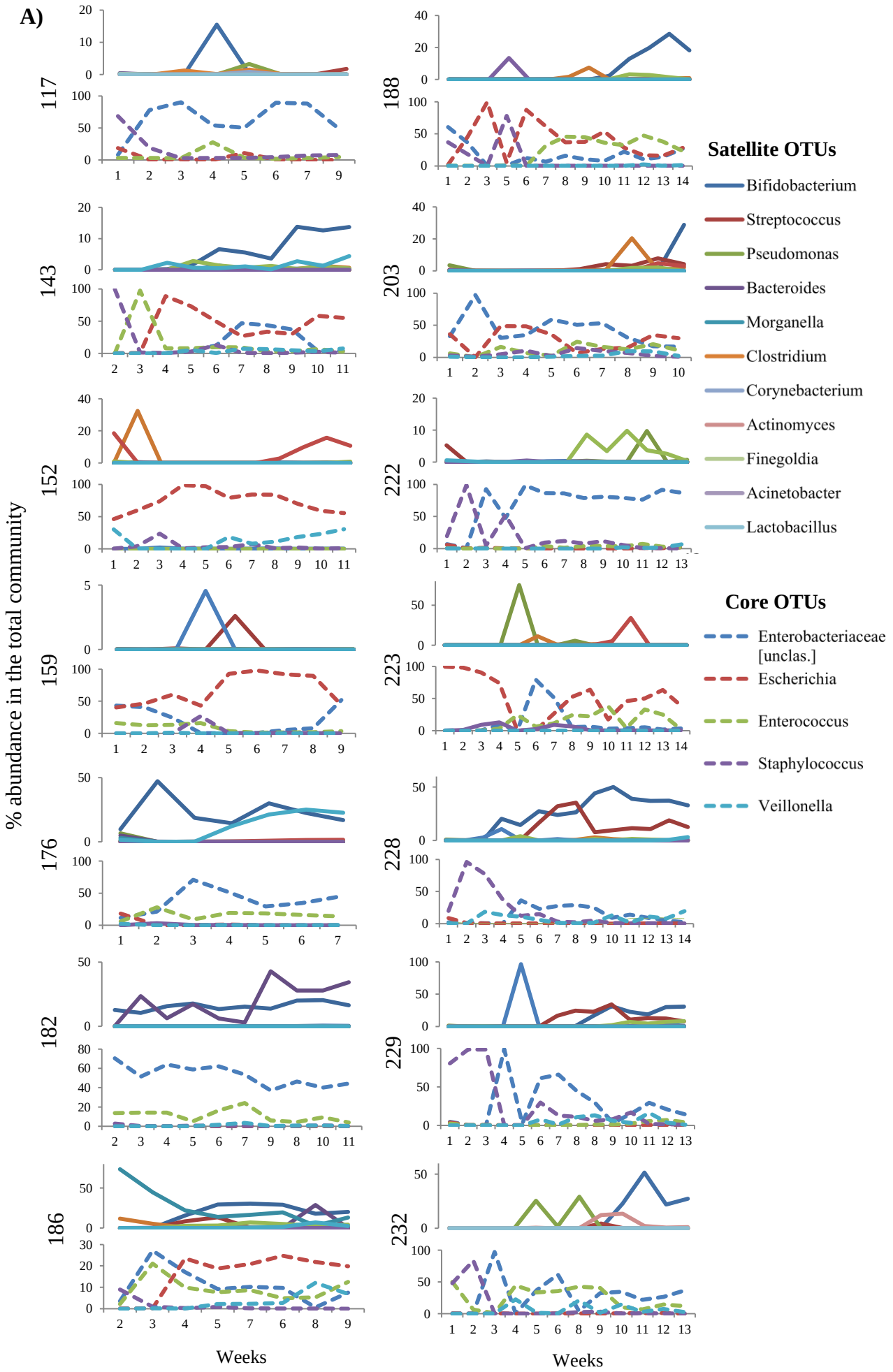
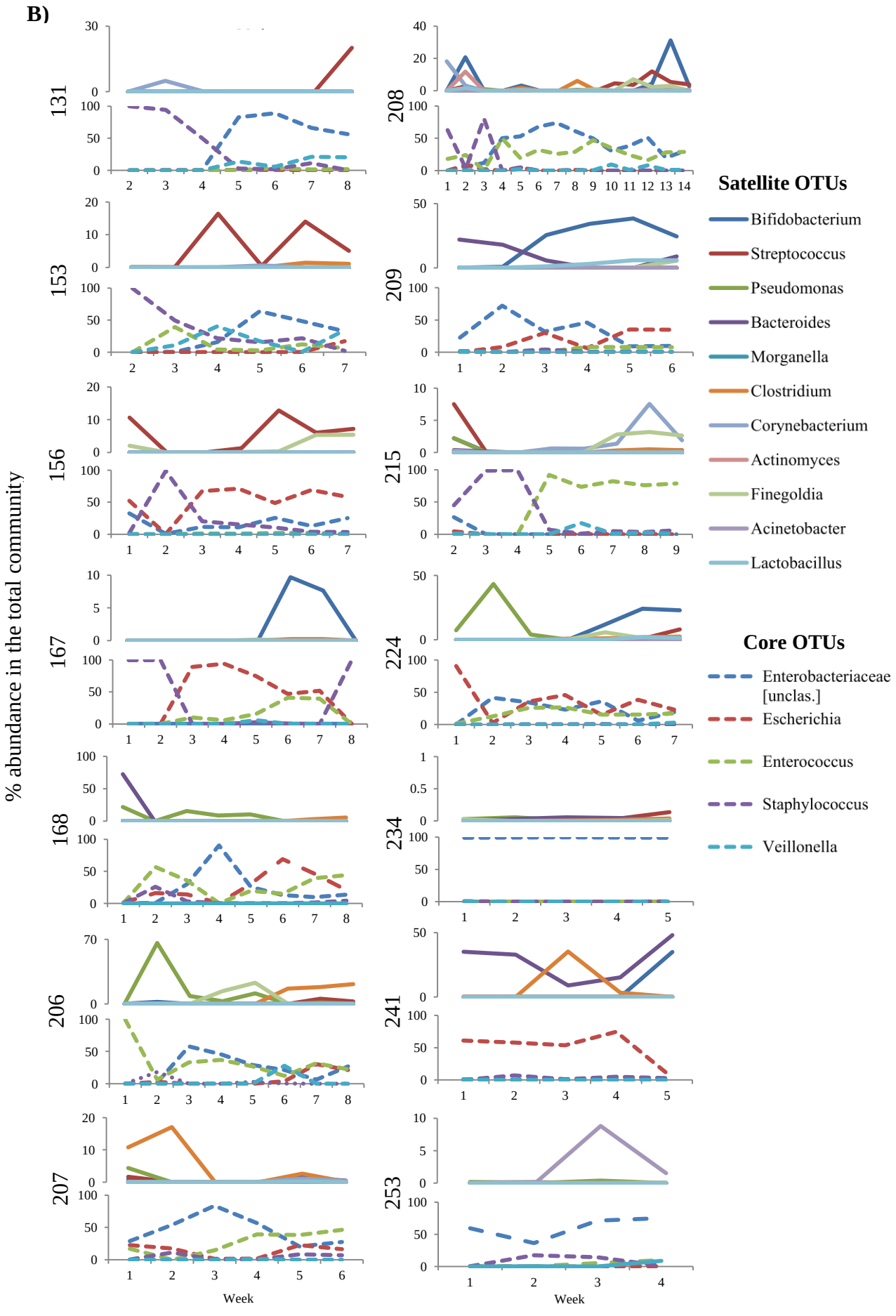

Actinobacteria
Aquificae
Bacteroides
Parabacteroides
Porphyromonadaceae
Bacteroidales
Porphyromonas
Prevotella
Alistipes
Deinococcus
Deinococci
Thermus
Acidobacteria
Tumebacillus
Bacillales
Staphylococcus
Facklamia
Aerococcaceae
Globicatella
Bacilli
Isobaculum
Lactobacillales
Enterococcus
Lactobacillus
Leuconostoc
Streptococcus
Firmicutes
Anaerofustis
Anaerostipes
Lachnospiraceae
Clostridiales
Dorea
Oribacterium
Peptostreptococcus
Ruminococcaceae
Megasphaera
Negativicoccus
Veillonellaceae
Veillonella
Zymophilus
Fusobacteria
Planctomycetes
Brevundimonas
Bradyrhizobium
Bacteria
Brucella
Methylobacterium
Rhizobiales
Alphaproteobacteria
Mesorhizobium
Phyllobacteriaceae
Phyllobacterium
Rhizobium
Roseomonas
Sphingomonas
Alcaligenaceae
Burkholderiaceae
Cupriavidus
Ralstonia
Burkholderiales
Comamonadaceae
Comamonas
Delftia
Betaproteobacteria
Oxalobacteraceae
Herbaspirillum
Massilia
Hydrogenophilaceae
Proteobacteria
Methylobacillus
Neisseria
Rhodocyclaceae
Bilophila
delta/epsilon subdivisions
Campylobacter
Aeromonas
Morganella
Enterobacteriaceae
Proteus
Yersinia
Halomonas
Haemophilus
Gammaproteobacteria
Acinetobacter
Moraxellaceae
Enhydrobacter
Pseudomonadales
Moraxella
Pseudomonas
Lysobacter
Xanthomonadaceae
Stenotrophomonas
Spirochaetes
Ureaplasma

Figure 7.3 – Development of the core and satellite gut microbiota in each healthy infant over the initial weeks of life. Legend shows satellite and core communities which are divided into the top and bottom graphs, respectively. Number indicates the individual patient number of the respective profiles (full patient data in appendix 11 - disc) A) Extremely preterm. B) Very preterm.


A)
% abundance in the total community
117
188
143
203
152
222
159
223
176
228
182
229
186
232
Weeks
Weeks
Satellite OTUs
Bifidobacterium
Streptococcus
Pseudomonas
Bacteroides
Morganella
Clostridium
Corynebacterium
Actinomyces
Finegoldia
Acinetobacter
Lactobacillus
Core OTUs
Enterobacteriaceae [unclas.]
Escherichia
Enterococcus
Staphylococcus
Veillonella


B)
% abundance in the total community
131
208
153
209
156
215
167
224
168
234
206
241
207
253
Week
Week
Satellite OTUs
Bifidobacterium
Streptococcus
Pseudomonas
Bacteroides
Morganella
Clostridium
Corynebacterium
Actinomyces
Finegoldia
Acinetobacter
Lactobacillus
Core OTUs
Enterobacteriaceae [unclas.]
Escherichia
Enterococcus
Staphylococcus
Veillonella

7.2.4 Development of NEC and sepsis

Due to the important significance between the groups based on gestational age, these groups were analysed independently for the association of the gut microbiota with NEC and sepsis pathogenesis. PLS-DA revealed a distinct bacterial community between the diseased samples, both pre and post onset, and the control samples (Fig 7.4). Within the diseased samples there was also separation of samples from NEC patients with sepsis patients. This applied to both gestational groups. Both the EP and VP groups also showed that a relatively large number of samples which predate NEC diagnosis fell outside the ellipse (indicating Hotellings T^2 range, at 95% confidence) and were thus significantly different. In the EP group, post sepsis samples grouped with post NEC samples, whereas in the VP group the post sepsis samples showed greatest similarity to the pre sepsis samples.

The Shannon diversity (H') was used to explore the diversity of diseased samples proceeding and following disease diagnosis, matched to two controls (Fig 7.5). Diversity variation was greatest in the initial weeks of life and increased from relatively low diversity initially to a more diverse community prior to discharge from the NICU. No consistent H' trend was observed prior to disease diagnosis within diseased infants compared to the control infants, although in general the H' decreased following disease diagnosis and subsequent antibiotic treatment.

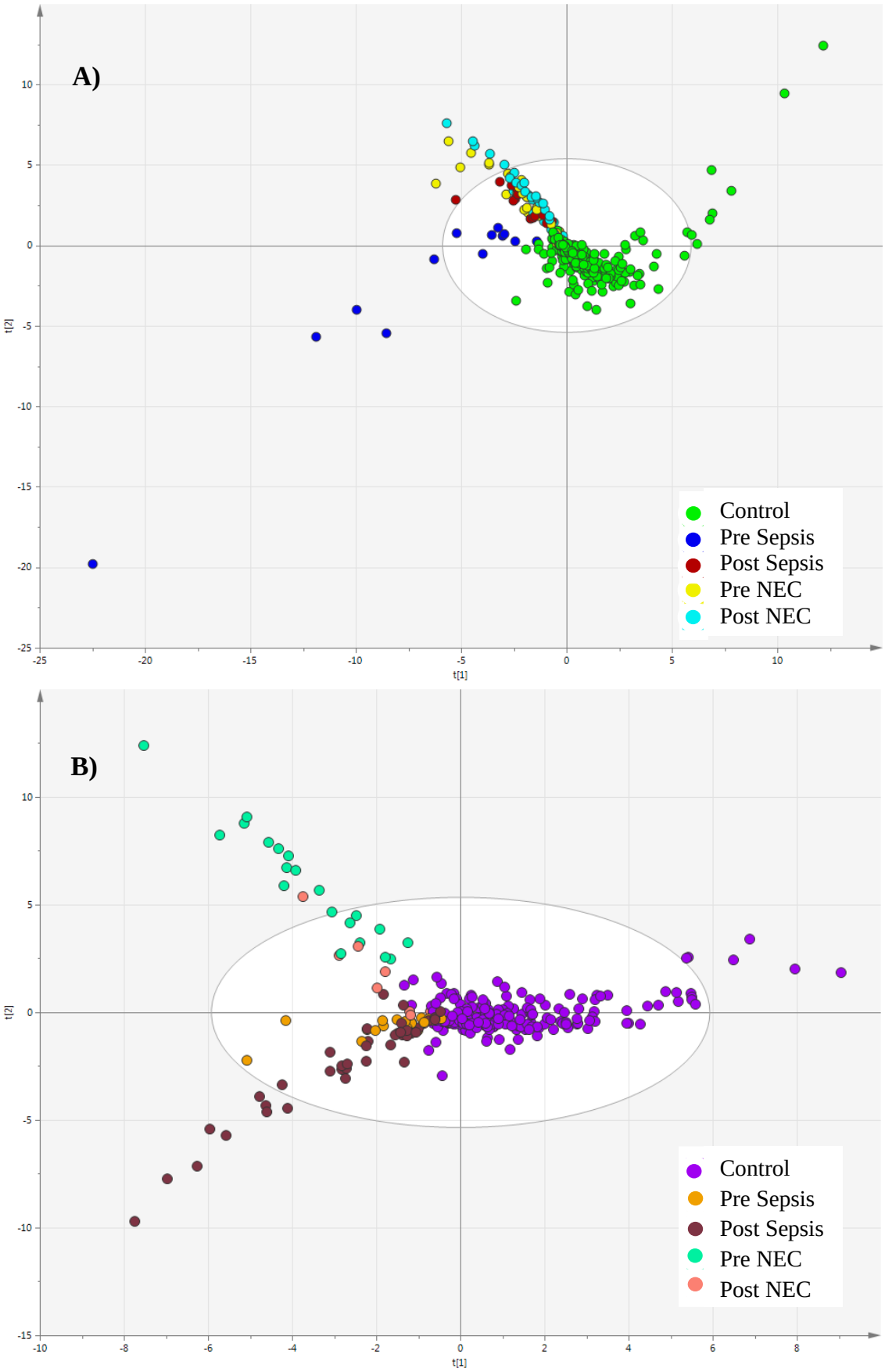
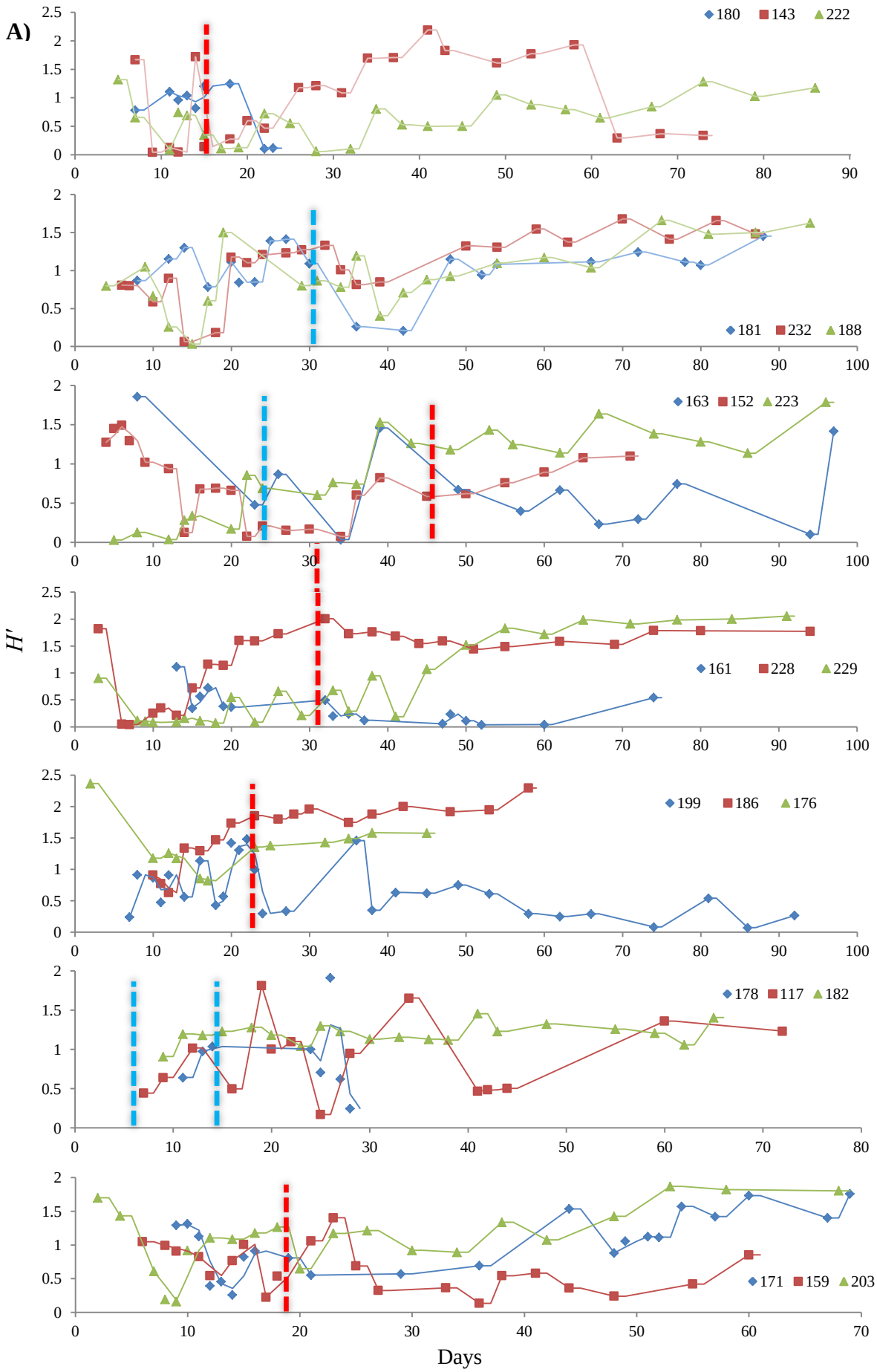
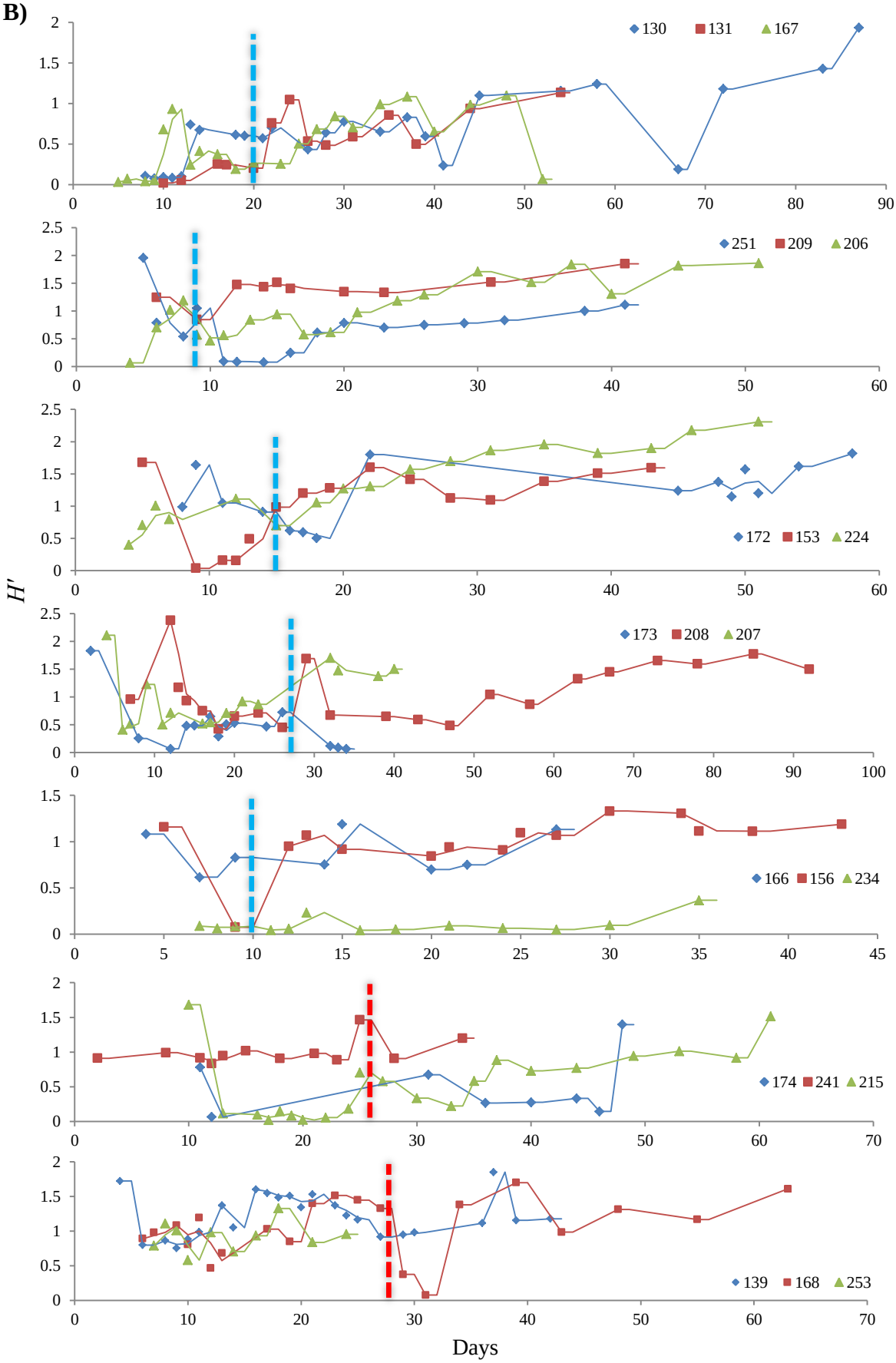


Figure 7.4 – PLS-DA plots comparing the bacterial profiles of all samples from each gestational group according to disease. A) Extremely preterm. B) Very preterm.

Figure 7.5 – Shannon diversity indices of each infant. Dotted blue lines represent sepsis diagnosis and dotted red lines represent NEC diagnosis. Disease infants are blue with each sample represented by triangles. Matched controls are either red with each sample represented by a square or green red with each sample represented by a triangle. A) Extremely preterm. B) Very preterm.


A)
H'
180 143 222
181 232 188
163 152 223
161 228 229
199 186 176
178 117 182
171 159 203
Days

B)


H'
130 131 167
251 209 206
172 153 224
173 208 207
166 156 234
174 241 215
139 168 253
Days

In a more focused exploration of the change in the total bacterial community in NEC and sepsis, area charts from the diseased patients were generated and antibiotic exposure mapped onto the graphs. Further details regarding the specific antibiotics and duration of administration are presented in Table 7.3. Antibiotic exposure was greatest in the EP group. *E. coli* was more prevalent in the EP group and the abundance of this organism increased prior to NEC in the majority of patients (Fig 7.6). Patient 171 from the EP group was an exception, where an increase in *K. Oxytoca* abundance prior to diagnosis of medical NEC was observed. Notably, this patient had the highest gestation from the EP group. Two patients (180 and 178), both in the EP group, died while on the NICU. In the final sample from both of these infants the community is dominated by *E. coli.*

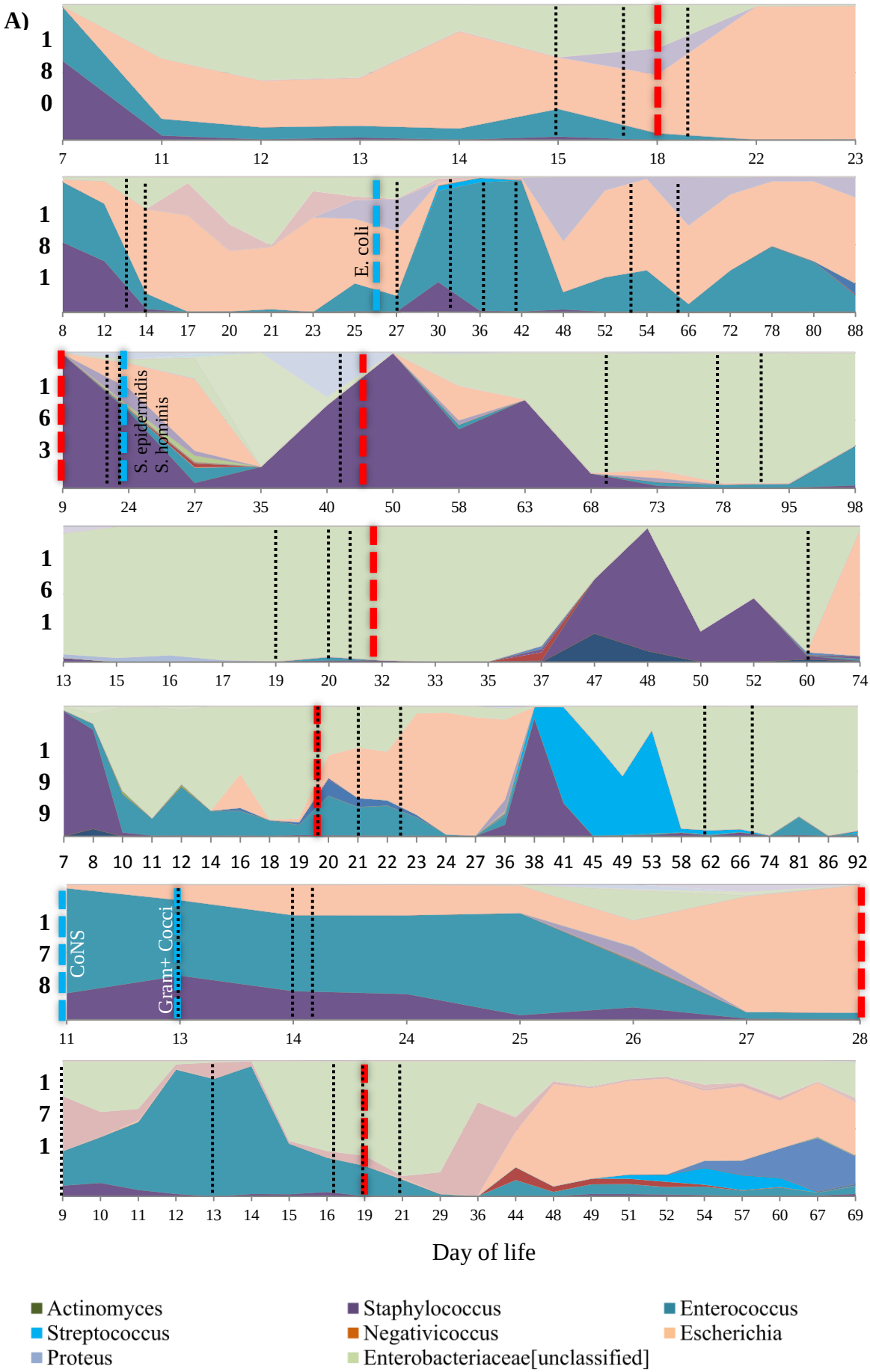
For NEC development in the VP group, *E. coli* abundance also increased in patient 139 prior to diagnosis and unique to this patient was a large abundance of *Actinomyces* spp. from day 16 of life. However, *E. coli* was not abundant in the other patient from the VP groups diagnosed with NEC (patient 174). Although no pre NEC sample was obtained 14 days prior to NEC diagnosis in this patient, the community remained stable from the preceding and post diagnosis samples.

In patients diagnosed with sepsis, the dominant organism in the gut prior to diagnosis often correlates with the species isolated in blood culture. In patient 130 there was a dominance of *Staphylococcus* prior to positive blood culture, in which *Staphylococcus aureus* was isolated. In other cases, the organism identified by blood culture appears in the community transiently. For example, *Streptococcus* only appears in patient 173 in the three days prior to positive blood culture with *Streptococcus agalactiae*. While shifts in the community were observed prior to and following disease diagnosis, no significant difference occurred in the ecological dominance of a single taxon from the bacterial community in NEC and sepsis patients, matched to controls (Fig 7.7).

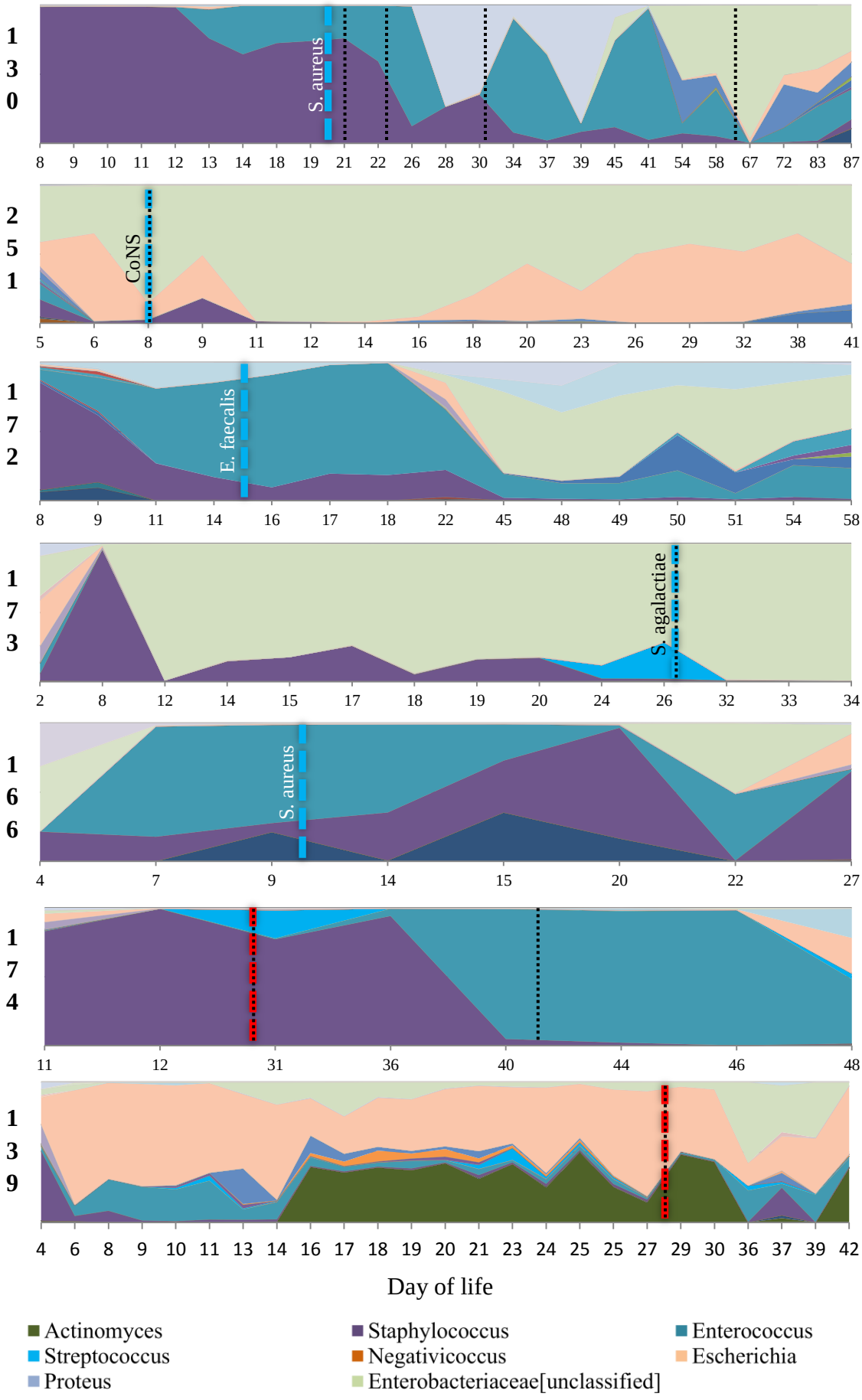
Table 7.3 – Antibiotic information patients with NEC and/or sepsis

	Patient Number	Abx start (days)	Abx 1 (days of usage)	Abx 2 (days of usage)	Abx 3 (days of usage)	Abx 4 (days of usage)	Abx 5 (days of usage)
Extremely preterm	180	0	P (2)	G (2)			
		15	A (3)	F (3)	G (3)		
		17	T (9)				
		19	M (18)				
		28	Me (10)	G (7)			
		43	M (2)	V (2)	C (2)		
	181	0	P (2)	G (2)			
		13	V (3)				
		14	F (9)				
		27	T (1)	C (1)	V (2)	G (4)	Me (17)
		33	G (3)				
		37	G (3)				
		41	G (3)				
		53	A (2)	F (2)	G (2)		
		61	A (2)	F (2)	G (2)		
	163	0	P (2)	G (2)			
		8	M (7)	V (7)	C (7)		
		20	A (2)	F (2)	G (2)		
		22	V (3)	C (5)			
		42	V (7)	M (7)	C (7)		
		69	V (2)	C (2)			
		77	V (2)	C (2)	M (2)		
		86	F (2)	G (2)			
	161	0	P (2)	G (2)			
		6	C (7)	V (7)	M (7)		
		19	C (6)	V (6)			
		20	M (5)				
		27	M (10)	A (10)	G (10)		
		61	A (5)				
	199	0	P (2)	G (2)			
		23	A (4)	F (4)	G (6)		
		25	M (10)				
		27	V (8)	C (8)			
		67	F (6)				
		74	C (2)	V (2)			
	178	0	P (2)	G (2)			
		6	C (3)	V (2)			
		13	V (2)	C (5)			
		14	M (9)				
		15	L (14)				
		29	C (1)	M (1)	V (1)		
	171	0	P (3)	G (3)			
		9	V (2)	C (2)			
		13	A (3)	G (3)	F (3)		
		17	A (4)	G (5)	F (4)		
		19	M (8)				
		21	V (6)	C (6)			
Very preterm	130	0	P (2)	G (2)			
		1	A (7)	M (7)			
		8	V (3)	C (3)			
		21	V (2)	C (2)			
		23	F (14)				
		31	G (6)				
		63	A (2)	F (2)	G (2)		
	251	0	P (2)	G (2)			
		8	V (7)	C (3)			
	172	0	P (2)	G (2)			
		15	F (5)	G (5)	A (5)		
	173	0	P (5)	G (5)			
		26	F (2)	A (8)	G (2)		
	166	Not available					
	174	0	P (5)	G (5)			
		5	V (2)	C (2)			
		25	A (2)	M (7)	C (7)		
		41	V (2)				
	139	0	P (2)	G (2)			
		28	F (4)	A (4)	G (4)	M (4)	

Figure 7.6 – Area charts of all diseased infants showing all phylotypes to genus level. Only the most abundant genera are in the legend for clarity. Dotted blue lines represent sepsis diagnosis and dotted red lines represent NEC diagnosis. Dotted black lines show antibiotic start day as per table 7.3. A) Extemely preterm. B) Very preterm.


A)
180
7 11 12 13 14 15 18 22 23
181
E. coli
8 12 14 17 20 21 23 25 27 30 36 42 48 52 54 66 72 78 80 88
163
S. epidermidis
S. hominis
9 24 27 35 40 50 58 63 68 73 78 95 98
161
13 15 16 17 19 20 32 33 35 37 47 48 50 52 60 74
199
7 8 10 11 12 14 16 18 19 20 21 22 23 24 27 36 38 41 45 49 53 58 62 66 74 81 86 92
178
CoNS
Gram+ Cocci
11 13 14 24 25 26 27 28
171
9 10 11 12 13 14 15 16 19 21 29 36 44 48 49 51 52 54 57 60 67 69
Day of life
Actinomyces
Staphylococcus
Enterococcus
Streptococcus
Negativicoccus
Escherichia
Proteus
Enterobacteriaceae[unclassified]

B)


130
S. aureus
8 9 10 11 12 13 14 18 19 21 22 26 28 30 34 37 39 45 41 54 58 67 72 83 87
251
CoNS
5 6 8 9 11 12 14 16 18 20 23 26 29 32 38 41
172
E. faecalis
8 9 11 14 16 17 18 22 45 48 49 50 51 54 58
173
S. agalactiae
2 8 12 14 15 17 18 19 20 24 26 32 33 34
166
S. aureus
4 7 9 14 15 20 22 27
174
11 12 31 36 40 44 46 48
139
4 6 8 9 10 11 13 14 16 17 18 19 20 21 23 24 25 25 27 29 30 36 37 39 42
Day of life
Actinomyces
Staphylococcus
Enterococcus
Streptococcus
Negativicoccus
Escherichia
Proteus
Enterobacteriaceae[unclassified]

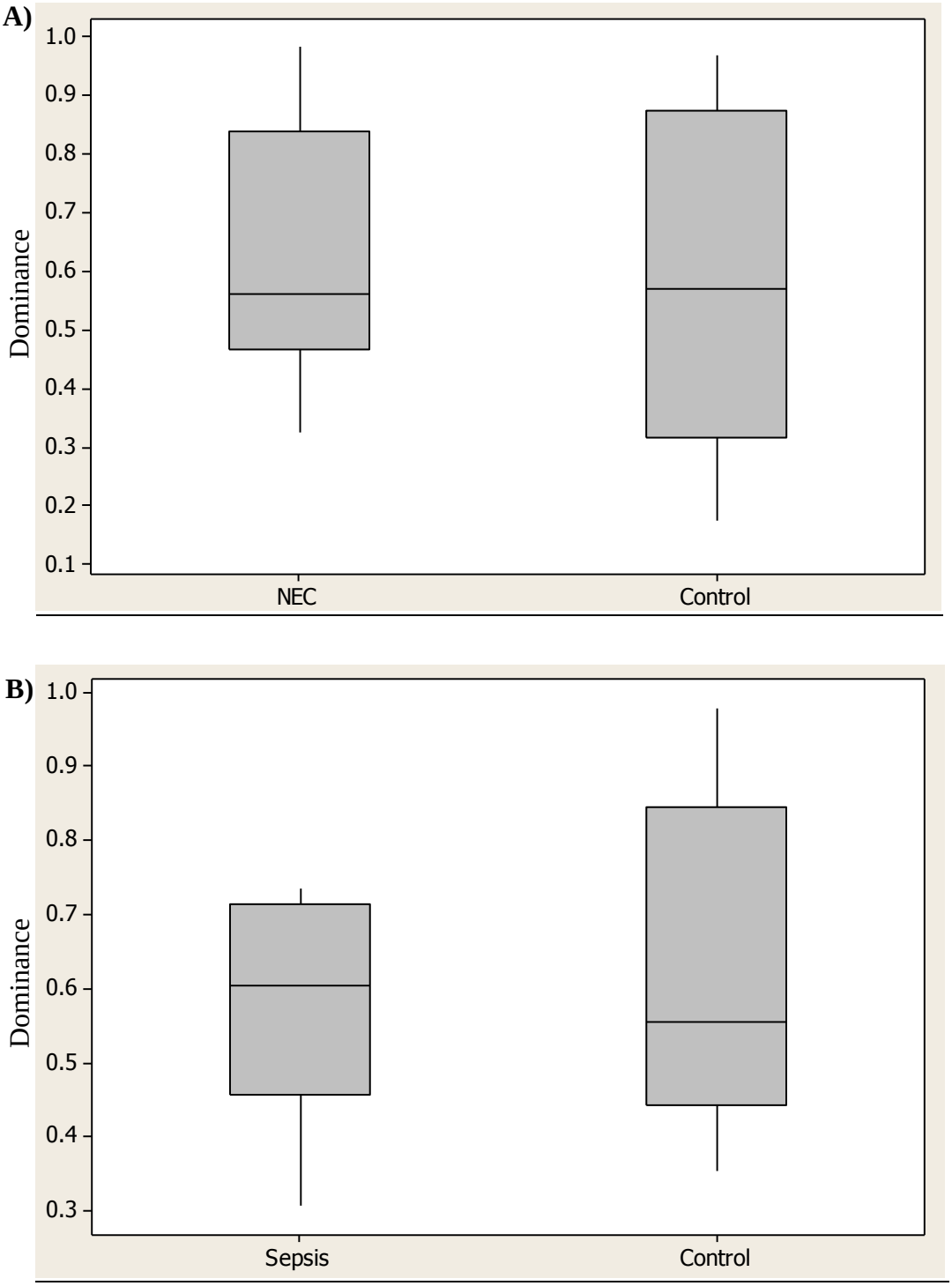


Figure 7.7 – Box plots of dominance of the bacterial community preceding disease diagnosis matched to controls. Box represents first and third quartile and the line with each box represents the median. Dominance scores of 1 represent a community consisting of a single taxon. A) Dominance in infants diagnosed with NEC. B) Dominance in infants diagnosed with sepsis.

7.3 Discussion

The gut microbiota has been previously shown to significantly differ between diseased patients and controls, suggesting a role in the etiology of NEC and sepsis. Elucidating the complex changes of this community prior to disease diagnosis will facilitate improved diagnostics, treatment, and prevention of significant morbidities. This study represents the largest such study to date, employing NGS to explore the gut microbiota in preterm infants while on the NICU. The sequence depth and subsequent coverage of the bacterial community in each sample involved in this study is far greater than previous studies (Wang *et al.* 2009; Mai *et al.* 2011), facilitating the detection of low abundance OTUs.

The influence of delivery mode on the gut microbiota of the neonate has received much attention. Significant differences have been reported based on delivery mode for the meconium and skin flora of neonates immediately following birth. Specifically, caesarean and vaginal delivery results in an initial gut microbiota of organisms that reflect the typical skin and vaginal microbiota, respectively (Biasucci *et al.* 2008; Dominguez-Bello *et al.* 2010). The results in this study support these findings with *Staphylococcus*, a common skin organism, being dominant from week 1 to 3 of life in caesarean delivered infants (Dominguez-Bello *et al.* 2010). This study advances on previous data in the neonate and shows that the difference observed between delivery modes is lost after week 3. By week 4 of life, infants developed an increasingly comparable gut microbiota, independent of delivery mode. This is in contrast to suggestions by others that the gut microbiota in infants born by caesarean delivery can be disrupted for up to 6 months (Grönlund *et al.* 1999). Using culture based approaches, differences were also reported in the gut microbiota 7 years after birth (Salminen *et al.* 2004). The difference in results between studies might be attributable to antibiotics,

specifically penicillin and gentamicin, which are administered to all neonates in this preterm cohort for 48 hours following birth. Thus, the antibiotics may prevent the long-term establishment of the pioneering organisms into the gut, allowing organisms which are introduced in subsequent weeks to colonise (Pérez-Cobas *et al.* 2012). This is further supported by the results of delivery mode on the gut microbiota where large numbers of pioneering organisms are no longer detectable after the initial week of life.

In association with previous studies, *K. oxytoca*, *E. coli*, *Staphylococcus*, *Enterococcus*, and *Bifidobacterium* were the most abundant genera throughout the entire cohort (Mshvildadze *et al.* 2010; Claud *et al.* 2013; Stewart *et al.* 2013). The most abundant OTU corresponded to an unclassified *Enterobacteriaceae* but no sequence matched this OTU by more than 97%, despite this family being particularly well studied. Mai *et al.* (2011) also detected an *Enterobacteriaceae* which did not match any sequence in Genbank by more than 97%. We were able to successfully isolate this OTU in culture and identify it as *K. oxytoca* suggesting further important annotation of databases is required. The abundant organisms largely reflected the core microbiome, which is representative of prevalent organisms but included *Veillonella* and excluded *Bifidobacterium*. *Veillonella* is a common member of the oral and gut flora and its prevalence might reflect the implications of this organism in lactate fermentation and biofilm formation (Periasamy & Kolenbrander 2010; Madan *et al.* 2012).

While members of the core microbiome dominated in the initial weeks, in the later weeks of life this dominance is less pronounced, with satellite organisms increasing in overall abundance. The bacterial profiles for each infant were specific to the individual. The abundance of *Bifidobacterium* colonisation in preterm neonates has varied considerably between studies but was found to be an abundant member of the community in this study. The levels of *Bifidobacterium* were observed to increase around week 10 of life and could represent a substantial proportion of the community in

some healthy patients. This may have important implication in probiotic therapy as increased levels of *Bifidobacterium* prevent dominance by a potentially pathogenic organism. Interestingly, *Lactobacillus* is also proposed as another potential probiotic candidate but was found in low abundance in all patients, in agreement with a previous study (Cox *et al.* 2010). Other OTUs from the satellite population which have important clinical implications were found in relatively high abundance in some patients, but appeared rather transiently. These satellite organisms, such as *Streptococcus* and *Pseudomonas*, are not well studied in the preterm gut but their presence warrants further consideration. *Pseudomonas aeruginosa* is an opportunistic pathogen which can cause healthcare-associated infections in the NICU, with increased risk in low birth weight preterm infants (Jefferies *et al.* 2012).

Gestational age, relative to prematurity, is known to be the most significant risk factor associated with NEC (Berrington *et al.* 2013). However, studies exploring the role of the gut microbiota in NEC pathogenesis neglect to consider that both the community itself and its influence in causing NEC may differ between EP and VP infants. This study represents the first to consider this and importantly we demonstrate a significant difference ($P = 0.001$) in the bacterial communities between these groups. While the dominant bacterial genera were comparable between the two groups, significant differences were reported for some lower abundant OTUs. Of particular clinical importance are *Lactobacillus* ($P = <0.001$), and *Ureaplasma* ($P = <0.001$) which were greater in VP and *Pseudomonas* ($P = <0.001$) which was greater in EP. The administration of antibiotics was increased in the EP group compared to the VP group which may account, in part, for the significance of gestational age in the development of the bacterial community in the preterm gut.

The significant difference between the gut microbiota of the EP and VP group warrants consideration when exploring the pathogenesis of disease. Although control samples

cluster distinctly from diseased samples in both groups, important differences were found between the groups. In the EP group, the post disease samples from both NEC and sepsis cluster together whereas the post disease samples are still comparable to the pre disease samples in the VP group. This may reflect a greater effect of antibiotic treatment in changing the gut microbiota in the more premature group (Lafeber *et al.* 2008). Increased antibiotic administration has been associated with the development of NEC and increased use of antibiotics in the EP group may contribute to the increased cases of NEC in this group. Antibiotics can significantly alter the gut microbiota and may cause dysbiosis (Hawrelak & Myers 2004). It is currently unclear whether a dysbiosis event occurs prior to disease diagnosis and, indeed, if this is causative or an effect of disease progression. For example, Wang *et al.* (2009) reported a reduction in diversity prior to NEC diagnosis but it is unclear if this was causative to NEC or a result of elevated inflammation. This is consistent with findings from other IBD conditions, including ulcerative colitis and Crohn's disease, where a decrease in diversity is observed prior to diagnosis (Gophna *et al.* 2006). However, in this study a reduction in diversity was not a factor in the predisposition for NEC or sepsis *per se*. Despite abundant OTUs in the community prior to diagnosis, the dominance of a single OTU did not consistently reduce when comparing diseased patients to matched controls. This is in accordance with a recent publication by Mai *et al.* (2011) which also employed NGS technology.

Temporal analysis with regular sampling is key to elucidating the changes of the dynamic preterm gut microbiota, which might be attributable to disease onset. The *Escherichia* genus has important clinical considerations in this study and the abundance of this genus was notably increased in the EP group, compared to the VP group. While diseased patients were found to harbour unique profiles, 6 of 8 NEC patients from the whole cohort had increase of *E. coli* prior to diagnosis. Within the EP group, two

patients (180 and 178) died while on the NICU and the final sample from both patients was dominated by *E. coli*. It should be noted that the *E. coli* was also detected in control infants. However, as outlined in chapter 6, this organism has received particular attention in recent studies exploring the role of the gut microbiota in inflammatory mediated conditions and the correlation between studies warrants further investigation (Gophna *et al.* 2006; Sasaki *et al.* 2007; Lepage *et al.* 2011).

In accordance with previous studies, the pathogenesis of sepsis was also associated with the gut microbiota (Madan *et al.* 2012; Mai *et al.* 2013; Stewart *et al.* 2013). However, unlike Madan *et al.* (2012) and Mai *et al.* (2013), sepsis was not associated with a reduction in diversity to the normal gut microbiota development in this study. In previous studies the sepsis cohort was more premature than the control group. We have shown prematurity to significantly influence the bacterial community development regardless of disease status, thus it is feasible that findings from studies where gestational age is not matched are confounded. In this study, the organisms detected by blood culture were always present in the gut microbiota and, in the majority of cases, were one of the abundant members of the community. The sepsis only case in the EP group (patient 181) was diagnosed by blood culture as *E. coli*, with this organism being the most abundant in the gut of this infant. *Staphylococcus* was causative to sepsis in the majority of cases and this organism was also abundant in the gut microbiota of these infants. This is in accordance with Madan *et al.* (2012), where *Staphylococcus* was dominant in the community prior to positive blood culture in which *Staphylococcus aureus* was isolated.

Probiotics have been proposed as a potentially useful supplement to preterm neonates. Evidence from this study suggests that a diverse gut microbiota may prevent the predisposition of NEC and sepsis to preterm infants. While current research into probiotic supplementation yields mixed results in terms of efficiency, it should be noted

that probiotics are not considered dangerous, nor have they been shown to have a detrimental effect on the host (Li *et al.* 2013; Nair & Soraisham 2013; Pärtty *et al.* 2013). Evidence suggests probiotics can be effective at colonising the gut as supplementing *Lactobacillus casei* increased the abundance of this organism in infant stool and stabilised the gut microbiota (Cox *et al.* 2010). They may further improve intestinal permeability and modulate the development and persistence of an appropriate mucosal immune response (Embleton & Yates 2008). Studies exploring the potential benefit of probiotic administration should consider the mechanistic effect on the gut microbiota, particularly the role in reducing dominance by potentially pathogenic organisms which is feasible based on existing evidence (Cox *et al.* 2010).

In summary, the preterm gut microbiota is a complex and dynamic community with a multitude of factors influencing its development. Mode of delivery initially results in different colonisation patterns and increased *Staphylococcus* in caesarean delivery. However, the community is more comparable regardless of delivery mode following week 3 of life. An important consideration in this study was the effect of gestational age on the bacterial community. Increased prematurity is regarded as the most significant risk factor in the care of neonates. A significant difference between the profiles of extremely preterm (>27 weeks gestation) and very preterm (27-30 weeks gestation) was reported. The differential development of the gut microbiota as a result of gestational age should be considered in future studies exploring the gut microbiota. While no consistent associations between reduced diversity and increased dominance prior to disease diagnosis were observed, *E. coli* was particularly abundant prior to diagnosis of NEC. A diverse community seems to be important to the health of a neonate supporting the notion of probiotics to stabilise the gut microbiota. Further evidence on the observed changes in the gut microbiota compared to the inflammatory state of the gut is

warranted to ascertain if the inflammation is driving these shifts and subsequent disease onset, or vice versa.

8. Concluding remarks

8.1 Summary

The development of the preterm neonatal gut microbiota is complex and highly individual. Assessing its association with clinical factors is challenging. This thesis has employed a range of genomic techniques to explore the total and viable bacterial and fungal communities. The fungal community was found to be patient specific, metabolically inactive and showed a relatively low diversity. There was also no association of the fungal community in the pathogenesis of NEC and sepsis. The bacterial community was also patient specific but more diverse than the fungal community. The viable bacterial community reflected the profiles of the total community. Thus subsequent investigations focused on the total bacterial community owning to ease of working with DNA compared to RNA.

The gut microbiota was dominated by two phyla; Proteobacteria and Firmicutes. Within these phyla the dominant genera were *Escherichia*, *Staphylococcus*, and *Enterococcus*. This dominance became less pronounced from the sixth week of life, with rarer taxa increasing in overall abundance. *Bifidobacterium* was abundant when exploring the V4 region using the MiSeq NGS platform but this was not the case in previous studies based on the PCR-DGGE of the V3 region and traditional culture. *Lactobacillus* was found in low abundance regardless of methodology or cohort.

The influence of delivery mode on the gut microbiota and the potential long term consequences are important. The results in this thesis show *Staphylococcus*, a common skin organism, being dominant from week 1 to 3 of life in caesarean delivered infants.

By week 4 of life, however, infants developed comparable gut microbiota to those delivered vaginally. This supports our early DGGE findings that birth mode did not significantly alter the bacterial community suggesting that birth mode does not influence the long-term establishment of the preterm gut microbiota.

Gestational age was demonstrated to alter the development of the bacterial community with significant differences between the abundances of clinically important OTUs. These differences might be attributable to antibiotics which are administered for 48 hours following birth preventing the long-term establishment of the pioneering organisms into the gut. The number of days of antibiotic administration is also increased in more preterm infants.

The power of the initial studies was limited by the size of the cohort and regularity of longitudinal sampling, but important differences between NEC and/or sepsis patients were found, compared to controls. Infants diagnosed with disease showed altered community development preceding and following disease diagnosis. In some cases this correlated with reduced diversity and increased dominance by a single OTU, but this did not apply to all cases. The *Escherichia* genus was associated with the pathogenesis of NEC with this genus increasing in dominance prior to NEC diagnosis, although, like the reduction in diversity, this observation was not seen universally in all patients. The pathogenesis of sepsis was also associated with the gut microbiota. Organisms detected by blood culture were present in the gut and, in the majority of cases, were one of the abundant members of the community.

Shifts in community structure and dominance by particular bacterial organisms might be causative to preterm disease pathogenesis. Overall, a stable and diverse community seems to be important to the health of a neonate.

8.2 Future work

The focus of this thesis has been on the microbial community present in the gut of preterm infants. This allows important insights into the ecology of microbes in this complex niche. To better understand the functional implications resulting from the shifts in the community or to elucidate if these shifts are driven by precursors such as inflammation it is important to adopt a systems biology approach. This involves implementation of a range of 'omic' techniques into experimental design, such as transcriptomics, proteomics, and metabolomics. Applying these techniques in parallel will help address the complexity involved in disease pathogenesis, especially relating to the functional aspects.

Proteomics provides information on which proteins are up- or down- regulated in response to demographic variables. This work is difficult on stool samples that are naturally rich in proteases (enzymes which cause the breakdown of proteins). However, proteomic analyses of serum is possible and serum represents a very useful tool, especially since bloods are taken from all infants on the NICU and, whereas stool samples can reduce around disease diagnosis, serum sampling increases (Embleton *et al.* 2013). Proteomic studies on necrotic gut tissues removed from patients who undergo surgery for NEC may provide useful insights into the events at the site affected by the disease. Because healthy tissue is also present either side of the necrotic area, a useful comparison between healthy and disease tissue is possible. A two-dimensional differential gel electrophoresis (2D DiGE) approach is often used. This involves the isoelectric focusing of samples on a strip which is then loaded into an acrylamide gel and electrophoresis carried out. Labelled proteins then appear as spots on the gels and the intensity of the spot relates to the abundance so the higher the intensity the more is

being produced. Spots of interest which are up or down regulated can then be removed from the gel and identified by mass spectroscopy (MS). Since samples can contain vast amounts of albumin it is necessary to run 2D gels. The large area relating to the albumin can be identified and removed from the analysis; otherwise the albumin peak in the MS spectrum would dominate preventing the detection of potentially important proteins.

Unlike genomics and proteomics which provide information on the genotype, metabolomics and the identification of low molecular weight compounds can be linked to phenotype. Stool samples that remain from the metagenomic studies could undergo extraction for metabolomics. These extractions can also be done on urine, serum, and tissue and involve homogenising the sample in solutions such as PBS and liquid chromatography mass spectrometry (LCMS) grade methanol before centrifugation and passing the supernatant through a 0.22 µm filter (Marchesi *et al.* 2007; Sellitto *et al.* 2012; Lee *et al.* 2013). These samples can then undergo LCMS analysis by passing the samples through a column. Fractioned metabolites are then read by MS and the resulting profiles can be searched against small molecule databases such as KEGG to identify the features. This will provide information pertaining to the host metabolite expression and when compared with demographic data can identify if metabolites of interest are up or down regulated. For example, specific metabolites may be present in higher amounts in patients with NEC compared to controls. This might have important implications in the prediction and prevention of the disease.

The work in this study was based on sampling from a single NICU and sampling only occurred while the patients were on the NICU. However, follow up studies would offer the opportunity to determine the impact of prematurity on long-term development of gut microbiota and pathologies associated with premature birth. To examine this, samples post discharge could be requested and sent back to the lab by post. Sampling from other NICUs would also allow comparisons of different clinical management to be assessed.

This would further increase the power of studies by increasing the number of diseased patients sampled and would eradicate the possibility that resulting conclusions are NICU specific.

Probiotics is currently an area of active debate with some studies demonstrating potential importance in reducing disease incidence and other studies reporting no effects. It will be important to determine the optimum 'cocktail' of species to use in the probiotic as well as the route of administration, dose, age at which to start and stop treatment, and the gestational cut-off of treatment. It is also currently unknown whether the probiotic strains are just transient colonisers during treatment or if the strains colonise long term. Information on the usefulness of prebiotics either instead of probiotics or in combination (synbiotics) is also limited. Future work should address these points and explore the immediate and long term effect of supplementation on the preterm gut microbiome.

Adherent-invasive *Escherichia coli* (AIEC) are commonly isolated from ileal biopsies in Crohn's disease (Small *et al.* 2013). AIEC are pathogenic and can colonise the intestinal mucosa by adhering to and invading intestinal epithelial cells to replicate intracellularly, as well as survive and replicate extensively within macrophages which induces the secretion of large amounts of TNF-α (Rolhion & Darfeuille-Michaud 2007). Potentially important to clarifying cause or effect, a recent report showed that the generation of nitrate by the host during the inflammatory response confers a growth advantage to commensal *E. coli* in anaerobic respiration (Winter *et al.* 2013). This ability to utilise the by-products of reactive oxygen and nitrogen species, produced during inflammation, may account for increase in abundance of this species in diseased patients owing to a growth advantage over other fermenting bacteria. Thus, the host inflammatory response can selectively enhance the growth of *E. coli* which should be considered when inferring the relationship between abundance of *E. coli* and NEC.

The dominant OTUs prior to disease diagnosis and the organisms isolated in positive blood culture are typically common residents of a normal healthy gut microbiota (Park *et al.* 2005). While no consistent differences occur between the diversity of diseased and control patients, it is plausible that dominant organisms in the community contribute to disease pathogenesis. One mechanism by which these otherwise normal members of the gut microbiota switch on pathogenesis is quorum sensing. Quorum sensing molecules (QSMs) are secreted by bacteria and when they reach a defined concentration they can activate bacterial proliferation and switch on a number of virulence genes (Chandran *et al.* 2003). Indeed, it has been proposed that the beneficial effect of antibiotics in the treatment of intestinal inflammation might be attributable, in part, to their effect on the quorum sensing related bacterial behaviour (Struss *et al.* 2012). When exploring the role of the gut microbiota it is crucial to know if the shifts in the bacterial community observed prior to diagnosis are causative to disease pathogenesis, or simply a subsidiarity effect of other factors such as inflammation. Interestingly, QSMs have been proposed as potential biomarkers to measure intestinal inflammatory activity (Kumari *et al.* 2008), which may help ascertain whether the bacterial community is driving disease pathogenesis, or vice versa.

9. References

Acinas, S. & Marcelino, L. (2004). Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. *J Bacteriol*, 186(9), pp.2629–2635.

Acker, J. Van & Smet, F. De (2001). Outbreak of necrotizing enterocolitis associated with Enterobacter sakazakii in powdered milk formula. *J Clin Microbiol*, 39(1), pp.293–97.

Van Acker, J., de Smet, F., Muyldermans, G., Bougatef, A., Naessens, A. & Lauwers, S. (2001). Outbreak of necrotizing enterocolitis associated with Enterobacter sakazakii in powdered milk formula. *J Clin Microbiol*, 39(1), pp.293–7.

Ader, F., Jawhara, S., Nseir, S., Kipnis, E., Faure, K., Vuotto, F., Chemani, C., Sendid, B., Poulain, D. & Guery, B. (2011). Short term Candida albicans colonization reduces Pseudomonas aeruginosa-related lung injury and bacterial burden in a murine model. *Crit Care care*, 15(3), p.R150.

Adessi, C., Matton, G., Ayala, G., Turcatti, G., Mermod, J.J., Mayer, P. & Kawashima, E. (2000). Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. *Nucleic Acids Res*, 28(20), p.E87.

Adlerberth, I. & Wold, a E. (2009). Establishment of the gut microbiota in Western infants. *Acta Paediatr*, 98(2), pp.229–38.

Alexander, V., Northrup, V. & Bizzarro, M. (2011). Antibiotic exposure in the newborn intensive care unit and the risk of necrotizing enterocolitis. *J Pediatr*, 159(3), pp.392–397.

Amend, A.S., Seifert, K. a & Bruns, T.D. (2010). Quantifying microbial communities with 454 pyrosequencing: does read abundance count? *Mol Ecol*, 19(24), pp.5555–65.

Amer, M., Hedlund, E., Rochester, J. & Caplan, M. (2004). Platelet-activating factor concentration in the stool of human newborns: effects of enteral feeding and neonatal necrotizing enterocolitis. *Biol Neonate*, 85(3), pp.159–166.

Ansorge, W. (2009). Next-generation DNA sequencing techniques. *N Biotechnol*, 25(4), pp.195–203.

Arboleya, S., Binetti, A., Salazar, N., Fernández, N., Solís, G., Hernández-Barranco, A., Margolles, A., Reyes-Gavilán, C. de los & Gueimonde, M. (2012). Establishment and development of intestinal microbiota in preterm neonates. *FEMS microbiol Ecol* , 79(3), pp.763–72.

Are, A., Aronsson, L., Wang, S., Greicius, G., Lee, Y.K., Gustafsson, J.-A., Pettersson, S. & Arulampalam, V. (2008). Enterococcus faecalis from newborn babies regulate endogenous PPARgamma activity and IL-10 levels in colonic epithelial cells. *Proc Natl Acad Sci U S A*, 105(6), pp.1943–8.

Atkinson, A.J., Colburn, W.A., DeGruttola, V.G., DeMets, D.L., Downing, G.J., Hoth, D.F., Oates, J.A., Peck, C.C., Schooley, R.T., Spilker, B.A., Woodcock, J. & Zeger, S.L. (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther*, 69(3), pp.89–95.

Bäckhed, F. (2011). Programming of Host Metabolism by the Gut Microbiota. *Ann Nutr Metab*, 58(s2), pp.44–52.

Bagci, S. & Eis-Hübinger, A. (2008). Detection of astrovirus in premature infants with necrotizing enterocolitis. *Pediatr Infect Dis J*, 27(4), pp.347–350.

Baker, G.C., Smith, J.J. & Cowan, D. a. (2003). Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods*, 55(3), pp.541–555.

Baxter, J. & Cummings, S.P. (2008). The degradation of the herbicide bromoxynil and its impact on bacterial diversity in a top soil. *J Appl Microbiol*, 104(6), pp.1605–16.

Benjamin, D., DeLong, E., Steinbach, W., Cotton, C.M., Walsh, T. & Clark, R.H. (2003). Empirical therapy for neonatal candidemia in very low birth weight infants. *Pediatrics*, 112(3), pp.543–47.

Benson, A.K., Kelly, S. a, Legge, R., Ma, F., Low, S.J., Kim, J., Zhang, M., Oh, P.L., Nehrenberg, D., Hua, K., Kachman, S.D., Moriyama, E.N., Walter, J., Peterson, D. a & Pomp, D. (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc Natl Acad Sci USA*, 107(44), pp.18933–8.

Bentley, D. & Balasubramanian, S. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456(7218), pp.53–59.

Berman, L. & Moss, R.L. (2011). Necrotizing enterocolitis: an update. *Semin Fetal Neonatal Med*, 16(3), pp.145–50.

Berrington, J., Stewart, C., Embleton, N. & Cummings, S. (2013). Gut microbiota in preterm infants: assessment and relevance to health and disease. *Arch Dis Child Fetal Neonatal Ed*, 98(4), pp.F286–90.

Berrington, J.E., Hearn, R.I., Bythell, M., Wright, C. & Embleton, N.D. (2012). Deaths in preterm infants: changing pathology over 2 decades. *J Pediatr*, 160(1), pp.49–53.

Bevins, C.L. & Salzman, N.H. (2011). The potter's wheel: the host's role in sculpting its microbiota. *Cell Mol Life Sci*, 68(22), pp.3675–85.

Bhandari, V., Wang, C., Rinder, C. & Rinder, H. (2008). Hematologic profile of sepsis in neonates: neutrophil CD64 as a diagnostic marker. *Pediatrics*, 121(1), pp.129–34.

Bhatia, J. (2010). Strategies to prevent necrotising enterocolitis. *Chin Med J (Engl)*, 123(20), pp.2759–2765.

Biasucci, G., Benenati, B., Morelli, L., Bessi, E. & Boehm, G. (2008). Cesarean Delivery May Affect the Early Biodiversity of Intestinal Bacteria. *J. Nutr.*, 138(9), p.1796S–1800.

Björkström, M. V, Hall, L., Söderlund, S., Håkansson, E.G., Håkansson, S. & Domellöf, M. (2009). Intestinal flora in very low-birth weight infants. *Acta Paediatr*, 98(11), pp.1762–7.

Blakey, J., Lubitz, L. & Barnes, G. (1982). Development of gut colonisation in pre-term neonates. *J Med Microbiol*, 15, pp.519–29.

Booijink, C.C.G.M., El-Aidy, S., Rajilić-Stojanović, M., Heilig, H.G.H.J., Troost, F.J., Smidt, H., Kleerebezem, M., De Vos, W.M. & Zoetendal, E.G. (2010). High temporal and inter-individual variation detected in the human ileal microbiota. *Environ Microbiol*, 12(12), pp.3213–27.

Booijink, C.C.G.M., Zoetendal, E.G., Kleerebezem, M. & de Vos, W.M. (2007). Microbial communities in the human small intestine: coupling diversity to metagenomics. *Future Microbiol.*, 2(3), pp.285–95.

Braak, C.J.F. Ter (1986). Canonical Correspondence Analysis: A New Eigenvector Technique for Multivariate Direct Gradient Analysis. *Ecology*, 67(5), pp.1167–79.

Braniste, V. & Pettersson, S. (2012). Enterococcus faecalis: a biological marker predicting the emergence of necrotizing enterocolitis. *Acta Paediatr*, 101(11), pp.1112–3.

Bray, J.R. & Curtis, J.T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol Monogr*, 27(4), p.325.

Breitbart, M., Haynes, M., Kelley, S., Angly, F., Edwards, R. a, Felts, B., Mahaffy, J.M., Mueller, J., Nulton, J., Rayhawk, S., Rodriguez-Brito, B., Salamon, P. & Rohwer, F. (2008). Viral diversity and dynamics in an infant gut. *Res Microbiol*, 159(5), pp.367–73.

Breitbart, M., Hewson, I., Felts, B., Mahaffy, J.M., Nulton, J., Salamon, P. & Rohwer, F. (2003). Metagenomic Analyses of an Uncultured Viral Community from Human Feces. *J Bacteriol*, 185(20), pp.6220–6223.

Brinkhoff, T. & Hannen, E. (2001). Use of silicone grease to avoid "smiling effect" in denaturing gradient gel electrophoresis. *J Rapid Meth Auto Microbio*, 9(4), pp.259–261.

Bucher, B.T., McDuffie, L. a, Shaikh, N., Tarr, P.I., Warner, B.B., Hamvas, A., White, F. V, Erwin, C.R. & Warner, B.W. (2011). Bacterial DNA content in the intestinal wall from infants with necrotizing enterocolitis. *J Pediatr Surg*, 46(6), pp.1029–33.

Bustin, S. a, Benes, V., Garson, J. a, Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J. & Wittwer, C.T. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem*, 55(4), pp.611–22.

Butel, M., Suau, A. & Campeotto, F. (2007). Conditions of bifidobacterial colonization in preterm infants: a prospective analysis. *J Pediatr Gastroenterol Nutr*, 44(5), pp.577–82.

Byun, R., Nadkarni, M.A., Chhour, K.L., Martin, F.E., Jacques, N.A. & Hunter, N. (2004). Quantitative analysis of diverse Lactobacillus species present in advanced dental caries. *J Clin Microbiol*, 42(7), pp.3128–3136.

Cabrera-Rubio, R. & Collado, M. (2012). The human milk microbiome changes over lactation and is shaped by maternal weight and mode of delivery. *Am J Clin Nutr*, 96(3), pp.544–551.

Cahan, H. & Deville, J.G. (2011). Outcomes of neonatal candidiasis: the impact of delayed initiation of antifungal therapy. *Int J Pediatr*, 2011, p.813871.

Caicedo, R. a, Schanler, R.J., Li, N. & Neu, J. (2005). The developing intestinal ecosystem: implications for the neonate. *Pediatr Res*, 58(4), pp.625–8.

Caplan, M. & Frost, B. (2011). Myth: necrotizing enterocolitis: probiotics will end the disease, and surgical intervention improves the outcome. *Semin Fetal Neonatal Med*, 16(5), pp.264–8.

Caplan, M. & Hedlund, E. (1997). The platelet-activating factor receptor antagonist WEB 2170 prevents neonatal necrotizing enterocolitis in rats. *J Pediatr Gastroenterol Nutr*, 24(3), pp.296–301.

Caplan, M.S., Sun, X.M., Hseuh, W. & Hageman, J.R. (1990). Role of platelet activating factor and tumor necrosis factor-alpha in neonatal necrotizing enterocolitis. *J Pediatr*, 116(6), pp.960–4.

Caporaso, J., Bittinger, K., Bushman, F., DeSantis, T., Andersen, G. & Knight, R. (2010). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*, 26(2), pp.266–7.

Caporaso, J., Kuczynski, J. & Stombaugh, J. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature*, 7(5), pp.335–336.

Carlisle, E.M., Poroyko, V., Caplan, M.S., Alverdy, J. a. & Liu, D. (2011). Gram Negative Bacteria Are Associated with the Early Stages of Necrotizing Enterocolitis. *PLoS One*, 6(3), p.e18084.

Carroccio, A., Iacono, G., Cottone, M., Di Prima, L., Cartabellotta, F., Cavataio, F., Scalici, C., Montalto, G., Di Fede, G., Rini, G., Notarbartolo, A. & Averna, M.R. (2003). Diagnostic accuracy of fecal calprotectin assay in distinguishing organic causes of chronic diarrhea from irritable bowel syndrome: a prospective study in adults and children. *Clin Chem*, 49(6 Pt 1), pp.861–7.

Çetinkaya, M., Özkan, H., Köksal, N., Akacı, O. & Özgür, T. (2011). Comparison of the efficacy of serum amyloid A, C-reactive protein, and procalcitonin in the diagnosis and follow-up of necrotizing enterocolitis in premature infants. *J Pediatr Surg*, 46(8), pp.1482–1489.

Chakravorty, S., Helb, D., Burday, M., Connell, N. & Alland, D. (2007). A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods*, 69(2), pp.330–9.

Chandran, P., Satthaporn, S., Robins, A. & Eremin, O. (2003). Inflammatory bowel disease: dysfunction of GALT and gut bacterial flora (II). *Surgeon*, 1(3), pp.125–36.

Chapman, R.L. (2007). Prevention and treatment of Candida infections in neonates. *Semin Perinatol*, 31(1), pp.39–46.

Chassard, C. & Bernalier-Donadille, A. (2006). H2 and acetate transfers during xylan fermentation between a butyrate-producing xylanolytic species and hydrogenotrophic microorganisms from the human gut. *FEMS Microbiol Lett*, 254(1), pp.116–22.

Chauhan, M., Henderson, G. & McGuire, W. (2008). Enteral feeding for very low birth weight infants: reducing the risk of necrotising enterocolitis. *Arch Dis Child Fetal Neonatal Ed*, 93(2), pp.F162–6.

Cilieborg, M.S., Boye, M. & Sangild, P.T. (2012). Bacterial colonization and gut development in preterm neonates. *Early Hum Dev*, 88 Suppl 1, pp.S41–9.

Claesson, M.J., Jeffery, I.B., Conde, S., Power, S.E., O'Connor, E.M., Cusack, S., Harris, H.M.B., Coakley, M., Lakshminarayanan, B., O'Sullivan, O., Fitzgerald, G.F., Deane, J., O'Connor, M., Harnedy, N., O'Connor, K., O'Mahony, D., van Sinderen, D., Wallace, M., Brennan, L., Stanton, C., Marchesi, J.R., Fitzgerald, A.P., Shanahan, F., Hill, C., Ross, R.P. & O'Toole, P.W. (2012). Gut microbiota composition correlates with diet and health in the elderly. *Nature*, 488(7410), pp.178–84.

Clarridge, J. (2004). Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Infect Dis*, 17(4), pp.840–862.

Claud, E.C., Keegan, K.P., Brulc, J.M., Lu, L., Bartels, D., Glass, E., Chang, E.B., Meyer, F. & Antonopoulos, D. (2013). Bacterial community structure and functional contributions to emergence of health or necrotizing enterocolitis in preterm infants. *Microbiome*, 1(1), p.20.

Coates, E.W., Karlowicz, M.G., Croitoru, D.P. & Buescher, E.S. (2005). Distinctive distribution of pathogens associated with peritonitis in neonates with focal intestinal perforation compared with necrotizing enterocolitis. *Pediatrics*, 116(2), pp.e241–6.

Coenye, T. & Vandamme, P. (2003). Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. *FEMS Microbiol Lett*, 228(1), pp.45–49.

Cogen, A., Nizet, V. & Gallo, R. (2008). Skin microbiota: a source of disease or defence? *Br J Dermatol*, 158(3), pp.442–455.

Collado, M.C., Delgado, S., Maldonado, A. & Rodríguez, J.M. (2009). Assessment of the bacterial diversity of breast milk of healthy women by quantitative real-time PCR. *Lett Appl Microbiol*, 48(5), pp.523–8.

Collier-Hyams, L.S. & Neish, a S. (2005). Innate immune relationship between commensal flora and the mammalian intestinal epithelium. *Cell Mol Life Sci*, 62(12), pp.1339–48.

Conte, M.P., Schippa, S., Zamboni, I., Penta, M., Chiarini, F., Seganti, L., Osborn, J., Falconieri, P., Borrelli, O. & Cucchiara, S. (2006). Gut-associated bacterial microbiota in paediatric patients with inflammatory bowel disease. *Gut*, 55(12), pp.1760–7.

Cordero, L. & Ayers, L. (2003). Duration of empiric antibiotics for suspected early-onset sepsis in extremely low birth weight infants. *Infect Control Hosp Epidemiol*, 24(9), pp.662–6.

Cotten, C.M., Taylor, S., Stoll, B., Goldberg, R.N., Hansen, N.I., Sánchez, P.J., Ambalavanan, N. & Benjamin, D.K. (2009). Prolonged duration of initial empirical antibiotic treatment is associated with increased rates of necrotizing enterocolitis and death for extremely low birth weight infants. *Pediatrics*, 123(1), pp.58–66.

Cox, M.J., Huang, Y.J., Fujimura, K.E., Liu, J.T., McKean, M., Boushey, H. a, Segal, M.R., Brodie, E.L., Cabana, M.D. & Lynch, S. V (2010). Lactobacillus casei abundance is associated with profound shifts in the infant gut microbiome. *PLoS One*, 5(1), p.e8745.

D'Incà, R., Dal Pont, E., Di Leo, V., Ferronato, A., Fries, W., Vettorato, M.G., Martines, D. & Sturniolo, G.C. (2007). Calprotectin and lactoferrin in the assessment of intestinal inflammation and organic disease. *Int J Colorectal Dis*, 22(4), pp.429–37.

Darfeuille-Michaud, A., Boudeau, J., Bulois, P., Neut, C., Glasser, A.-L., Barnich, N., Bringer, M.-A., Swidsinski, A., Beaugerie, L. & Colombel, J.-F. (2004). High prevalence of adherent-invasive Escherichia coli associated with ileal mucosa in Crohn's disease. *Gastroenterology*, 127(2), pp.412–421.

Deshpande, G., Rao, S., Patole, S. & Bulsara, M. (2010). Updated meta-analysis of probiotics for preventing necrotizing enterocolitis in preterm neonates. *Pediatrics*, 125(5), pp.921–30.

Deshpande, G., Rao, S., Patole, S. & Edward, K. (2007). Probiotics for prevention of necrotising enterocolitis in preterm neonates with very low birthweight: a systematic review of randomised controlled trials. *Lancet*, 369(9573), pp.1614–20.

Dethlefsen, L. & Relman, D. (2011). Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc Natl Acad Sci U S A*, 108(1), pp.4554–61.

Dhas, D.B.B., Bhat, B.V. & Gane, D.B. (2012). Role of Calprotectin in Infection and Inflammation. *Curr Pediatr Res*, 16(2), pp.83–94.

Dominguez-Bello, M.G., Costello, E.K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N. & Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci U S A*, 107(26), pp.11971–5.

Donachie, S.P., Foster, J.S. & Brown, M. V (2007). Culture clash: challenging the dogma of microbial diversity. *ISME J*, 1(2), pp.97–99.

Dowd, S., Sun, Y., Secor, P. & Rhoads, D. (2008). Survey of bacterial diversity in chronic wounds using pyrosequencing, DGGE, and full ribosome shotgun sequencing. *BMC Microbiol*, 8, p.43.

Dubourg, G., Lagier, J.C., Armougom, F., Robert, C., Hamad, I., Brouqui, P. & Raoult, D. (2013). The gut microbiota of a patient with resistant tuberculosis is more comprehensively studied by culturomics than by metagenomics. *Eur J Clin Microbiol Infect Dis*, 32(5), pp.637–45.

Durbán, A., Abellán, J.J., Jiménez-Hernández, N., Ponce, M., Ponce, J., Sala, T., D'Auria, G., Latorre, A., Moya, A., Ana, D., Abellan, J. & Moy, A. (2011). Assessing gut microbial diversity from feces and rectal mucosa. *Microb Ecol*, 61(1), pp.123–133.

Echarri, P.P., Graciá, C.M., Berruezo, G.R., Vives, I., Ballesta, M., Solís, G., Morillas, I.V., Reyes-Gavilán, C.G.D.L., Margolles, A. & Gueimonde, M. (2011). Assessment of intestinal microbiota of full-term breast-fed infants from two different geographical locations. *Early Hum Dev*, 87(7), pp.511–3.

Eckburg, P., Lepp, P. & Relman, D. (2003). Archaea and their potential role in human disease. *Infect Immun*, 71(2), pp.591–596.

Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E. & Relman, D.A. (2005). Diversity of the human intestinal microbial flora. *Science*, 308(5728), pp.1635–8.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19), pp.2460–1.

ELFIN, T.I.G. (2013). Lactoferrin immunoprophylaxis for very preterm infants. *Arch Dis Child Fetal Neonatal Ed*, 98(1), pp.F2–F4.

Embleton, N.D., Turnbull, E., Turner, S. & Berrington, J.E. (2013). Successful blood salvaging from preterm infants: maximizing opportunities, minimizing interventions. *Acta Paediatr*.

Embleton, N.D. & Yates, R. (2008). Probiotics and other preventative strategies for necrotising enterocolitis. *Semin Fetal Neonatal Med*, 13(1), pp.35–43.

Ercolini, D. (2004). PCR-DGGE fingerprinting: novel strategies for detection of microbes in food. *J Microbiol Methods*, 56(3), pp.297–314.

Eriksson, L., Johansson, E. & Kettaneh-Wold, N. (2006). *Multi-and Megavariate Data Analysis, Part 2, Advanced Applications and Method Extensions*, MKS Umetrics AB.

Fairchild, K.D., Tomkoria, S., Sharp, E.C. & Mena, F. V (2002). Neonatal Candida glabrata sepsis: clinical and laboratory features compared with other Candida species. *Pediatr Infect Dis J*, 21(1), pp.39–43.

Fanos, V., Barberini, L., Antonucci, R. & Atzori, L. (2011). Metabolomics in neonatology and pediatrics. *Clin Biochem*, 44(7), pp.452–4.

Favier, C.F.C.F., Vaughan, E.E.E.E., De Vos, W.M., Akkermans, A.D.L.A.D.L. & Vos, Willem M De (2002). Molecular monitoring of succession of bacterial communities in human neonates. *Appl Environ Microbiol*, 68(1), p.219.

Feja, K.N., Wu, F., Roberts, K., Loughrey, M., Nesin, M., Larson, E., Della-Latta, P., Haas, J., Cimiotti, J. & Saiman, L. (2005). Risk factors for candidemia in critically ill infants: a matched case-control study. *J Pediatr*, 147(2), pp.156–161.

Finegold, S.M., Molitoris, D., Song, Y., Liu, C., Vaisanen, M.-L., Bolte, E., McTeague, M., Sandler, R., Wexler, H., Marlowe, E.M., Collins, M.D., Lawson, P. a, Summanen, P., Baysallar, M., Tomzynski, T.J., Read, E., Johnson, E., Rolfe, R., Nasir, P., Shah, H., Haake, D. a, Manning, P. & Kaul, A. (2002). Gastrointestinal microflora studies in late-onset autism. *Clin Infect Dis*, 35(1), pp.S6–S16.

Fogel, G.B., Collins, C.R., Li, J. & Brunk, C.F. (1999). Prokaryotic Genome Size and SSU rDNA Copy Number: Estimation of Microbial Relative Abundance from a Mixed Population. *Microb Ecol*, 38(2), pp.93–113.

Fox, T.P. & Godavitarne, C. (2012). What really causes necrotising enterocolitis? *ISRN Gastroenterol*, 2012, p.628317.

France, M., Cochetière, D. La, Darmaun, D., Rozé, J.C., France de La Cochetiere, M., Rougé, C., Christophe Roze, J., Potel, G. & Gras Leguen, C. (2007). Intestinal microbiota in neonates and preterm infants: a review. *Curr Pediatr Rev*, 3(1), pp.21–34.

Frank, D.N., Zhu, W., Sartor, R.B. & Li, E. (2011). Investigating the biological and clinical significance of human dysbioses. *Trends Microbiol*, 19(9), pp.427–434.

Fromin, N., Hamelin, J., Tarnawski, S., Roesti, D., Jourdain-Miserez, K., Forestier, N., Teyssier-Cuvelle, S., Gillet, F., Aragno, M. & Rossi, P. (2002). Statistical analysis of denaturing gel electrophoresis (DGE) fingerprinting patterns. *Environ Microbiol*, 4(11), pp.634–643.

Fukatsu, T. (2012). Next-generation sequencing sheds light on intricate regulation of insect gut microbiota. *Mol Ecol*, 21(24), pp.5908–10.

Fusunyan, R.D., Nanthakumar, N.N., Baldeon, M.E. & Walker, W. a (2001). Evidence for an innate immune response in the immature human intestine: toll-like receptors on fetal enterocytes. *Pediatr Res*, 49(4), pp.589–93.

Garrido, E., Carrera, E., Manzano, R. & Lopez-Sanroman, A. (2013). Clinical significance of cytomegalovirus infection in patients with inflammatory bowel disease. *World J Gastroenterol*, 19(1), pp.17–25.

Van der Gast, C.J., Walker, A.W., Stressmann, F. a, Rogers, G.B., Scott, P., Daniels, T.W., Carroll, M.P., Parkhill, J. & Bruce, K.D. (2011). Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. *ISME J*, 5(5), pp.780–91.

Gerritsen, J., Smidt, H., Rijkers, G.T. & de Vos, W.M. (2011). Intestinal microbiota in human health and disease: the impact of probiotics. *Genes Nutr*, 6(3), pp.209–40.

Gewolb, I.H., Schwalbe, R.S., Taciak, V.L., Harrison, T.S. & Panigrahi, P. (1999). Stool microflora in extremely low birthweight infants. *Arch Dis Child Fetal Neonatal Ed*, 80(3), pp.F167–F173.

Gibbs, K., Lin, J. & Holzman, I.R. (2007). Necrotising enterocolitis: the state of the science. *Indian J Pediatr*, 74(1), pp.67–72.

Gibson, N. (2006). The use of real-time PCR methods in DNA sequence variation analysis. *Clin Chim Acta*, 363, pp.32–47.

Gophna, U., Sommerfeld, K., Gophna, S., Doolittle, W.F. & Veldhuyzen van Zanten, S.J.O. (2006). Differences between tissue-associated intestinal microfloras of patients with Crohn's disease and ulcerative colitis. *J Clin Microbiol*, 44(11), pp.4136–41.

Green, S.J., Leigh, M.B. & Neufeld, J.D. (2009). *Denaturing gradient gel electrophoresis (DGGE) for microbial community analysis* Microbiolo. K. N. Timmis, ed., Heidelberg: Springer.

Grewe, M., Tsiotos, G.G., Luque de-Leon, E. & Sarr, M.G. (1999). Fungal infection in acute necrotizing pancreatitis. *J Am Coll Surg*, 188(4), pp.408–414.

Grönlund, M.M., Lehtonen, O.P., Eerola, E. & Kero, P. (1999). Fecal microflora in healthy infants born by different methods of delivery: permanent changes in intestinal flora after cesarean delivery. *J Pediatr Gastroenterol Nutr*, 28(1), pp.19–25.

Gude, D. (2012). Biomarkers in sepsis: A comprehensive review. *IJCCI*, 3(2), pp.66–80.

Guthrie, S.O., Gordon, P. V, Thomas, V., Thorp, J. a, Peabody, J. & Clark, R.H. (2003). Necrotizing enterocolitis among neonates in the United States. *J Perinatol*, 23(4), pp.278–85.

Haas, B., Gevers, D. & Earl, A. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res*, 21(3), pp.494–504.

Hall, M. a, Cole, C.B., Smith, S.L., Fuller, R. & Rolles, C.J. (1990). Factors influencing the presence of faecal lactobacilli in early infancy. *Arch Dis Child*, 65(2), pp.185–8.

Halpern, M. & Holubec, H. (2002). Up-regulation of IL-18 and IL-12 in the ileum of neonatal rats with necrotizing enterocolitis. *Pediatr Res*, 51(6), pp.24–26.

Hamady, M. & Knight, R. (2009). Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Res*, 19(7), pp.1141–52.

Hamady, M., Lozupone, C. & Knight, R. (2009). Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J*, 4(1), pp.17–27.

Hammer, Ø. & Harper, D. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeo Electronica*, 4(1), pp.1–9.

Harpavat, S., Pammi, M. & Gilger, M. (2012). Novel treatments for NEC: keeping IBD in mind. *Curr Gastroenterol Rep*, 14(5), pp.373–9.

Harris, K. a. (2003). Development of broad-range 16S rDNA PCR for use in the routine diagnostic clinical microbiology service. *J Med Microbiol*, 52(8), pp.685–691.

Hawrelak, J.A. & Myers, S.P. (2004). The causes of intestinal dysbiosis: a review. *Altern Med Rev*, 9(2), pp.180–97.

Hernandez, G., Mondala, T. & Head, S. (2009). Assessing a novel room temperature RNA storage medium for compatibility in microarray gene expression analysis. *Biotechniques*, 47(2), pp.667–670.

Hoy, C., Millar, M., MacKay, P., Godwin, P., Langdale, V. & Levene, M. (1990). Quantitative changes in faecal microflora preceding necrotising enterocolitis in premature neonates. *Arch Dis Child*, 65(10), pp.1057–59.

Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402), pp.207–214.

Hunt, K.M., Foster, J. a, Forney, L.J., Schütte, U.M.E., Beck, D.L., Abdo, Z., Fox, L.K., Williams, J.E., McGuire, M.K. & McGuire, M.A. (2011). Characterization of the diversity and temporal stability of bacterial communities in human milk. *PLoS One*, 6(6), p.e21313.

Hunter, C.J., Singamsetty, V.K., Chokshi, N.K., Boyle, P., Camerini, V., Grishin, A.V., Upperman, J.S., Ford, H.R. & Prasadarao, N.V. (2008). Enterobacter sakazakii enhances epithelial cell injury by inducing apoptosis in a rat model of necrotizing enterocolitis. *J Infect Dis*, 198(4), p.586.

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

Huse, S.M., Dethlefsen, L., Huber, J. a, Mark Welch, D., Welch, D.M., Relman, D. a & Sogin, M.L. (2008). Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genet*, 4(11), p.e1000255.

Isenbarger, T. a, Carr, C.E., Johnson, S.S., Finney, M., Church, G.M., Gilbert, W., Zuber, M.T. & Ruvkun, G. (2008). The most conserved genome segments for life detection on Earth and other planets. *Orig Life Evol Biosph*, 38(6), pp.517–33.

Jacquot, A., Neveu, D., Aujoulat, F., Mercier, G., Marchandin, H., Jumas-Bilak, E. & Picaud, J.-C. (2011). Dynamics and clinical evolution of bacterial gut microflora in extremely premature patients. *J Pediatr*, 158(3), pp.390–6.

Jalanka-Tuovinen, J., Salonen, A., Nikkilä, J., Immonen, O., Kekkonen, R., Lahti, L., Palva, A. & de Vos, W.M. (2011). Intestinal microbiota in healthy adults: temporal analysis reveals individual and common core and relation to intestinal symptoms. *PLoS One*, 6(7), p.e23035.

Janse, I., Bok, J. & Zwart, G. (2004). A simple remedy against artifactual double bands in denaturing gradient gel electrophoresis. *J Microbiol Methods*, 57(2), pp.279–81.

Jarrell, K.F., Walters, A.D., Bochiwal, C., Borgia, J.M., Dickinson, T. & Chong, J.P.J. (2011). Major players on the microbial stage: why archaea are important. *Microbiology*, 157(Pt 4), pp.919–36.

Jefferies, J.M.C., Cooper, T., Yam, T. & Clarke, S.C. (2012). Pseudomonas aeruginosa outbreaks in the neonatal intensive care unit--a systematic review of risk factors and environmental sources. *J Med Microbiol*, 61(Pt 8), pp.1052–61.

Jilling, T., Simon, D., Lu, J. & Meng, F. (2006). The roles of bacteria and TLR4 in rat and murine models of necrotizing enterocolitis. *J Immunol*, 177(5), pp.3273–3282.

Jiménez, E., Marín, M.L., Martín, R., Odriozola, J.M., Olivares, M., Xaus, J., Fernández, L. & Rodríguez, J.M. (2008). Is meconium from healthy newborns actually sterile? *Res Microbiol*, 159(3), pp.187–93.

Jordan, J. a, Jones-Laughner, J. & Durso, M.B. (2009). Utility of pyrosequencing in identifying bacteria directly from positive blood culture bottles. *J Clin Microbiol*, 47(2), pp.368–72.

Kang, Y., Cheng, J., Mei, L., Hu, J., Piao, Z. & Yin, S. (2010). Multiple copies of 16S rRNA gene affect the restriction patterns and DGGE profile revealed by analysis of genome database. *Mikrobiologii*, 79(5), pp.655–662.

Kaper, J.B., Nataro, J.P. & Mobley, H.L. (2004). Pathogenic Escherichia coli. *Nat Rev Microbiol*, 2(2), pp.123–40.

Karlen, Y., McNair, A., Perseguers, S., Mazza, C. & Mermod, N. (2007). Statistical significance of quantitative PCR. *BMC bioinformatics*, 8.

Kaufman, D. & Fairchild, K.D. (2004). Clinical microbiology of bacterial and fungal sepsis in very-low-birth-weight infants. *Clin Microbiol Rev*, 17(3), p.638.

Kaur, P., Chakraborti, A. & Asea, A. (2010). Enteroaggregative Escherichia coli: An Emerging Enteric Food Borne Pathogen. *Interdiscip Perspect Infect Dis*, 2010, pp.1–10.

Klappenbach, J. a, Dunbar, J.M. & Schmidt, T.M. (2000). rRNA operon copy number reflects ecological strategies of bacteria. *Appl Environ Microbiol*, 66(4), pp.1328–33.

Klappenbach, J. a, Saxman, P.R., Cole, J.R. & Schmidt, T.M. (2001). rrndb: the Ribosomal RNA Operon Copy Number Database. *Nucleic Acids Res*, 29(1), pp.181–4.

Kliegman, R., Walker, W. & Yolken, R. (1993). Necrotizing enterocolitis: research agenda for a disease of unknown etiology and pathogenesis. *Pediatr Res*, 34(6).

Koenig, J.E., Spor, A., Scalfone, N., Fricker, A.D., Stombaugh, J., Knight, R., Angenent, L.T. & Ley, R.E. (2010). Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A*, 108, pp.4578–4585.

Koskinen, R., Ali-Vehmas, T., Kämpfer, P., Laurikkala, M., Tsitko, I., Kostyal, E., Atroshi, F. & Salkinoja-Salonen, M. (2000). Characterization of Sphingomonas isolates from Finnish and Swedish drinking water distribution systems. *J Appl Microbiol*, 89(4), pp.687–96.

Kulik, E. & Sandmeier, H. (2001). Identification of archaeal rDNA from subgingival dental plaque by PCR amplification and sequence analysis. *FEMS Microbiol Lett*, 196(2), pp.129–33.

Kumari, A., Pasini, P. & Daunert, S. (2008). Detection of bacterial quorum sensing N-acyl homoserine lactones in clinical samples. *Anal Bioanal Chem*, 391(5), pp.1619–27.

Lafeber, H.N., Westerbeek, E. a M., van den Berg, A., Fetter, W.P.F. & van Elburg, R.M. (2008). Nutritional factors influencing infections in preterm infants. *J Nutr*, 138(9), p.1813S–1817S.

Lane, D. (1995). 16S/23S rRNA sequencing. In M. Stackebrandt, E. and Goodfellow, ed. *Nucleic Acid Techniques in Bacterial Systematics*. New York: NY, John Wiley and Sons, pp. 115–175.

LaTuga, M.S., Ellis, J.C., Cotton, C.M., Goldberg, R.N., Wynn, J.L., Jackson, R.B. & Seed, P.C. (2011). Beyond Bacteria: A Study of the Enteric Microbial Consortium in Extremely Low Birth Weight Infants A. Driks, ed. *PLoS One*, 6(12), p.e27858.

Lauber, C.L., Zhou, N., Gordon, J.I., Knight, R. & Fierer, N. (2010). Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. *FEMS Microbiol Lett*, 307(1), pp.80–6.

Lee, C., Kim, J., Ahn, S. & Lee, S. (2013). Metabolomic Study of a Diagnostic Model for the Metabolites of Stool Fat. *Korean J Gastroenterol*, 61(1), pp.9–16.

Lee, J. (2011). An update on necrotizing enterocolitis: pathogenesis and preventive strategies. *Korean J Pediatr*, 54(9), pp.368–372.

Lepage, P., Häsler, R., Spehlmann, M.E., Rehman, A., Zvirbliene, A., Begun, A., Ott, S., Kupcinskas, L., Doré, J., Raedler, A. & Schreiber, S. (2011). Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis. *Gastroenterology*, 141(1), pp.227–36.

Ley, R., Bäckhed, F., Turnbaugh, P., Lozupone, C., Knight, R. & Gordon, J. (2005). Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A*, 102(31), pp.11070–5.

Li, D., Rosito, G. & Slagle, T. (2013). Probiotics for the prevention of necrotizing enterocolitis in neonates: an 8-year retrospective cohort study. *J Clin Pharm Ther*.

Lin, H.-C., Hsu, C.-H., Chen, H.-L., Chung, M.-Y., Hsu, J.-F., Lien, R., Tsao, L.-Y., Chen, C.-H. & Su, B.-H. (2008). Oral probiotics prevent necrotizing enterocolitis in very low birth weight preterm infants: a multicenter, randomized, controlled trial. *Pediatrics*, 122(4), pp.693–700.

Lin, P. & Stoll, B. (2006). Necrotising enterocolitis. *Lancet*, 368(9543), pp.1271–1283.

Lin, P.W., Nasr, T.R. & Stoll, B.J. (2008). Necrotizing enterocolitis: recent scientific advances in pathophysiology and prevention. *Semin Perinatol*, 32(2), pp.70–82.

Lindberg, M. (2010). The Human Gastric Microbiota in Health and Disease.

Liu, Y. & Whitman, W.B. (2008). Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. *Ann N Y Acad Sci*, 1125, pp.171–89.

Lodha, A., de Silva, N., Petric, M. & Moore, A.M. (2005). Human torovirus: a new virus associated with neonatal necrotizing enterocolitis. *Acta Paediatr*, 94(8), pp.1085–8.

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. *Nature biotechnology*, 30(5), pp.434–9.

Luo, C., Tsementzi, D., Kyrpides, N., Read, T. & Konstantinidis, K.T. (2012). Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS One*, 7(2), p.e30087.

MacDonald, L. & Baker C (1998). Risk factors for candidemia in a children's hospital. *Clin Infect Dis*, 26(3), pp.642–645.

Macia, L., Thorburn, A.N., Binge, L.C., Marino, E., Rogers, K.E., Maslowski, K.M., Vieira, A.T., Kranich, J. & Mackay, C.R. (2012). Microbial influences on epithelial integrity and immune function as a basis for inflammatory diseases. *Immunol Rev*, 245(1), pp.164–76.

Madan, J.C., Salari, R.C., Saxena, D., Davidson, L., O'Toole, G. a., Moore, J.H., Sogin, M.L., Foster, J. a., Edwards, W.H., Palumbo, P. & Hibberd, P.L. (2012). Gut

microbial colonisation in premature neonates predicts neonatal sepsis. *Arch Dis Child Fetal Neonatal Ed*, 97(6), pp.F456–62.

Magne, F., Suau, A., Pochart, P. & Desjeux, J.-F. (2005). Fecal microbial community in preterm infants. *J Pediatr Gastroenterol Nutr*, 41(4), pp.386–92.

Mai, V., Torrazza, R.M., Ukhanova, M., Wang, X., Sun, Y., Li, N., Shuster, J., Sharma, R., Hudak, M.L. & Neu, J. (2013). Distortions in development of intestinal microbiota associated with late onset sepsis in preterm infants. *PLoS One*, 8(1), p.e52876.

Mai, V., Ukhanova, M. & Baer, D.J. (2010). Understanding the Extent and Sources of Variation in Gut Microbiota Studies; a Prerequisite for Establishing Associations with Disease. *Diversity*, 2(9), pp.1085–1096.

Mai, V., Young, C.M., Ukhanova, M., Wang, X., Sun, Y., Casella, G., Theriaque, D., Li, N., Sharma, R., Hudak, M. & Neu, J. (2011). Fecal microbiota in premature infants prior to necrotizing enterocolitis. *PLoS One*, 6(6), p.e20647.

Mannoia, K., Boskovic, D.S., Slater, L., Plank, M.S., Angeles, D.M. & Gollin, G. (2011). Necrotizing enterocolitis is associated with neonatal intestinal injury. *J Pediatr Surg*, 46(1), pp.81–5.

Manzoni, P., Jacqz-Aigrain, E., Rizzollo, S., Franco, C., Stronati, M., Mostert, M. & Farina, D. (2011). Antifungal prophylaxis in neonates. *Early Hum Dev*, 87(Suppl 1), pp.S59–60.

Manzoni, P., Kaufman, D.A., Mostert, M. & Farina, D. (2008). Neonatal Candida spp. infections: an update. *Pediatric Health*, 2, pp.79–88.

Manzoni, P., Rinaldi, M., Cattani, S., Pugni, L., Romeo, M.G., Messner, H., Stolfi, I., Decembrino, L., Laforgia, N., Vagnarelli, F., Memo, L., Bordignon, L., Saia, O.S., Maule, M., Gallo, E., Mostert, M., Magnani, C., Quercia, M., Bollani, L., Pedicino, R., Renzullo, L., Betta, P., Mosca, F., Ferrari, F., Magaldi, R., Stronati, M. & Farina, D. (2009). Bovine lactoferrin supplementation for prevention of late-onset sepsis in very low-birth-weight neonates: a randomized trial. *JAMA*, 302(13), pp.1421–8.

Marchesi, J.R., Holmes, E., Khan, F., Kochhar, S., Scanlan, P., Shanahan, F., Wilson, I.D. & Wang, Y. (2007). Rapid and Noninvasive Metabonomic Characterization of Inflammatory Bowel Disease research articles. *J Proteome Res*, pp.546–551.

Mardis, E. (2008). Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet*, 9, pp.387–402.

Mardis, E.R. (2008). The impact of next-generation sequencing technology on genetics. *Trends Genet*, 24(3), pp.133–41.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L. a, Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X. V, Godwin, B.C., He, W., Helgesen, S., Ho, Chun Heen, Ho, Chun He,

Irzyk, G.P., Jando, S.C. et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057), pp.376–80.

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. *Curr Opin Biotechnol*, 21(2), pp.149–56.

Martin, A. (2002). Phylogenetic approaches for describing and comparing the diversity of microbial communities. *Appl Environ Microbiol*, 68(8), pp.3673–82.

Martin, C.R. & Walker, W.A. (2006). Intestinal immune defences and the inflammatory response in necrotising enterocolitis. *Semin Fetal Neonatal Med*, 11(5), pp.369–77.

Martin, C.R. & Walker, W.A. (2008). Probiotics: role in pathophysiology and prevention in necrotizing enterocolitis. *Semin Perinatol*, 32(2), pp.127–37.

Martín, R., Jiménez, E., Heilig, H., Fernández, L., Marín, M.L., Zoetendal, E.G. & Rodríguez, J.M. (2009). Isolation of bifidobacteria from breast milk and assessment of the bifidobacterial population by PCR-denaturing gradient gel electrophoresis and quantitative real-time PCR. *Appl Environ Microbiol*, 75(4), pp.965–9.

Matsuda, K., Tsuji, H., Asahara, T., Matsumoto, K., Takada, T. & Nomoto, K. (2009). Establishment of an analytical system for the human fecal microbiota, based on reverse transcription-quantitative PCR targeting of multicopy rRNA molecules. *Appl Environ Microbiol*, 75(7), pp.1961–9.

Michu, E., Mráčková, M., Vyskot, B. & Žlůvová, J. (2010). Reduction of heteroduplex formation in PCR amplification. *Biologia Plantarum*, 54(1), pp.173–176.

Miller, T.L. & Wolin, M.J. (1986). Methanogens in human and animal intestinal Tracts. *Systematic and Applied Microbiology*, 7(2-3), pp.223–229.

Morgan, J.L., Darling, A.E. & Eisen, J. a (2010). Metagenomic sequencing of an in vitro-simulated microbial community. *PLoS One*, 5(4), p.e10209.

Morowitz, M.J., Poroyko, V., Caplan, M., Alverdy, J. & Liu, D.C. (2010). Redefining the role of intestinal microbes in the pathogenesis of necrotizing enterocolitis. *Pediatrics*, 125(4), pp.777–85.

Morowitz, M.J.M.J., Denef, V.J.V.J., Costello, E.K.E.K., Thomas, B.C.B.C., Poroyko, V., Relman, D.A.D. a & Banfield, J.F.J.F. (2011). Strain-resolved community genomic analysis of gut microbial colonization in a premature infant. *Proc Natl Acad Sci U S A*, 108(3), pp.1128–34.

Morrow, A.L., Lagomarcino, A.J., Schibler, K.R., Taft, D.H., Yu, Z., Wang, B., Altaye, M., Wagner, M., Gevers, D., Ward, D. V, Kennedy, M. a, Huttenhower, C. & Newburg, D.S. (2013). Early microbial and metabolomic signatures predict later onset of necrotizing enterocolitis in preterm infants. *Microbiome*, 1(1), p.13.

Mshvildadze, M., Neu, J., Shuster, J., Theriaque, D., Li, N. & Mai, V. (2010). Intestinal microbial ecology in premature infants assessed with non-culture-based techniques. *J Pediatr*, 156(1), pp.20–5.

Mullis, K., Faloona, F. & Scharf, S. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol.*, 51, pp.263–73.

Mussap, M., Noto, A., Cibecchini, F. & Fanos, V. (2012). The importance of biomarkers in neonatology. *Semin Fetal Neonatal Med*, 18, pp.56–64.

Muyzer, G. & Smalla, K. (1998). Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie van Leeuwenhoek*, 73(1), pp.127–41.

Muyzer, G., Teske, A., Wirsen, C.O. & Jannasch, H.W. (1995). Phylogenetic relationships of Thiomicrospira species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Arch Microbiol*, 164(3), pp.165–72.

Muyzer, G., de Waal, E.C. & Uitterlinden, A.G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol*, 59(3), pp.695–700.

Na, H., Kim, O.-S., Yoon, S.-H., Kim, Y. & Chun, J. (2011). Comparative approach to capture bacterial diversity of coastal waters. *J Microbiol*, 49(5), pp.729–40.

Nair, V. & Soraisham, A. (2013). Probiotics and prebiotics: role in prevention of nosocomial sepsis in preterm infants. *Int J pediatr*, 2013.

Nakamura, N., Gaskins, H.R., Collier, C.T., Nava, G.M., Rai, D., Petschow, B., Russell, W.M., Harris, C., Mackie, R.I., Wampler, J.L. & Walker, D.C. (2009). Molecular ecological analysis of fecal bacterial populations from term infants fed formula supplemented with selected blends of prebiotics. *Appl Environ Microbiol*, 75(4), pp.1121–8.

Nam, Y.-D., Chang, H.-W., Kim, K.-H., Roh, S.W., Kim, M.-S., Jung, M.-J., Lee, S.-W., Kim, J.-Y., Yoon, J.-H. & Bae, J.-W. (2008). Bacterial, archaeal, and eukaryal diversity in the intestines of Korean people. *J Microbiol*, 46(5), pp.491–501.

Nanthakumar, N., Meng, D., Goldstein, A. & Zhu, W. (2011). The mechanism of excessive intestinal inflammation in necrotizing enterocolitis: an immature innate immune response J. Neu, ed. *PLoS One*, 6(3), p.e17776.

Nanthakumar, N.N., Fusunyan, R.D., Sanderson, I. & Walker, W. a (2000). Inflammation in the developing human intestine: A possible pathophysiologic contribution to necrotizing enterocolitis. *Proc Natl Acad Sci U S A*, 97(11), pp.6043–8.

Neish, A.S. (2009). Microbes in gastrointestinal health and disease. *Gastroenterology*, 136(1), pp.65–80.

Nelson, A., De Soyza, A., Bourke, S.J., Perry, J.D. & Cummings, S.P. (2010). Assessment of sample handling practices on microbial activity in sputum samples from patients with cystic fibrosis. *Lett Appl Microbiol*, 51(3), pp.272–7.

Nelson, A., De Soyza, A., Perry, J.D., Sutcliffe, I.C. & Cummings, S.P. (2012). Polymicrobial challenges to Koch's postulates: ecological lessons from the bacterial vaginosis and cystic fibrosis microbiomes. *Innate Immun*, 18(5), pp.774–83.

Neu, J. (2005). Neonatal necrotizing enterocolitis: an update. *Acta Paediatr*, 94(449), pp.100–5.

Neu, J. (2011). Routine probiotics for premature infants: let's be careful! *J Pediatr.*, 158(4), pp.672–4.

Ng, P., Ang, I., Chiu, R. & Li, K. (2010). Host-response biomarkers for diagnosis of late-onset septicemia and necrotizing enterocolitis in preterm infants. *J Clin Invest*, 120(8), pp.2989–3000.

Ng, P.C., Li, G., Chui, K.M., Chu, W.C.W., Li, K., Wong, R.P.O., Chik, K.W., Wong, E. & Fok, T.F. (2004). Neutrophil CD64 is a sensitive diagnostic marker for early-onset neonatal infection. *Pediatr Res*, 56(5), pp.796–803.

Ng, P.C., Li, K., Wong, R.P.O., Chui, K., Wong, E., Li, G. & Fok, T.F. (2003). Proinflammatory and anti-inflammatory cytokine responses in preterm infants with systemic infections. *Arch Dis Child Fetal Neonatal Ed*, 88(3), pp.F209–13.

Nübel, U., Engelen, B., Felske, A., Snaidr, J., Wieshuber, A., Amann, R.I., Ludwig, W. & Backhaus, H. (1996). Sequence heterogeneities of genes encoding 16S rRNAs in Paenibacillus polymyxa detected by temperature gradient gel electrophoresis. *J Bacteriol*, 178(19), pp.5636–43.

O'Sullivan, D. (2000). Methods for analysis of the intestinal microflora. *Curr Issues Intest Microbiol*, 1(2), pp.39–50.

Ochoa, T.J., Brown, E.L., Guion, C.E., Chen, J.Z., McMahon, R.J. & Cleary, T.G. (2006). Effect of lactoferrin on enteroaggregative E. coli (EAEC). *Biochem Cell Biol.*, 84(3), pp.369–76.

Palmer, C., Bik, E.M., DiGiulio, D.B., Relman, D.A. & Brown, P.O. (2007). Development of the human infant intestinal microbiota. *PLoS Biol*, 5(7), pp.1556–73.

Park, H.-K., Shim, S.-S., Kim, S.-Y., Park, J.-H., Park, S.-E., Kim, H.-J., Kang, B.-C. & Kim, C.-M. (2005). Molecular analysis of colonized bacteria in a human newborn infant gut. *J Microbiol*, 43(4), pp.345–53.

Pärtty, A., Luoto, R., Kalliomäki, M., Salminen, S. & Isolauri, E. (2013). Effects of Early Prebiotic and Probiotic Supplementation on Development of Gut Microbiota and Fussing and Crying in Preterm Infants: A Randomized, Double-Blind, Placebo-Controlled Trial. *J Pediatr*.

Patel, R., Lin, M., Laney, M., Kurn, N., Rose, S. & Ullman, E.F. (1996). Formation of chimeric DNA primer extension products by template switching onto an annealed downstream oligonucleotide. *Proc Natl Acad Sci U S A*, 93(7), pp.2969–74.

Penders, J., Thijs, C., Vink, C., Stelma, F.F., Snijders, B., Kummeling, I., van Den Brandt, P.A. & Stobberingh, E.E. (2006). Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics*, 118(2), pp.511–21.

Penders, J., Vink, C., Driessen, C., London, N., Thijs, C. & Stobberingh, E.E. (2005). Quantification of Bifidobacterium spp., Escherichia coli and Clostridium difficile in faecal samples of breast-fed and formula-fed infants by real-time PCR. *FEMS Microbiol Lett*, 243(1), pp.141–7.

Pérez-Cobas, A.E., Gosalbes, M.J., Friedrichs, A., Knecht, H., Artacho, A., Eismann, K., Otto, W., Rojo, D., Bargiela, R., von Bergen, M., Neulinger, S.C., Däumer, C., Heinsen, F.-A., Latorre, A., Barbas, C., Seifert, J., Dos Santos, V.M., Ott, S.J., Ferrer, M. & Moya, A. (2012). Gut microbiota disturbance during antibiotic therapy: a multi-omic approach. *Gut*, pp.1–11.

Periasamy, S. & Kolenbrander, P.E. (2010). Central role of the early colonizer Veillonella sp. in establishing multispecies biofilm communities with initial, middle, and late colonizers of enamel. *J Bacteriol*, 192(12), pp.2965–72.

Peris-Bondia, F., Latorre, A., Artacho, A., Moya, A. & D'Auria, G. (2011). The Active Human Gut Microbiota Differs from the Total Microbiota M. M. Heimesaat, ed. *PLoS One*, 6(7), p.e22448.

Petrosino, J.F., Highlander, S., Luna, R.A., Gibbs, R. a & Versalovic, J. (2009). Metagenomic pyrosequencing and microbial identification. *Clin Chem*, 55(5), pp.856–66.

Picard, C., Fioramonti, J., Francois, A., Robinson, T., Neant, F. & Matuchansky, C. (2005). Review article: bifidobacteria as probiotic agents -- physiological effects and clinical benefits. *Aliment Pharmacol Ther*, 22(6), pp.495–512.

Pierrakos, C. & Vincent, J. (2010). Sepsis biomarkers: a review. *Crit Care*, 14(R15), pp.1–18.

Price, M., Dehal, P. & Arkin, A. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One*, 5(3), p.e9490.

Procianoy, R. & Silveira, R. (2004). The role of sample collection timing on interleukin-6 levels in early-onset neonatal sepsis. *J Pediatr (Rio J)*, 80(5), pp.407–410.

Prosser, J.I., Bohannan, B.J.M., Curtis, T.P., Ellis, R.J., Firestone, M.K., Freckleton, R.P., Green, J.L., Green, L.E., Killham, K., Lennon, J.J., Osborn, a M., Solan, M., van der Gast, C.J. & Young, J.P.W. (2007). The role of ecological theory in microbial ecology. *Nat Rev Microbiol*, 5(5), pp.384–92.

Rajilić-Stojanović, M., Smidt, H. & de Vos, W.M. (2007). Diversity of the human gastrointestinal tract microbiota revisited. *Environ Microbiol*, 9(9), pp.2125–36.

Rakoff-Nahoum, S. & Paglino, J. (2004). Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell*, 118, pp.229–241.

Rieu-Lesme, F., Delbès, C. & Sollelis, L. (2005). Recovery of partial 16S rDNA sequences suggests the presence of Crenarchaeota in the human digestive ecosystem. *Curr Microbiol*, 51(5), pp.317–21.

Riskin, A., Hochwald, O., Bader, D., Srugo, I., Naftali, G., Kugelman, A., Cohen, E., Mor, F., Kaufman, B. & Shaoul, R. (2010). The effects of lactulose supplementation to enteral feedings in premature infants: a pilot study. *J Pediatr*, 156(2), pp.209–14.

Roger, L.C., Costabile, A., Holland, D.T., Hoyles, L. & McCartney, A.L. (2010). Examination of faecal Bifidobacterium populations in breast- and formula-fed infants during the first 18 months of life. *Microbiology*, 156(Pt 11), pp.3329–41.

Rolhion, N. & Darfeuille-Michaud, A. (2007). Adherent-invasive Escherichia coli in inflammatory bowel disease. *Inflamm Bowel Dis*, 13(10), pp.1277–83.

Ronaghi, M. (2001). Pyrosequencing sheds light on DNA sequencing. *Genome Res*, 11(1), pp.3–11.

Ronaghi, M. & Karamohamed, S. (1996). Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem*, 242(1), pp.84–89.

Sakata, H., Yoshioka, H. & Fujita, K. (1985). Development of the intestinal flora in very low birth weight infants compared to normal full-term newborns. *Eur J Pediatr*, 144(2), pp.186–90.

Salminen, S., Gibson, G.R., McCartney, A.L. & Isolauri, E. (2004). Influence of mode of delivery on gut microbiota composition in seven year old children. *Gut*, 53(9), pp.1388–9.

Samuel, B.S. & Gordon, J.I. (2006). A humanized gnotobiotic mouse model of host-archaeal-bacterial mutualism. *Proc Natl Acad Sci U S A*, 103(26), pp.10011–6.

Samuelson, J. (1999). Why metronidazole is active against both bacteria and parasites. *Antimicrob Agents Chemother*, 43(7), pp.1533–41.

Sandhu, G.S., Kline, B.C., Stockman, L. & Roberts, G.D. (1995). Molecular probes for diagnosis of fungal infections. *J Clin Microbiol*, 33(11), pp.2913–9.

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

Santulli, T. V., Schullinger, J.N., Heird, W.C., Gongaware, R.D., Wigger, J., Barlow, B., Blanc, W.A. & Berdon, W.E. (1975). Acute Necrotizing Enterocolitis in Infancy: A Review of 64 Cases. *Pediatrics*, 55(3), pp.376–387.

Sasaki, M., Sitaraman, S. V, Babbin, B. a, Gerner-Smidt, P., Ribot, E.M., Garrett, N., Alpern, J. a, Akyildiz, A., Theiss, A.L., Nusrat, A. & Klapproth, J.-M. a (2007).

Invasive Escherichia coli are a feature of Crohn's disease. *Lab Invest*, 87(10), pp.1042–54.

Savino, F., Roana, J., Mandras, N., Tarasco, V., Locatelli, E. & Tullio, V. (2011). Faecal microbiota in breast-fed infants after antibiotic therapy. *Acta Paediatr*, 100(1), pp.75–8.

Scanlan, P.D., Shanahan, F. & Marchesi, J.R. (2008). Human methanogen diversity and incidence in healthy and diseased colonic groups using mcrA gene analysis. *BMC Microbiol*, 8, p.79.

Scanlan, P.D., Shanahan, F., O'Mahony, C. & Marchesi, J.R. (2006). Culture-independent analyses of temporal variation of the dominant fecal microbiota and targeted bacterial subgroups in Crohn's disease. *J Clin Microbiol*, 44(11), pp.3980–8.

Schatz, M.C., Delcher, A.L. & Salzberg, S.L. (2010). Assembly of large genomes using second-generation sequencing. *Genome Res*, 20(9), pp.1165–73.

Schloss, P. (2009). A high-throughput DNA sequence aligner for microbial ecology studies. *PLoS One*, 4(12), p.e8230.

Schloss, P.D., Gevers, D. & Westcott, S.L. (2011). Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies. *PLoS One*, 6(12), p.e27310.

Schwiertz, A., Gruhl, B., Löbnitz, 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), pp.393–9.

Seekatz, A.M., Panda, A., Rasko, D.A., Toapanta, F.R., Eloe-Fadrosh, E.A., Khan, A.Q., Liu, Z., Shipley, S.T., Detolla, L.J., Sztein, M.B. & Fraser, C.M. (2013). Differential response of the cynomolgus macaque gut microbiota to Shigella infection. R. M. R. II, ed. *PLoS One*, 8(6), p.e64212.

Sekirov, I., Russell, S., Antunes, L. & Finlay, B. (2010). Gut microbiota in health and disease. *Physiol Rev*, 90(3), pp.859–904.

Sellitto, M., Bai, G., Serena, G., Fricke, W.F., Sturgeon, C., Gajer, P., White, J.R., Koenig, S.S.K., Sakamoto, J., Boothe, D., Gicquelais, R., Kryszak, D., Puppa, E., Catassi, C., Ravel, J. & Fasano, A. (2012). Proof of concept of microbiome-metabolome analysis and delayed gluten exposure on celiac disease autoimmunity in genetically at-risk infants. *PLoS One*, 7(3), p.e33387.

Seng, P., Drancourt, M., Gouriet, F., La Scola, B., Fournier, P.-E., Rolain, J.M. & Raoult, D. (2009). Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin Infect Dis*, 49(4), pp.543–51.

Shanahan, F. (2010). Probiotics in perspective. *Gastroenterology*, 139(6), pp.1808–12.

Sharma, S., Szele, Z. & Schilling, R. (2006). Influence of freeze-thaw stress on the structure and function of microbial communities and denitrifying populations in soil. *Appl Environ Microbiol*, 72(3), pp.2148–2154.

Shendure, J. & Ji, H. (2008). Next-generation DNA sequencing. *Nature biotechnology*, 26(10), pp.1135–1145.

Shuldiner, A., Nirula, A. & Roth, J. (1989). Hybrid DNA artifact from PCR of closely related target sequences. *Nucleic Acids Res*, 17(11), p.4409.

Sim, K., Cox, M.J., Wopereis, H., Martin, R., Knol, J., Li, M.-S., Cookson, W.O.C.M., Moffatt, M.F. & Kroll, J.S. (2012). Improved Detection of Bifidobacteria with Optimised 16S rRNA-Gene Based Pyrosequencing. *PLoS One*, 7(3), p.e32543.

Simões, C. & Maukonen, J. (2013). Habitual Dietary Intake Is Associated with Stool Microbiota Composition in Monozygotic Twins. *J Nutr*, 143(3), pp.417–23.

Siqueira, J.F., Fouad, A.F. & Rôças, I.N. (2012). Pyrosequencing as a tool for better understanding of human microbiomes. *J Oral Microbiol*, 4, pp.1–15.

Small, C.-L.N., Reid-Yu, S.A., McPhee, J.B. & Coombes, B.K. (2013). Persistent infection with Crohn's disease-associated adherent-invasive Escherichia coli leads to chronic inflammation and intestinal fibrosis. *Nat Commun*, 4, p.1957.

Smith, B., Bodé, S., Skov, T.H., Mirsepasi, H., Greisen, G. & Krogfelt, K.A. (2012). Investigation of the early intestinal microflora in premature infants with/without necrotizing enterocolitis using two different methods. *Pediatr Res*, 71(1), pp.115–120.

Smith, C. & Osborn, A. (2009). Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. *FEMS Microbiol Ecol*, 67(1), pp.6–20.

Smith, S.D., Tagge, E.P., Miller, J., Cheu, H., Sukarochana, K. & Rowe, M.I. (1990). The hidden mortality in surgically treated necrotizing enterocolitis: fungal sepsis. *J Pediatr Surg*, 25(10), pp.1030–3.

Sobhani, I., Tap, J., Roudot-Thoraval, F., Roperch, J.P., Letulle, S., Langella, P., Corthier, G., Tran Van Nhieu, J. & Furet, J.P. (2011). Microbial dysbiosis in colorectal cancer (CRC) patients. *PLoS One*, 6(1), p.e16393.

Standage, S. & Wong, H. (2011). Biomarkers for pediatric sepsis and septic shock. *Expert Rev Anti Infect Ther*, 9(1), pp.71–79.

Stark, P. & Lee, A. (1982). The microbial ecology of the large bowel of breastfed and formula-fed infants during the first year of life. *Med Microbiol*, 15(2), pp.189–203.

Stenger, M.R., Reber, K.M., Giannone, P.J. & Nankervis, C. a (2011). Probiotics and prebiotics for the prevention of necrotizing enterocolitis. *Curr Infect Dis Rep*, 13(1), pp.13–20.

Stewart, C.J., Nelson, A., Scribbins, D., Marrs, E.C.L., Lanyon, C., Perry, J.D., Embleton, N.D., Cummings, S.P. & Berrington, J.E (2013). Bacterial and fungal viability in the preterm gut: NEC and sepsis. *Arch Dis Child Fetal Neonatal Ed*, 98(4), pp.F298–303.

Stewart, C.J., Marrs, E.C.L., Nelson, A., Lanyon, C., Perry, J.D., Embleton, N.D., Cummings, S.P. & Berrington, J.E. (2013). Development of the Preterm Gut Microbiome in Twins at Risk of Necrotising Enterocolitis and Sepsis Y. Sanz, ed. *PLoS One*, 8(8), p.e73465.

Stewart, J., Chadwick, V. & Murray, A. (2005). Investigations into the influence of host genetics on the predominant eubacteria in the faecal microflora of children. *J Med Microbiol*, 54(12), pp.1239–42.

Stoll, B.J. (1994). Epidemiology of necrotizing enterocolitis. *Clin Perinatol*, 21(2), pp.205–18.

Stoll, B.J., Hansen, N.I., Adams-Chapman, I., Fanaroff, A. a, Hintz, S.R., Vohr, B. & Higgins, R.D. (2004). Neurodevelopmental and growth impairment among extremely low-birth-weight infants with neonatal infection. *JAMA*, 292(19), pp.2357–65.

Streimish, I. & Bizzarro, M. (2012). Neutrophil CD64 as a diagnostic marker in neonatal sepsis. *Pediatr Infect Dis J*, 31(7), pp.777–781.

Struss, A.K., Pasini, P., Flomenhoft, D., Shashidhar, H. & Daunert, S. (2012). Investigating the effect of antibiotics on quorum sensing with whole-cell biosensing systems. *Anal Bioanal Chem*, 402(10), pp.3227–36.

Sud, I.J. & Feingold, D.S. (1982). Action of antifungal imidazoles on Staphylococcus aureus. *Antimicrob Agents Chemother*, 22(3), pp.470–4.

Sun, Y., Cai, Y., Liu, L., Yu, F., Farrell, M.L., McKendree, W. & Farmerie, W. (2009). ESPRIT: estimating species richness using large collections of 16S rRNA pyrosequences. *Nucleic Acids Res*, 37(10), pp.e76–e76.

Sutcliffe, I.C., Trujillo, M.E. & Goodfellow, M. (2012). A call to arms for systematists: revitalising the purpose and practises underpinning the description of novel microbial taxa. *Antonie van Leeuwenhoek*, 101(1), pp.13–20.

Swidsinski, A., Loening-Baucke, V., Vaneechoutte, M. & Doerffel, Y. (2008). Active Crohn's disease and ulcerative colitis can be specifically diagnosed and monitored based on the biostructure of the fecal flora. *Inflamm Bowel Dis*, 14(2), pp.147–61.

Szajewska, H. (2010). Probiotics and prebiotics in preterm infants: where are we? Where are we going? *Early Hum Dev*, 86 Suppl 1, pp.81–6.

Tamrakar, R., Yamada, T., Furuta, I., Cho, K., Morikawa, M., Yamada, H., Sakuragi, N. & Minakami, H. (2007). Association between Lactobacillus species and bacterial vaginosis-related bacteria, and bacterial vaginosis scores in pregnant Japanese women. *BMC Infect Dis*, 7(128), pp.1–8.

Tannock, G.W., Munro, K., Bibiloni, R., Simon, M.A., Hargreaves, P., Gopal, P., Harmsen, H. & Welling, G. (2004). Impact of consumption of oligosaccharide-containing biscuits on the fecal microbiota of humans. *Appl Environ Microbiol*, 70(4), p.2129.

Temmerman, R. & Masco, L. (2003). Development and validation of a nested-PCR-denaturing gradient gel electrophoresis method for taxonomic characterization of bifidobacterial communities. *Appl Environ Microbiol*, 69(11), pp.6380–6385.

Thingstad, T. & Lignell, R. (1997). Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat Microb Ecol*, 13, pp.19–27.

Thingstad, T.F. (2000). Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol Oceanog*, 45(6), pp.1320–1328.

Thompson, J.R., Marcelino, L. a & Polz, M.F. (2002). Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by "reconditioning PCR". *Nucleic Acids Res*, 30(9), pp.2083–8.

Thompson-Chagoyán, O.C., Maldonado, J. & Gil, A. (2007). Colonization and impact of disease and other factors on intestinal microbiota. *Dig Dis Sci*, 52(9), pp.2069–77.

Thuijls, G., Derikx, J.P.M., van Wijck, K., Zimmermann, L.J.I., Degraeuwe, P.L., Mulder, T.L., Van der Zee, D.C., Brouwers, H. a a, Verhoeven, B.H., van Heurn, L.W.E., Kramer, B.W., Buurman, W. a & Heineman, E. (2010). Non-invasive markers for early diagnosis and determination of the severity of necrotizing enterocolitis. *Ann Surg*, 251(6), pp.1174–80.

Thurber, R. V, Haynes, M., Breitbart, M., Wegley, L. & Rohwer, F. (2009). Laboratory procedures to generate viral metagenomes. *Nat Protoc*, 4(4), pp.470–83.

Tirodker, U.H., Nataro, J.P., Smith, S., LasCasas, L. & Fairchild, K.D. (2003). Detection of fungemia by polymerase chain reaction in critically ill neonates and children. *J Perinatol*, 23(2), pp.117–22.

Tourlomousis, P. & Kemsley, E. (2010). electrophoresis of complex microbial communities: A two-step approach to address the effect of gel-to-gel variation and allow valid comparisons across a large dataset. *Microb Ecol*, 59(4), pp.776–86.

Tran, L., Ferris, M., Norori, J., Stark, M., Craver, R., Dowd, S. & Penn, D. (2013). Necrotizing enterocolitis and cytomegalovirus infection in a premature infant. *Pediatrics*, 131(1), pp.e318–22.

Trosvik, P., Stenseth, N.C. & Rudi, K. (2010). Convergent temporal dynamics of the human infant gut microbiota. *ISME J*, 4(2), pp.151–8.

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), pp.804–10.

Valasek, M. & Repa, J. (2005). The power of real-time PCR. *Adv Physiol Educ*, 29(3), pp.151–159.

Vaughan, E.E., Schut, F., Heilig, H., Zoetendal, E.G., de Vos, W.M. & Akkermans, A.D.L. (2000). A molecular view of the intestinal ecosystem. *Curr Issues Intest Microbiol*, 1(1), pp.1–12.

Vaz-Moreira, I., Egas, C., Nunes, O.C. & Manaia, C.M. (2011). Culture-dependent and culture-independent diversity surveys target different bacteria: a case study in a freshwater sample. *Antonie van Leeuwenhoek*, 100(2), pp.245–57.

Venkatesh, M.P. & Abrams, S. a (2010). Oral lactoferrin for the prevention of sepsis and necrotizing enterocolitis in preterm infants. *Cochrane Database Syst Rev*, 5(6), p.CD007137.

Vergnano, S., Menson, E., Kennea, N., Embleton, N., Russell, A.B., Watts, T., Robinson, M.J., Collinson, A. & Heath, P.T. (2011). Neonatal infections in England: the NeonIN surveillance network. *Arch Dis Child Fetal Neonatal Ed*, 96(1), pp.F9–F14.

Vianna, M. & Conrads, G. (2006). Identification and quantification of archaea involved in primary endodontic infections. *J Clin Microbiol*, 44(4), pp.1274–1282.

Wang, Q., Garrity, G.M., Tiedje, J.M. & Cole, J.R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol*, 73(16), pp.5261–7.

Wang, Y., Hoenig, J.D., Malin, K.J., Qamar, S., Petrof, E.O., Sun, J., Antonopoulos, D.A., Chang, E.B. & Claud, E.C. (2009). 16S rRNA gene-based analysis of fecal microbiota from preterm infants with and without necrotizing enterocolitis. *ISME J*, 3(8), pp.944–54.

Wang, Y. & Qian, P.-Y. (2009). Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PLoS One*, 4(10), p.e7401.

Westra-Meijer, C.M., Degener, J.E., Dzoljic-Danilovic, G., Michel, M.F. & Mettau, J.W. (1983). Quantitative study of the aerobic and anaerobic faecal flora in neonatal necrotising enterocolitis. *Arch Dis Child*, 58(7), pp.523–528.

Winter, S.E., Winter, M.G., Xavier, M.N., Thiennimitr, P., Poon, V., Keestra, A.M., Laughlin, R.C., Gomez, G., Wu, J., Lawhon, S.D., Popova, I.E., Parikh, S.J., Adams, L.G., Tsolis, R.M., Stewart, V.J. & Bäumler, A.J. (2013). Host-derived nitrate boosts growth of E. coli in the inflamed gut. *Science*, 339(6120), pp.708–11.

Wu, S.-F., Caplan, M. & Lin, H.-C. (2012). Necrotizing Enterocolitis: Old Problem with New Hope. *Pediatr Neonatol*, 53(3), pp.158–163.

Young, C., Sharma, R., Handfield, M., Mai, V. & Neu, J. (2009). Biomarkers for infants at risk for necrotizing enterocolitis: clues to prevention? *Pediatr Res*, 65(5), pp.91–97.

Yue, J.C. & Clayton, M.K. (2005). A Similarity Measure Based on Species Proportions. *Commun stat A-Theor*, 34(11), pp.2123–2131.

Zaoutis, T.E., Heydon, K., Localio, R., Walsh, T.J. & Feudtner, C. (2007). Outcomes attributable to neonatal candidiasis. *Clin Infect Dis*, 44(9), pp.1187–93.

Zoetendal, E., Akkermans, A., Akkermans-van Vliet, W., Visser, J. & Vos, W. (2001). The host genotype affects the bacterial community in the human gastronintestinal tract. *Microb Ecol Health Dis*, 13(3), pp.129–134.

Zoetendal, E.G., Vaughan, E.E. & de Vos, W.M. (2006). A microbial world within us. *Mol Microbiol*, 59(6), pp.1639–50.

Appendices

Appendix 1 - Enzymatic lysis buffer

Pre-treatment of Gram-positive bacteria (DNeasy Blood and Tissue Kit) enzymatic lysis buffer:

20 mM Tris·Cl, pH 8.0

2 mM sodium EDTA

1.2% Triton® X-100

Immediately before use, add lysozyme to 20 mg/mL

Appendix 2 – 50× TAE Buffer

TAE buffer was prepared at 50x concentrate then diluted as required.

Step 1 – 200 mL EDTA pH 8.0

37.22 g EDTA

200 mL dH_2O

A beaker containing the EDTA and ~150 mL dH_2O was placed on to a magnetic stirrer and the pH was measured throughout. Sodium hydroxide pellets were added to the solution until the solution was at pH 8.0. dH_2O was added to achieve a final volume of 200 mL.

Step 2 – Make up 2 L

484g Tris base ultrapure

114.2 mL Glacial acetic acid

200 mL EDTA pH 8.0

dH_2O to 2L

The Tris base was weighed and placed into a 1L Duran bottle along with the glacial acetic acid and the EDTA which was prepared fresh as described above. dH_2O was added to achieve a final volume of 2 L.

To make 1 L of 1× TAE: dilute 20 mL of 50x TAE in 980 mL dH2O.

Appendix 3 – Dyes

Bromophenol blue (6x concentrate)

Bromophenol blue was prepared at 6× concentrate and diluted appropriately with the sample as required.

0.025 g Bromophenol blue

4.0 g Sucrose

dH_2O to 1 L

DCode Dye

DCode dye was added to the high denaturing solution so that the efficacy of mixing between denaturing solutions in DGGE gel pouring could be established.

0.05 g Bromophenol blue

0.05 g Xylene cyanol

1× TAE to 10 mL

DGGE loading dye (2× concentrate)

DGGE loading dye was prepared at 2× concentrate and diluted appropriately with the sample as required.

Step 1 - 2% (w/v) solution of bromophenol blue and xylene cyanol

A 2% (w/v) solution of bromophenol blue and a 2% (w/v) solution of xylene cyanol were prepared by dissolving 0.002g of each solid in 1 mL dH_2O.

Step 2 – Make up 10 mL

0.25ml 2% (w/v) bromophenol blue

0.25ml 2% (w/v) xylene cyanol

7.0ml 100% glycerol

2.5ml dH2O

Appendix 4 – DGGE denaturing solutions

Table showing how to prepare each denaturing solution

Reagent	Bacterial		Fungal	
	34%	55%	40%	60%
40% (v/v) acrylamide (37.5:1 acrylamide:bisacrylamide)	30 mL	30 mL	30 mL	30 mL
50x TAE	2 mL	2 mL	2 mL	2 mL
Deionised formamide	13.6 mL	22 mL	16 mL	24 mL
Urea (electrophoresis grade)	14.28 g	23.1 g	16.8 g	25.2 g
dH2O	To 100 mL	To 100 mL	To 100 mL	To 100m L

Appendix 5 – SOC media

Step 1 – Prepare solutions

1M NaCl

0.5844 g NaCl

dH_2O to 10 mL

1M KCl

0.7455 g KCl

DH_2O to 10 mL

2M Mg2+ stock

2.330 g $MgCl_2 \cdot 6H_2O$

2.465 g $MgSO_4 \cdot 7H_2O$

dH_2O to 10 mL filter sterilise with a 0.22 μM filter

2M glucose

3.603 g Glucose

dH_2O to 10 mL filter sterilise with a 0.22 μM filter

Step 2 – Make the media

To make the media add;

2.0 g Tryptone

0.5 g Yeast extract

1 mL 1M NaCl

1 mL 1M KCl

dH_2O to 100 mL

Autoclave and allow to cool to room temperature. Then add;

1 mL 2M Mg^{2+}

1 mL 2M Glucose

Check the pH is 7.0, adjust accordingly if it is not.

Appendix 5 – Luria-Bertani media

Basic recipe (per 1 L)

Tryptone 10 g

Yeast Extract 5 g

Sodium Chloride 5 g

Agar 15 g (For broth omit agar from the recipe)

Autoclave and allow to cool to 50 °C and pour ~20 mL in to each Petri plate

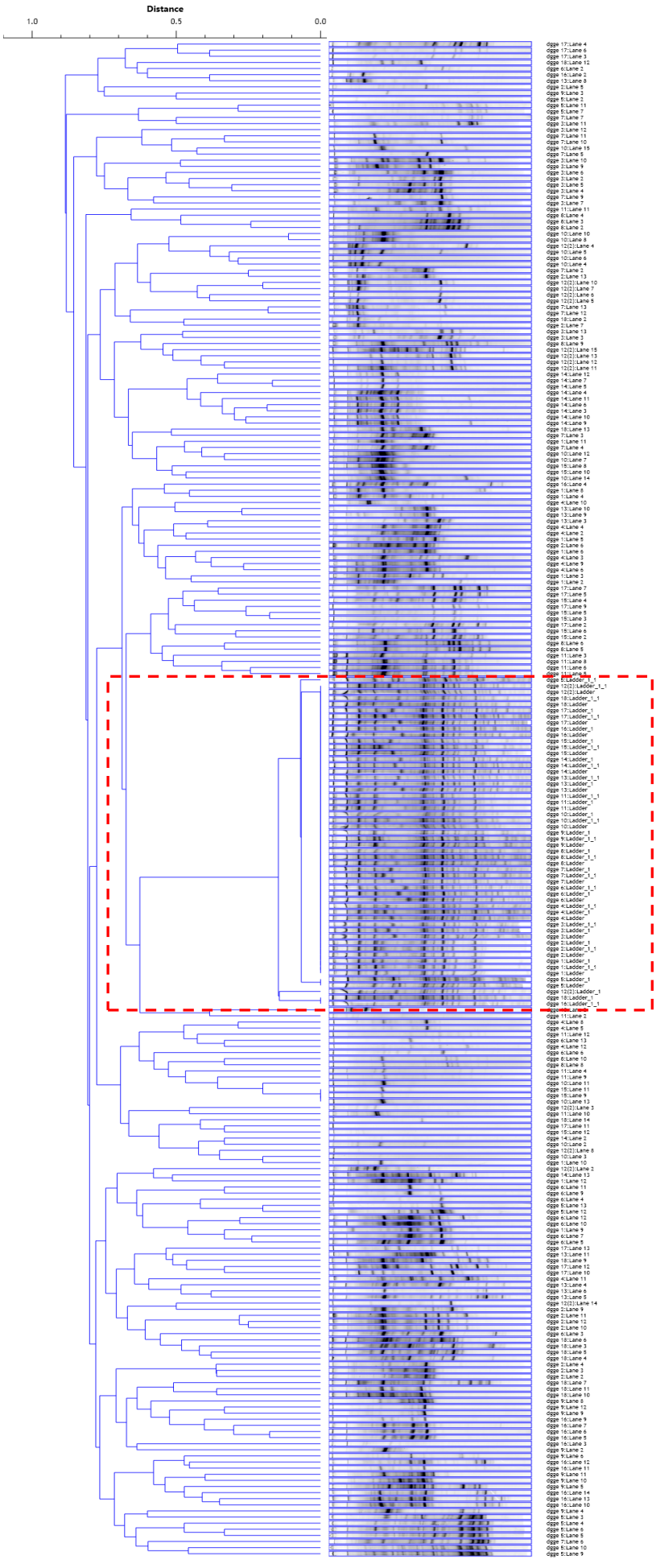
Antibiotic selection media

Proceed as described above to make up the basic recipe (omit agar for broths). When the media has cooled to 50 °C add Ampicillin (100 μg/mL). For plates pour as described above and for broths dispense 5mL aliquots into sterile glass universals.

JM109 LB plates (LB/ampicillin/IPTG/X-Gal)

Once the media from the basic recipe has cooled to 50 °C add Ampicillin (100 μg/mL), IPTG (0.5 mM), and X-Gal (80 μg/mL) to the media and pour as described above.

Appendix 6 - Dendrogram (DICE coefficient) to confirm clustering of ladder lanes


Distance
1.0
0.5
0.0

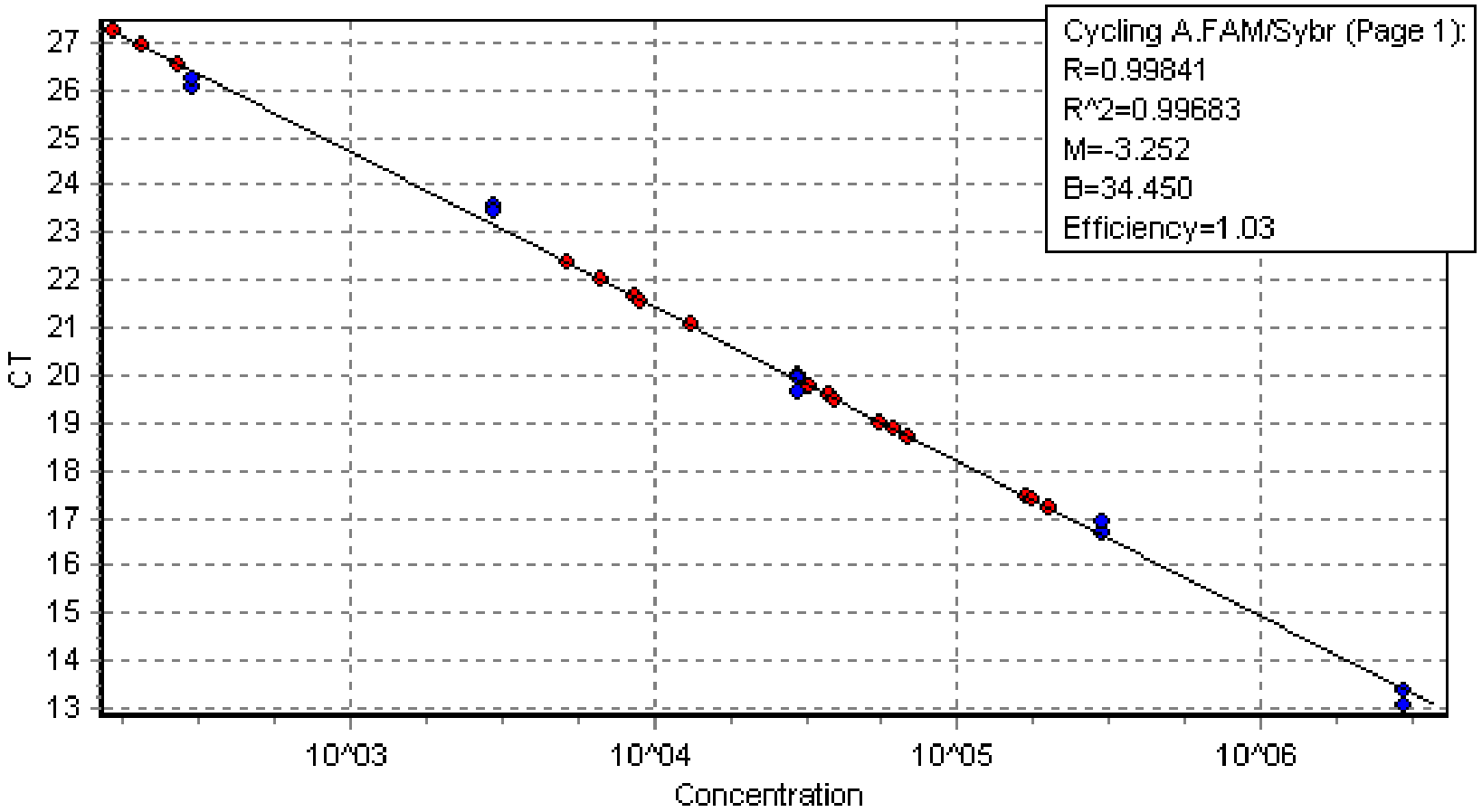
Appendix 7 – Sequence identities of excised DGGE bands from chapters 3 - 5

Table showing the DGGE band sequence identities for chapters 3-5

	Rf Value	Closest Match	Max Identity
Most abundant[a]	0.022	Enterococcus faecalis	98%
	0.036	Enterococcus faecalis	100%
	0.089	Enterococcus faecalis	100%
	0.17	Propionibacterium acnes	100%
	0.267	Streptococcus salivarius	99%
	0.325	Enterococcus faecalis	100%
	0.391	Streptococcus mutans	98%
	0.48	Escherichia coli	100%
	0.501	Escherichia coli	100%
NEC Only[b]	0.102	Staphylococcus epidermidis	100%
	0.136	Staphylococcus epidermidis	99%
	0.18	Staphylococcus epidermidis	100%
	0.19	Enterobacter cloacae	98%
	0.298	Propionibacterium acnes	100%
	0.348	Bacteroides fragilis	98%
	0.374	Bifidobacterium longum	96%
	0.404	Streptococcus mutans	99%
	0.456	Sphingomonas aromaticivorans	95%
	0.469	Enterobacter ludwigii	99%
	0.549	Enterobacter cloacae	85%
	0.647	Methylobacterium populi	100%
NEC and sepsis[c]	0.027	Flavobacteria symbiont	99%
	0.149	Streptococcus salivarius	100%
	0.204	Enterococcus faecalis	98%
	0.514	Veillonella atypica	98%
	0.607	Enterococcus faecalis	100%

[a]Bands present in the 25% most abundant (not already sequenced due to involvement in NEC); [b]Bands associated with patients diagnosed with NEC; [c]bands associated with patients diagnosed with both NEC and sepsis

Appendix 8 – Comparison of culture dependent and molecular techniques in elucidating the gut microbiota of preterm infants



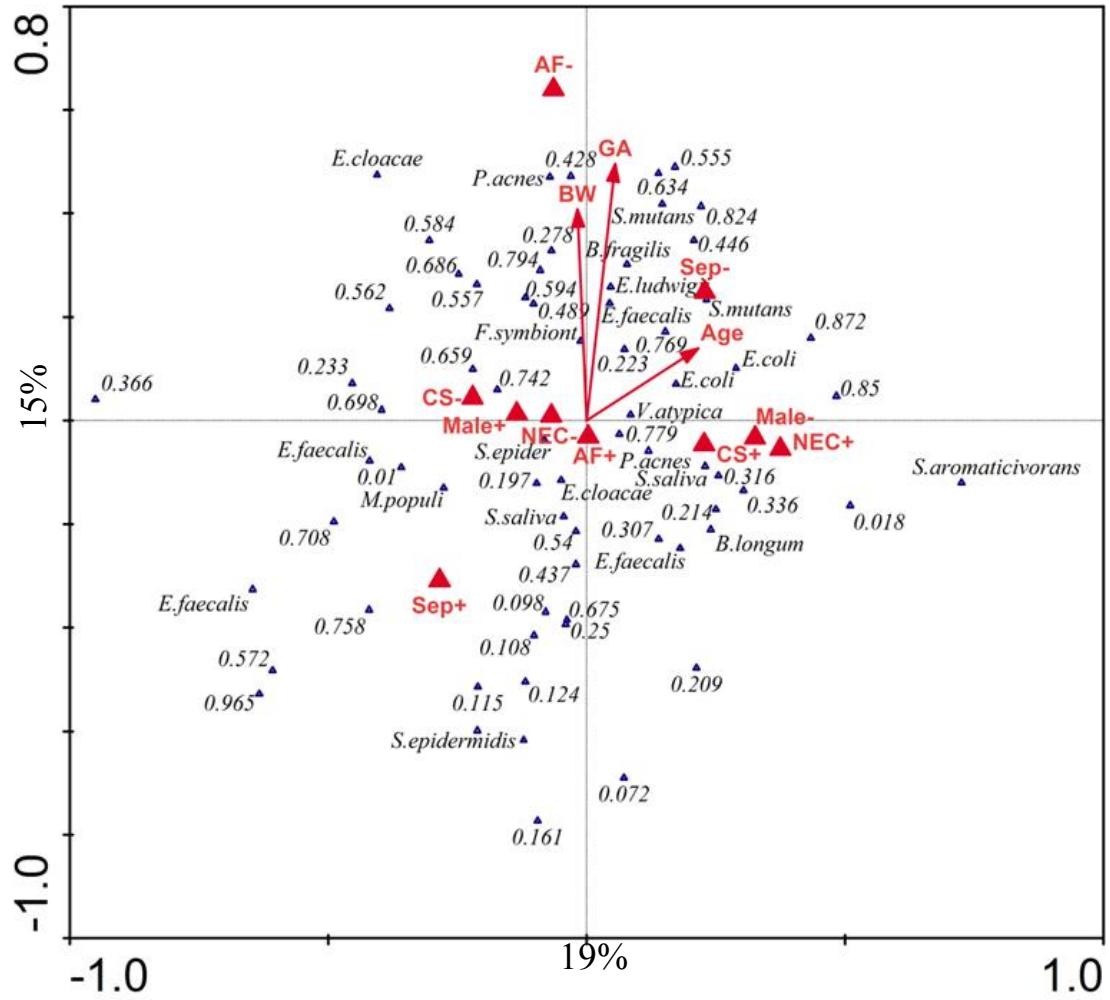
Example of qPCR standard curve

Appendix 9 – Bacterial and fungal viability in the preterm gut: necrotising enterocolitis and sepsis

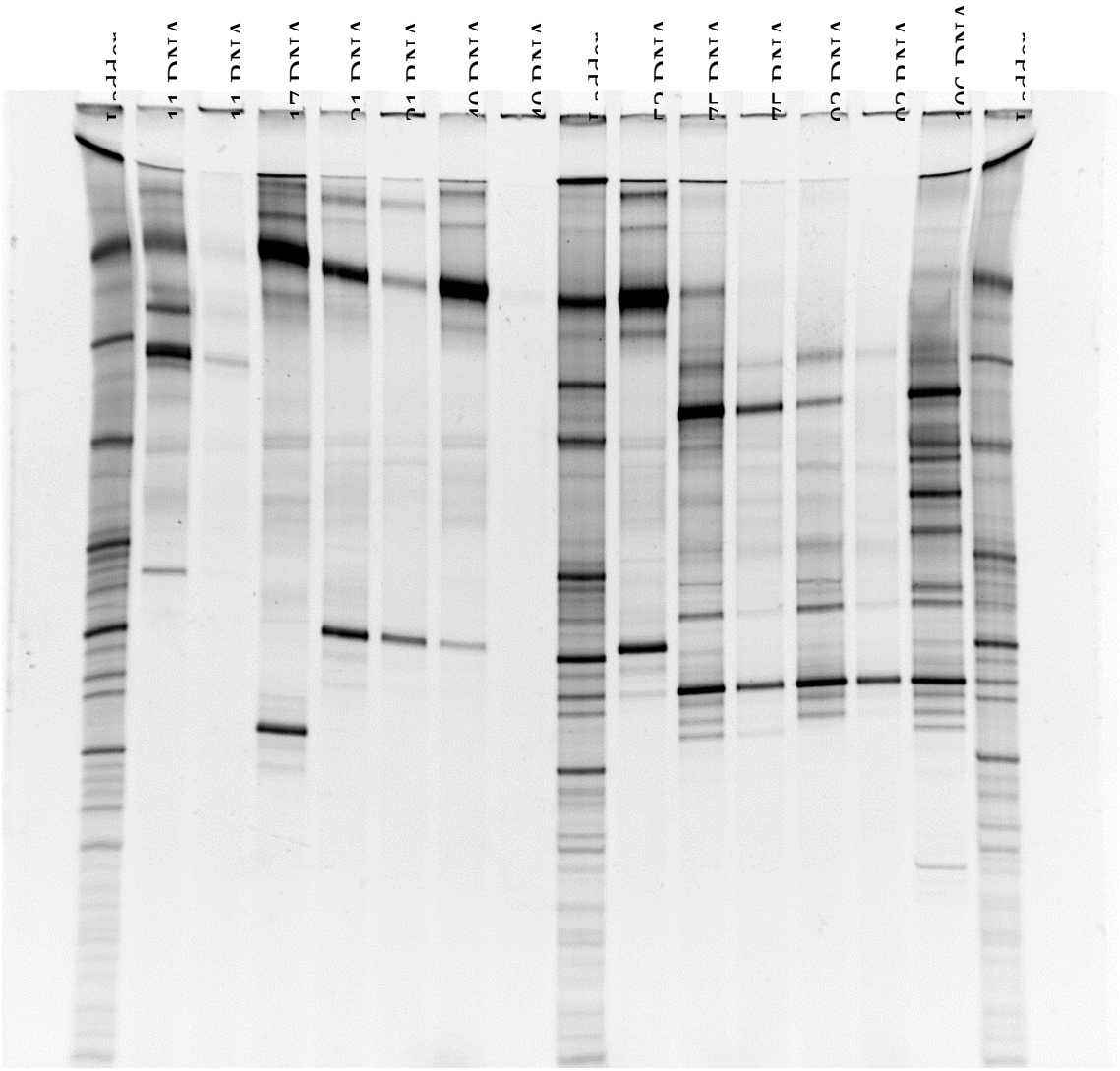
Table showing disease state and antibiotic regime for each individual infant

Patient	Disease State	Antibiotic Course																											
		1			2			3			4			5			6			7			8			9			10		
		DOL	Abx	Dur	DOL	Abx	Dur	DOL	Abx	Dur	DOL	Abx	Dur	DOL	Abx	Dur	DOL	Abx	Dur	DOL	Abx	Dur	DOL	Abx	Dur	DOL	Abx	Dur	DOL	Abx	Dur
1	Healthy	0	P,G	2	4	A,G,F	3	25	A,G,F	2																				
10	Sepsis	0	P,G	2	0	F	6	10	V,C,M	6	22	C,V	3	22	Mr,V	14	57	A,F,G	2											
16	Healthy	0	A,G,M	5	5	V	3	16	A,G	5	42	A,G	2																	
23	Healthy	0	P,G	2	8	V,C	2	11	V,C,M	5	17	V,C,M	2																	
26	Healthy	0	P,G	2																										
28	Healthy	0	P,G	2	16	A,F,G	2	18	C,V	4	18	M	7	22	L,Mr	4	34	V,C	5											
34	Sepsis	0	P,G	5	7	A,G,M	4	18	A,F,G	5	21	V,C	7	78	F,A,G	3	102	F,A,G	3	140	F	3								
39	Sepsis	0	P,G	2	13	F	14	15	A,G	2	17	V	3	24	V	5	33	V,C	3	49	A,G,F	3								
40	Healthy	0	P,G	2	67	A,F,G	2																							
41	Sepsis	0	P,G	2	5	V	1	15	A,F,G	3	31	V,C	5	42	A,G,F	5	50	A,G,F	2	61	A,G,F	3								
42	Sepsis	0	P,G	2	17	M,V,Cx	7	17	F,G	6	22	Ap, Mr	18	22	V	14														
43	Healthy	0	A,G,M	2	18	A,G,F	3																							
44	Healthy	0	A,G,M	2																										
45	Healthy	0	P,G	2																										
46	Healthy	0	P,G	2	3	A,F	2	8	A,F,G	3	17	C	7	17	V	3	30	Mr	5											
47	NEC	0	P,G	2	4	A,F	2	5	V	9	9	C	24																	
48	Healthy	0	A,G,M	10	23	A,F,G	2	43	A,G	3	43	F	5																	
49	Healthy	0	P,G	2	20	A,F,G	2																							
50	Healthy	0	P,G	2	3	A,F,M	5	13	V,C,M	5																				
51	NEC	0	P,G	2	16	V,C,M	7	15	V	3	40	A	3	40	F,G	7	56	Mr,G	6	92	G,A,F	10								
52	Healthy	0	P,G	2																										
54	NEC	0	A,G,M	5	11	C,V	2	13	M,C	20	35	V	10																	
55	NEC	0	P,G	2	3	A,F,M,G	4	10	V,C	10	25	A,V,G	5																	
60	Sepsis	0	P,G	2	3	G,A,F	3	5	V	7	52	Cx,F,M	8																	
61	Healthy	0	P,G	2	3	V	2	4	C	1	24	A,G,M	3	40	A,F,G	3	67	F	11											
63	NEC	0	P,G	5	5	Cx,V,M	7	28	V	10	28	C	2																	
65	Sepsis	0	P,G	2	9	A,F,G	3	11	V,C	3	13	Mr	7	31	Mr	4	43	V,C	7	58	C,V,M	6								
67	Healthy	0	P,G	2	9	A,F,G,V	2	11	C	2	11	M	6	12	Mr	5	33	V	5	67	G,T	5	98	F,A,G	3					
68	Healthy	0	P,G	2	3	V	5	3	C	2	5	Mr	3	16	V,C	2	22	A,F,G	3											
70	Sepsis	0	P,G	2	3	F	3	4	A,G	1	5	Mr,V	5	16	F	8	33	A,G	2	33	F	4	41	A,G,F	3					
78	NEC	0	P,G	2	3	Cx	18	5	M,G	14																				
81	NEC	0	P,G	2	5	F,G	5	17	M	16	17	C	4	17	V	7	20	Mr	7	24	G	10	68	M,V,C	3	79	A,G,F	5	84	C,V	5

Abbreviations: DOL = day of life, Abx = antibiotic, Dur = duration (days). Antibiotics include A = Amoxicillin, Ap = Amphotericin, C = Ceftazadime, Cx = Cefotaxime, F = Flucloxacillin, G = Gentamicin, L = Linezolid, M = Metronidazole, Mr = Meropenem, P = Penicillin, T = Tazocin, V = Vancomycin.



CCA of the DNA profiles based on the same subset of samples that underwent RNA analysis



Example DGGE gel demonstrating the DNA and RNA profiles of samples from patient 42

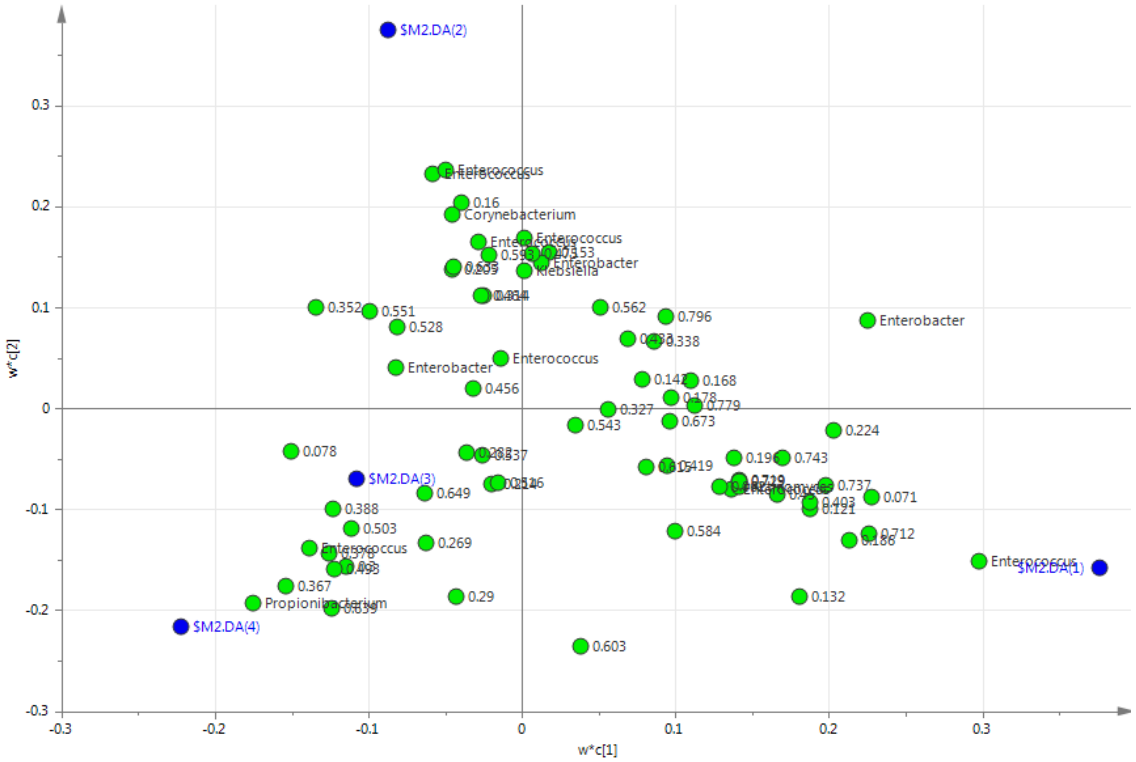
Appendix 10 – Development of the preterm gut microbiota in twins at risk of necrotising enterocolitis and sepsis

See attached disc in rear of thesis for:

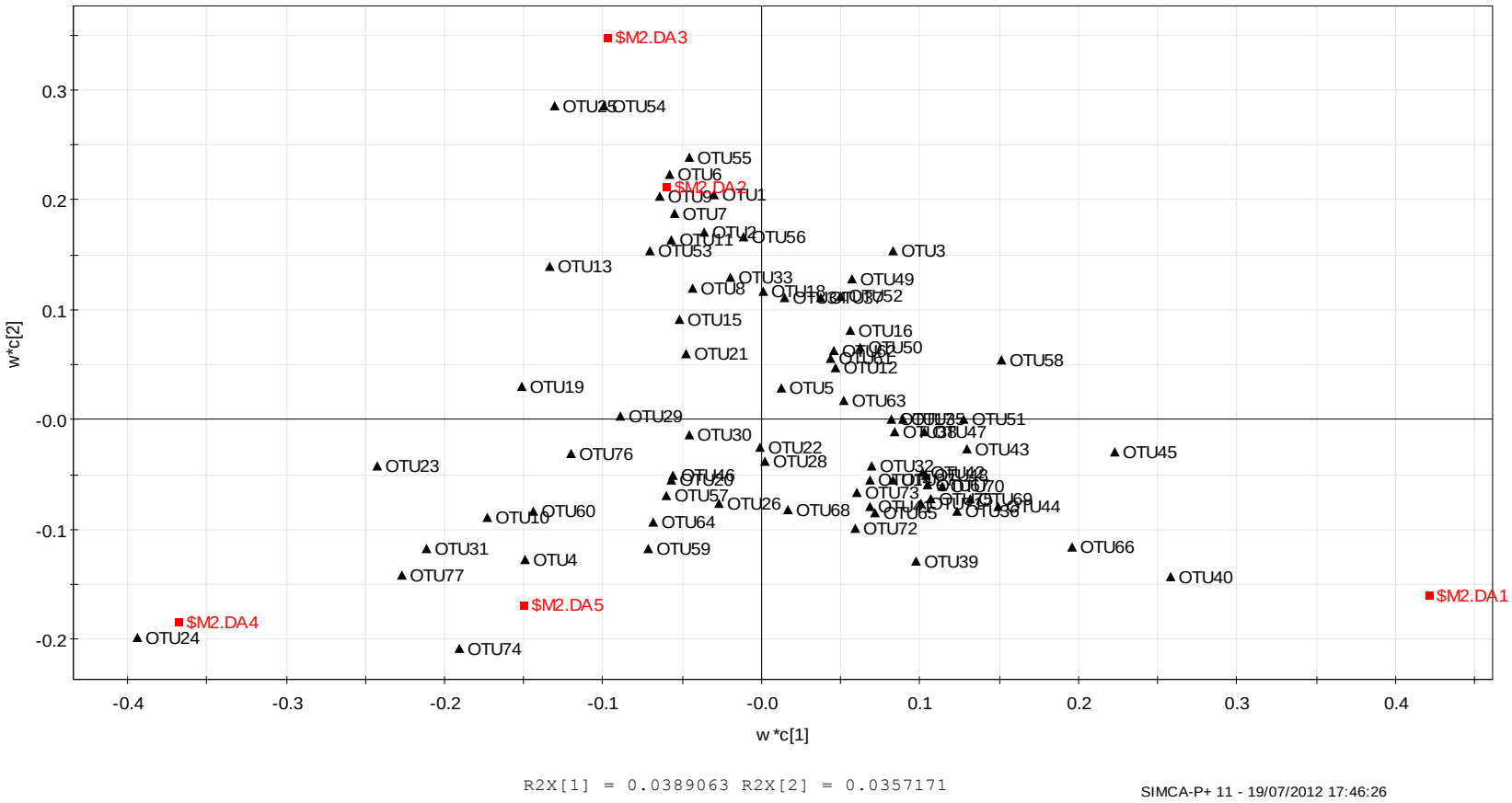
Table showing the full demographic information inclusive of every sample in the twin study

Loadings plot generated in SIMCA based on the DGGE stool data to accompany

Figure 1. Grouping based on sets of multiples.



S5 - Loadings plot generated in SIMCA based on the DGGE expressed breast milk data matched to respective stool to accompany Figure 1. Grouping based on sets of multiples.



Loadings plot generated in SIMCA to identify which bands were associated with late onset infection. Grouped by disease state; 1 = control, 2 = Pre NEC, 3 = Pre sepsis, 4 = Post sepsis, 5 = Post NEC.

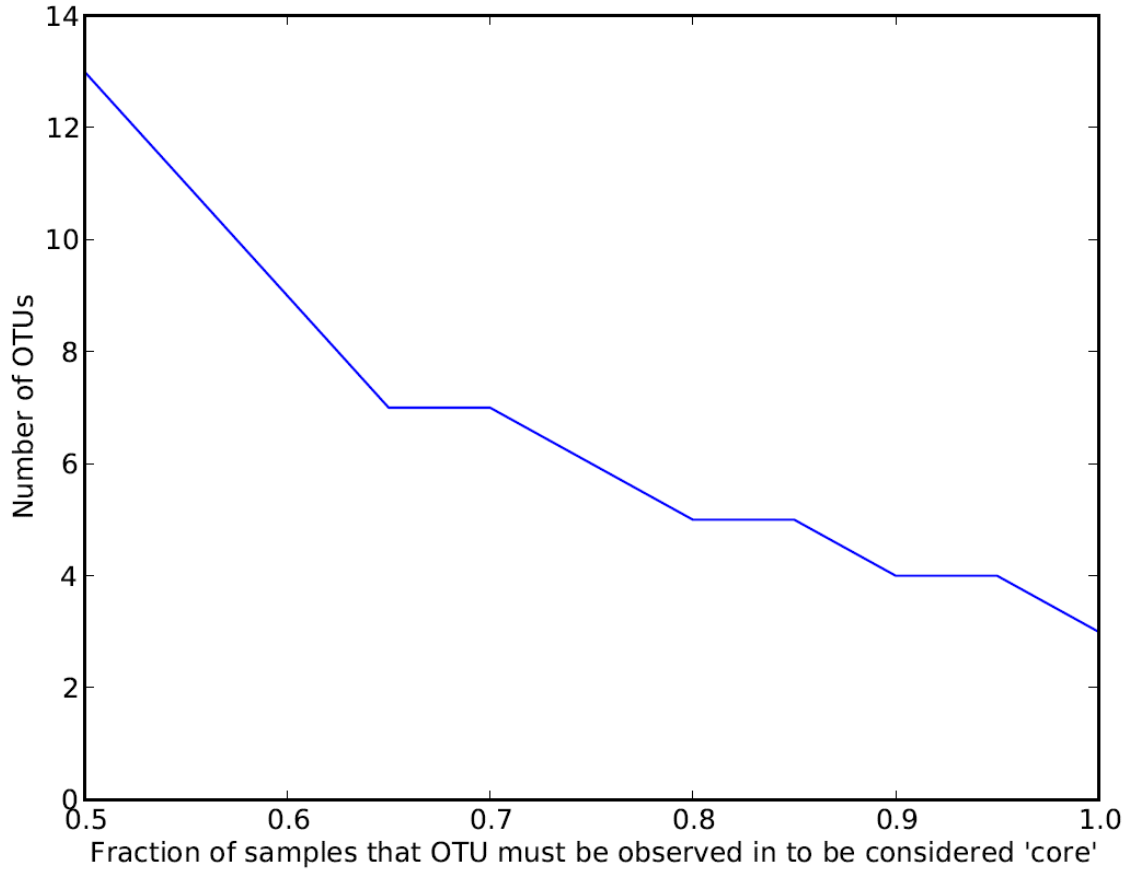
Table showing DGGE band BLASTn sequence identities

	OTU	Closest Match	Accession Number
Pre NEC	24	*Enterococcus*	JX304745.1
Post NEC	23	*Enterococcus*	JX304745.1
	10	*Corynebacterium*	DQ778040.1
	77	*Corynebacterium*	GQ260084.1
	60	*Enterococcus*	HE979846.1
	74	*Enterococcus*	JX304745.1
	4	*Enterobacter*	JN886722.1
Pre Sepsis	25	*Propionibacterium*	JX262688.1
	54	*Actinomyces*	AJ243894.1
Post Sepsis	55	*Klebsiella*	DQ303436.1
	6	*Enterobacter*	JX847659.1
	9	*Enterobacter*	JX847659.1
	1	*Enterococcus*	JX304745.1
	7	*Enterococcus*	JX304745.1
Control	40	*Enterobacter*	FR773881.1
	45	*Enterococcus*	JX304745.1
	66	*Enterococcus*	JX304745.1

Appendix 11 – Next generation sequencing of the gut microbiota in preterm neonates: A case-control study

See attached disc in rear of thesis for:

Table showing the full demographic and MiSeq data summary for each sample



Number of OTUs present in all samples from 50% - 100%.

All significantly different OTUs between EP and VP group determined by MetaStats

Taxonomy	P value
Firmicutes; Bacilli; Lactobacillales; Streptococcaceae; Streptococcus	<0.001
Proteobacteria; Betaproteobacteria; Burkholderiales; Sutterellaceae; Sutterella	<0.001
Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; unclassified	<0.001
Firmicutes; Clostridia; Clostridiales; Ruminococcaceae; unclassified	<0.001
Tenericutes; Mollicutes; Mycoplasmatales; Mycoplasmataceae; Ureaplasma	<0.001
Actinobacteria; Actinobacteria; Bifidobacteriales; Bifidobacteriaceae; Gardnerella	<0.001
Actinobacteria; Actinobacteria; Bifidobacteriales; Bifidobacteriaceae; Bifidobacterium	<0.001
Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; unclassified	<0.001
Firmicutes; Bacilli; Lactobacillales; Lactobacillaceae; Lactobacillus	<0.001
Actinobacteria; Actinobacteria; Actinomycetales; Corynebacteriaceae; Corynebacterium	<0.001
Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae; Megasphaera	<0.001
Firmicutes; Bacilli; Lactobacillales; Lactobacillaceae; Lactobacillus	<0.001
Firmicutes; Clostridia; Clostridiales; Eubacteriaceae; Anaerofustis	<0.001
Actinobacteria; Actinobacteria; Bifidobacteriales; Bifidobacteriaceae; Bifidobacterium	<0.001
Actinobacteria; Actinobacteria; Actinomycetales; Actinomycetaceae; Actinomyces	<0.001
Firmicutes; Bacilli; Lactobacillales; Carnobacteriaceae; unclassified	<0.001
Firmicutes; Bacilli; unclassified; unclassified; unclassified	<0.001
Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas	<0.001
Bacteroidetes; Bacteroidia; Bacteroidales; Prevotellaceae; Prevotella	<0.001
Bacteroidetes; Bacteroidia; Bacteroidales; Prevotellaceae; Prevotella	<0.001
Actinobacteria; Actinobacteria; Actinomycetales; Micrococcaceae; Rothia	<0.001
Firmicutes; Bacilli; Lactobacillales; Leuconostocaceae; Leuconostoc	<0.001
Firmicutes; Bacilli; Lactobacillales; Lactobacillaceae; Lactobacillus	<0.001
Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Rhizobium	<0.001
Actinobacteria; Actinobacteria; Bifidobacteriales; Bifidobacteriaceae; Alloscardovia	<0.001
Firmicutes; Bacilli; Bacillales; Staphylococcaceae; unclassified	<0.001
Actinobacteria; Actinobacteria; Actinomycetales; Actinomycetaceae; Actinomyces	<0.001
Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae; Veillonella	<0.001
Proteobacteria; Epsilonproteobacteria; Campylobacterales; Campylobacteraceae; Campylobacter	<0.001
Fusobacteria; Fusobacteria; Fusobacteriales; Leptotrichiaceae; Leptotrichia	<0.001
Bacteroidetes; Bacteroidia; Bacteroidales; Prevotellaceae; Prevotella	<0.001
Firmicutes; Bacilli; Lactobacillales; Enterococcaceae; unclassified	<0.001
Firmicutes; Clostridia; Clostridiales; Clostridiales_Incertae_Sedis_XI; Peptoniphilus	<0.001
Firmicutes; Bacilli; Lactobacillales; Carnobacteriaceae; Alloiococcus	<0.001
Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; unclassified	<0.001
Actinobacteria; Actinobacteria; Bifidobacteriales; Bifidobacteriaceae; Bifidobacterium	<0.001
Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; unclassified	<0.001
Proteobacteria; Alphaproteobacteria; unclassified; unclassified; unclassified	<0.001
Firmicutes; Bacilli; Bacillales; unclassified; unclassified	<0.001
Firmicutes; Bacilli; unclassified; unclassified; unclassified	<0.001
Proteobacteria; Alphaproteobacteria; Rhizobiales; Methylobacteriaceae; Methylobacterium	<0.001
Proteobacteria; Alphaproteobacteria; Rhizobiales; Phyllobacteriaceae; Phyllobacterium	<0.001

Actinobacteria; Actinobacteria; Actinomycetales; unclassified; unclassified	<0.001
Actinobacteria; Actinobacteria; Actinomycetales; Corynebacteriaceae; Corynebacterium	<0.001
Firmicutes; Clostridia; Clostridiales; Clostridiales_Incertae_Sedis_XI; Anaerococcus	<0.001
Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; Comamonas	<0.001
Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae; Sphingomonas	<0.001
Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides	<0.001
Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; unclassified	<0.001
Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia	<0.001
Actinobacteria; Actinobacteria; Coriobacteriales; Coriobacteriaceae; Slackia	<0.001
Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; Dorea	<0.001
Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; unclassified	<0.001
Proteobacteria; Gammaproteobacteria; Pseudomonadales; Moraxellaceae; Acinetobacter	<0.001
Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; Clostridium_XlVa	<0.001
Firmicutes; Bacilli; Lactobacillales; Streptococcaceae; Streptococcus	<0.001
Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides	<0.001
Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides	<0.001
Actinobacteria; Actinobacteria; Actinomycetales; Propionibacteriaceae; Propionibacterium	<0.001
Firmicutes; Bacilli; unclassified; unclassified; unclassified	<0.001
Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Tatumella	<0.001
Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae; Dialister	<0.001
Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; Delftia	<0.001
Firmicutes; Clostridia; Clostridiales; Clostridiales_Incertae_Sedis_XI; Anaerococcus	0.001
Firmicutes; Bacilli; Lactobacillales; unclassified; unclassified	0.001
Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae; Paracoccus	0.001
Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; Acidovorax	0.001
Firmicutes; Bacilli; unclassified; unclassified; unclassified	0.001
Actinobacteria; Actinobacteria; Actinomycetales; Dermacoccaceae; Dermacoccus	0.001
Firmicutes; Bacilli; Lactobacillales; Enterococcaceae; Enterococcus	0.001
Actinobacteria; Actinobacteria; Bifidobacteriales; Bifidobacteriaceae; Bifidobacterium	0.001
Firmicutes; Bacilli; Lactobacillales; Streptococcaceae; Streptococcus	0.001
Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Morganella	0.001
Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas	0.001
Actinobacteria; Actinobacteria; Actinomycetales; unclassified; unclassified	0.001
Firmicutes; Bacilli; Lactobacillales; Streptococcaceae; Streptococcus	0.001
Actinobacteria; Actinobacteria; Actinomycetales; Actinomycetaceae; Actinomyces	0.001
Actinobacteria; Actinobacteria; Actinomycetales; Corynebacteriaceae; Corynebacterium	0.001
Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae; Negativicoccus	0.001
Bacteroidetes; Bacteroidia; Bacteroidales; Porphyromonadaceae; Parabacteroides	0.001
Proteobacteria; Gammaproteobacteria; Aeromonadales; Aeromonadaceae; Aeromonas	0.001
Firmicutes; Clostridia; Clostridiales; Clostridiaceae_1; Clostridium_sensu_stricto	0.001
Firmicutes; Clostridia; Clostridiales; Clostridiaceae_1; Clostridium_sensu_stricto	0.001
Fusobacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae; Fusobacterium	0.001
Firmicutes; Bacilli; Lactobacillales; Lactobacillaceae; Lactobacillus	0.001
Actinobacteria; Actinobacteria; Actinomycetales; Corynebacteriaceae; Corynebacterium	0.001

Proteobacteria; Deltaproteobacteria; Desulfovibrionales; Desulfovibrionaceae; Bilophila	0.001
Actinobacteria; Actinobacteria; Actinomycetales; Actinomycetaceae; Actinobaculum	0.001
Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides	0.001
Actinobacteria; Actinobacteria; Coriobacteriales; Coriobacteriaceae; Olsenella	0.001
Firmicutes; Bacilli; Lactobacillales; Aerococcaceae; Facklamia	0.001
Firmicutes; Clostridia; Clostridiales; unclassified; unclassified	0.001
Firmicutes; Bacilli; Lactobacillales; Streptococcaceae; Streptococcus	0.001
Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides	0.001
Actinobacteria; Actinobacteria; Coriobacteriales; Coriobacteriaceae; Atopobium	0.001
Firmicutes; Bacilli; Lactobacillales; Aerococcaceae; Globicatella	0.001
Proteobacteria; Gammaproteobacteria; unclassified; unclassified; unclassified	0.001
Firmicutes; Clostridia; Clostridiales; Clostridiales_Incertae_Sedis_XI; Anaerococcus	0.002
Proteobacteria; Alphaproteobacteria; Caulobacterales; Caulobacteraceae; Caulobacter	0.002
Proteobacteria; Gammaproteobacteria; unclassified; unclassified; unclassified	0.002
Proteobacteria; Betaproteobacteria; Rhodocyclales; Rhodocyclaceae; unclassified	0.002
Firmicutes; Bacilli; Lactobacillales; Lactobacillaceae; Lactobacillus	0.002
Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae; Veillonella	0.002
Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; unclassified	0.002
Firmicutes; Bacilli; Lactobacillales; unclassified; unclassified	0.002
Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae; Sphingomonas	0.002
Firmicutes; Bacilli; Bacillales; Bacillales_Incertae_Sedis_XI; Gemella	0.003
Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; unclassified	0.004
Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; unclassified	0.004
Proteobacteria; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; unclassified	0.004
Tenericutes; Mollicutes; Mycoplasmatales; Mycoplasmataceae; Mycoplasma	0.004
Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; Cupriavidus	0.004
Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae; unclassified	0.004
Firmicutes; Clostridia; Clostridiales; Clostridiales_Incertae_Sedis_XI; Anaerococcus	0.004
Firmicutes; Clostridia; Clostridiales; unclassified; unclassified	0.004
Actinobacteria; Actinobacteria; Actinomycetales; Actinomycetaceae; Trueperella	0.004
Firmicutes; Clostridia; Clostridiales; Clostridiales_Incertae_Sedis_XI; Anaerococcus	0.004
Actinobacteria; Actinobacteria; Actinomycetales; Actinomycetaceae; Actinomyces	0.005
Firmicutes; Clostridia; Clostridiales; Ruminococcaceae; Butyricicoccus	0.006
Actinobacteria; Actinobacteria; Actinomycetales; Corynebacteriaceae; Corynebacterium	0.007
Actinobacteria; Actinobacteria; Bifidobacteriales; Bifidobacteriaceae; Bifidobacterium	0.007
Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae; Veillonella	0.007
Bacteroidetes; Flavobacteria; Flavobacteriales; Flavobacteriaceae; Chryseobacterium	0.007
Firmicutes; Clostridia; Clostridiales; Peptostreptococcaceae; Clostridium_XI	0.007
Actinobacteria; Actinobacteria; Actinomycetales; Micrococcaceae; Micrococcus	0.008
Actinobacteria; Actinobacteria; Actinomycetales; Actinomycetaceae; Actinomyces	0.008
Actinobacteria; Actinobacteria; Actinomycetales; Nocardiaceae; Rhodococcus	0.008
Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae; Veillonella	0.009
Actinobacteria; Actinobacteria; Actinomycetales; unclassified; unclassified	0.009
Proteobacteria; Gammaproteobacteria; Xanthomonadales; Xanthomonadaceae; Stenotrophomonas	0.010

Firmicutes; Clostridia; Clostridiales; Clostridiales_Incertae_Sedis_XI; Finegoldia	0.011
Firmicutes; Bacilli; Bacillales; Bacillaceae_1; Bacillus	0.011
Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; Ralstonia	0.012
Firmicutes; Bacilli; unclassified; unclassified; unclassified	0.013
Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae; Veillonella	0.013
Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia	0.014
Firmicutes; Bacilli; Bacillales; Staphylococcaceae; Staphylococcus	0.014
Firmicutes; Bacilli; Lactobacillales; Streptococcaceae; Streptococcus	0.015
Actinobacteria; Actinobacteria; Bifidobacteriales; Bifidobacteriaceae; unclassified	0.015
Firmicutes; Bacilli; Lactobacillales; unclassified; unclassified	0.015
Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae; Veillonella	0.015
Actinobacteria; Actinobacteria; Actinomycetales; Corynebacteriaceae; Corynebacterium	0.016
Firmicutes; Clostridia; Clostridiales; Clostridiales_Incertae_Sedis_XI; Anaerococcus	0.016
Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; unclassified	0.016
Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae; Veillonella	0.016
Firmicutes; Clostridia; Clostridiales; Clostridiales_Incertae_Sedis_XIII; Mogibacterium	0.016
Firmicutes; Bacilli; Bacillales; Bacillaceae_2; unclassified	0.016
Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; Clostridium_XlVa	0.018
Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; unclassified	0.019
Proteobacteria; Alphaproteobacteria; Rhizobiales; Brucellaceae; Brucella	0.020
Bacteroidetes; Bacteroidia; Bacteroidales; Prevotellaceae; Prevotella	0.020
Proteobacteria; Gammaproteobacteria; Oceanospirillales; Halomonadaceae; Halomonas	0.023
Firmicutes; Clostridia; Clostridiales; unclassified; unclassified	0.023
Firmicutes; Bacilli; Bacillales; Alicyclobacillaceae; Tumebacillus	0.029
Firmicutes; Bacilli; Lactobacillales; Enterococcaceae; Enterococcus	0.030
Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae; Veillonella	0.030
Proteobacteria; Gammaproteobacteria; unclassified; unclassified; unclassified	0.030
Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae; Veillonella	0.030
Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Morganella	0.030
Firmicutes; Bacilli; unclassified; unclassified; unclassified	0.030
Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; unclassified	0.030
Bacteroidetes; Flavobacteria; Flavobacteriales; Flavobacteriaceae; unclassified	0.030
Actinobacteria; Actinobacteria; Coriobacteriales; Coriobacteriaceae; Atopobium	0.030
Proteobacteria; Gammaproteobacteria; Pseudomonadales; Moraxellaceae; Psychrobacter	0.030
Proteobacteria; Alphaproteobacteria; Rhizobiales; Methylobacteriaceae; Methylobacterium	0.030
Bacteroidetes; Bacteroidia; Bacteroidales; Prevotellaceae; Prevotella	0.030
Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; unclassified	0.031
Firmicutes; Bacilli; Bacillales; unclassified; unclassified	0.031
Fusobacteria; Fusobacteria; Fusobacteriales; Leptotrichiaceae; Leptotrichia	0.031
Bacteroidetes; Flavobacteria; Flavobacteriales; Flavobacteriaceae; Chryseobacterium	0.031
Actinobacteria; Actinobacteria; Coriobacteriales; Coriobacteriaceae; Eggerthella	0.031
Aquificae; Aquificae; Aquificales; Hydrogenothermaceae; Sulfurihydrogenibium	0.031
Firmicutes; Bacilli; Bacillales; Bacillaceae_1; Geobacillus	0.031
unclassified; unclassified; unclassified; unclassified; unclassified	0.035

Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Proteus	0.037
Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides	0.037
Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia	0.037
Bacteroidetes; Sphingobacteria; Sphingobacteriales; Sphingobacteriaceae; Pedobacter	0.039
unclassified; unclassified; unclassified; unclassified; unclassified	0.040
Actinobacteria; Actinobacteria; Coriobacteriales; Coriobacteriaceae; Eggerthella	0.041
Firmicutes; Clostridia; Clostridiales; Clostridiaceae_1; Clostridium_sensu_stricto	0.042
Proteobacteria; Alphaproteobacteria; Rhizobiales; Methylobacteriaceae; Methylobacterium	0.044
Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; Clostridium_XlVa	0.046
Firmicutes; Bacilli; unclassified; unclassified; unclassified	0.047
Firmicutes; Bacilli; Bacillales; unclassified; unclassified	0.049

Publications

1. Stewart CJ, Marrs ECL, Magorrian S, Nelson A, Lanyon C, Perry JD, Embleton ND, Cummings SP, Berrington JE. (2013) Development of the Preterm Gut Microbiome in Twins at Risk of Necrotising Enterocolitis and Sepsis. *PLoS One*. 8(8): e73465
2. Berrington JE, Hearn RI, Hall C, Stewart CJ, Cummings SP, Embleton ND. (2013) Proportionate reduction in uncertainty of late onset infection in preterm infants by neutrophil CD64 measurement. *Fetal and Pediatric Pathology*. In press.
3. Embleton ND, Berrington JE, McGuire W, Stewart CJ, Cummings SP. (2013) Lactoferrin: Antimicrobial activity and therapeutic potential. *Semin Fetal Neonatal Med*. Online ahead of print: doi: 10.1016/j.siny.2013.02.001
4. Stewart CJ, Nelson A, Scribbins D, Marrs ECL, Lanyon C, Perry JD, Embleton ND, Cummings SP, Berrington JE. (2012) Bacterial and fungal viability in the preterm gut: NEC and sepsis. *Arch Dis Child Fetal Neonatal Ed*. 98:F298-F303
5. Berrington JE, Stewart CJ, Embleton ND, Cummings SP. (2012) Gut microbiota in preterm infants: assessment and relevance to health and disease. *Arch Dis Child Fetal Neonatal Ed*. 98(4):F286-90
6. Stewart CJ, Marrs ECL, Magorrian S, Nelson A, Lanyon C, Perry JD, Embleton ND, Cummings SP, Berrington JE. (2012) The preterm gut microbiota: changes associated with necrotising enterocolitis and infection. *Acta paediatrica.* 101(11):1121-7