Mapping the Gut Microbiome to Preterm Neonatal Outcomes

Thesis Presented for the Degree of Doctor of Philosophy

By

Emma Margaret Mitchell, MSc Molecular Genetics

In Candidature for the Degree of Doctor of Philosophy

Infection and Inflammation Division

School of Medicine

Cardiff University

September 2019

Word Count 65,963

Summary

The development of the gut microbiome in preterm infants can have a substantial impact on health, such as the development of the common preterm disease Necrotizing Enterocolitis (NEC). This thesis aimed to identify further areas in which the gut microbiome could be contributing to the development of disease in preterm infants.

Experimental methodology included, 16S rRNA gene metataxonomics, to map the preterm gut microbiome. In addition, protease activity and inhibition assays were implemented to assess total faecal protease activity and identify families of proteases present. Moreover, ELISAs were used to investigate inflammatory content of preterm infant stool. Finally, data from a project, by Dr David Gallacher, into the lung microbiome of preterm infants was analysed with the data from this project to establish links between the development of the gut and lung microbiomes of preterm infants.

The results of this thesis found that the preterm gut microbiome shifts from a *Firmicute* dominated community to a *Proteobacteria* one, during the first 30 days of life. In addition, associations between gender, mode of delivery, antibiotics and sampling site were found. Secondly, no significant changes in protease activity were found over time, however, protease activity during the first 30 days of life varied between individuals. Thirdly, no inflammatory response was detected in the stool of preterm infants. Finally, no significant associations between the bacterial communities of the gut and the lung of preterm infants.

In conclusion, novel findings of this thesis have shown that gender, antibiotics and sampling site have a significant effect of the development of the gut microbiome during the first 30 days of life. Moreover, protease and inflammatory activity of preterm infant stool was not significant. Lastly, development of the gut and lung microbiomes of preterm ventilated infants progress along very different courses.

Statements and Declaration

STATEMENT 1
This thesis is being submitted in partial fulfilment of the requirements for the degree of … (*insert PhD, MD, MPhil, etc., as appropriate*)

Signed _____

Date _____

STATEMENT 2
This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is it being submitted concurrently for any other degree or award (outside of any formal collaboration agreement between the University and a partner organisation)

Signed _____

Date _____

STATEMENT 3
I hereby give consent for my thesis, if accepted, to be available in the University's Open Access repository (or, where approved, to be available in the University's library and for inter-library loan), and for the title and summary to be made available to outside organisations, subject to the expiry of a University-approved bar on access if applicable.

Signed _____

Date _____

DECLARATION
This thesis is the result of my own independent work, except where otherwise stated, and the views expressed are my own. Other sources are acknowledged by explicit references. The thesis has not been edited by a third party beyond what is permitted by Cardiff University's Use of Third Party Editors by Research Degree Students Procedure.

Signed _____

Date _____

WORD COUNT _____

(Excluding summary, acknowledgements, declarations, contents pages, appendices, tables, diagrams and figures, references, bibliography, footnotes and endnotes)

Use of third party editors by research degree students procedure

If you choose to use a third party editor to improve the presentation of your thesis, you must do so in accordance with the University's Policy on the Submission and Presentation of Research Degree Theses (Use of Third Party Editors) and the Procedure outlined below.

You should provide the third party with a copy of the following instructions before they undertake any work.

a The third party should provide comments or amendments via track changes and the research degree candidate should choose whether to accept or reject these changes in a master copy.

b A third party editor cannot be used**:**

- to change the text of the thesis so as to clarify and/or develop the ideas and arguments
- to reduce the length of the thesis so it falls within the specified word limit
- to correct information within the thesis
- to change ideas and arguments put forward within the thesis
- to translate the text into English/Welsh.

c A third party editor can be used to offer advice on:

- spelling and punctuation
- formatting and sorting of footnotes and endnotes for consistency and order
- ensuring the thesis follows the conventions of grammar and syntax in written English/Welsh
- the breaking up of overly long sentences and/or paragraphs
- changing active and personal usages into passive and impersonal, as may be appropriate
- improving the positioning of tables and illustrations and the grammar, spelling and punctuation of any text in or under tables and illustrations
- ensuring consistency of page numbers, headers and footers.

The declaration that is required for inclusion in the thesis when it is submitted for examination includes reference to this Procedure.

DECLARATION
This thesis is the result of my own independent work, except where otherwise stated, and the views expressed are my own. Other sources are acknowledged by explicit references. The thesis has not been edited by a third party beyond what is permitted by Cardiff University's Use of Third Party Editors by Research Degree Students Procedure.

Signed _____
Date _____

Table of Contents

Acknowledgements

Firstly, I would like to thank my supervisor Prof. Sailesh Kotecha for his constant support and encouragement throughout my PhD. I'd also like to thank my co-supervisor Prof. Julian Marchesi for his knowledge on all things Microbiome.

Secondly, there have been several colleagues who have made studying for a PhD a once in a lifetime experience. I'd like to thank Dr Ann Smith, Dr Laura Osborne, Dr Jane Mikhail, Alexandra Campbell, Kyle Murphey, Dr Guilia Masetti and Duaa Al Dulaimy from the MMI Group. Moreover, I have had the pleasure of working with two groups of amazing people, the second of which I'd like to thank are from the Child Health Department at Cardiff University. Thank you, Dr Sarah Kotecha, Dr John Watkins, Dr John Lowe, Kylie Hart, Dr Mike Cousins, Gillian Willets and Louise Yendle, for the countless cups of coffee and friendship. In particular, this project would not have been possible without the efforts of Dr David Gallacher, so I'd like to take the opportunity to say thank you for all of your help and support over the past 4 years.

Of course, I cannot forget to thank the love of my life, Robert Field, for his unyielding love and support. If it were not for this PhD, we would never have met. You were there for me at the best and worst times, and I couldn't have done this without you. You always made me feel like I could accomplish anything, even when I didn't believe you. I know we have an amazing future ahead of us I can't wait to tackle our next challenge with you.

Finally, I would never live it down if I didn't thank my unfathomably strong, loving and brilliant parents, Diane and David Mitchell. Your support, guidance and love has been infinite throughout my life and I know for certain that I could not have gotten as far in my career without you both. Also, I would like to thank my wider family, some of whom have passed, who's love, and encouragement has given me strength over the last 4 years.

Thank you.

List of Figures and Tables

List of Abbreviations

AEBSF	4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride
AJC	Apical Junction Complex
BAL	Bronchoalveolar Lavage
CD	Crohn's Disease
COPD	Chronic Obstructive Pulmonary Disease
CS	Caesarean Section
DNA	Deoxyribosenucleic Acid
ECM	Extracellular Matrix
EHEC	Enterohemorrhagic *E. coli*
ELBW	Extremely Low Birth Weight
EOS	Early Onset Sepsis
FISH	Fluorescence *in situ* Hybridization
GI	Gastrointestinal
HMO	Human Milk Oligosaccharides
HMP	Human Microbiome Project
HTRA	High Temperature Requirement A
IBD	Inflammatory Bowel Disease
IECs	Intestinal Epithelial Cells
IL	Interleukin
LBW	Low Birth Weight
LOS	Late Onset Sepsis
LPS	Lipopolysaccharide
MLC	Myosin Light Chain
MMP	Matrix Metalloprotease

NEC	Necrotizing Enterocolitis
NICU	Neonatal Intensive Care Unit
NMDS	Non Metric Multidimensional Scaling
NPA	Nasopharyngeal Aspirate
OTU	Operational Taxonomic Unit
PAR	Proteinase Activated Receptor
PCoA	Principal Coordinate Analysis
PCR	Polymerase Chain Reaction
RNA	Ribosenucleic Acid
rRNA	Ribosomal Ribosenucleic Acid
S	Svedburg Unit
SOBs	Species Observed
TAF	Tracheal Aspirate Fluid
TIMPs	Tissue Inhibitors of Metalloproteinases
TJ	Tight Junction
UC	Ulcerative Colitis
VLBW	Very Low Birth Weight

Chapter 1

Introduction

Chapter 1. Introduction

1.1 General Introduction

In this section the general concepts of microbiome research will be introduced. Moreover, the most used terms during microbiome research will be defined and clarified. Lastly, the main topics covered in this thesis will also be summarized.

1.1.1 Key Definitions

When referring to disease, research into the bacterial cause of disease has been conducted for hundreds of years. However, during the 20th century, focus has shifted to that of a deeper understanding of the entire non host, predominantly bacterial, content of the human body. Ultimately, once the content is known, links to disease can be established through investigations into differences during and after disease processes, such as infection. As with every area of research, it is given a name that encompasses key aspects of roles within that area. The research noted in this paragraph, into non host organisms on and within the human body, has been termed the **microbiome**.

As with any fledging concept it has been misused and taken out of context by both specialists and general public, alike. This confusion has led to a loss of understanding and miscommunication. A recent editorial by Marchesi and Ravel in 2015 has provided clarification and is a useful resource for terms consistently used throughout this thesis.

Microbiota refers to the collection of microorganisms contained within a defined environment or niche, which in this thesis will be the preterm infant gut. Information regarding the microbiota present in the preterm infant gut will be ascertained via a **metataxonomic** process. This process involves a high throughput sequencing method such as, Illumina Mi-Seq, and subsequent analysis to establish the content and relationships within the samples tested. It will not be investigated as part of this thesis, but the **metagenome** refers to the entire genetic content of the sample area (Marchesi and Ravel 2015).

The term **microbiome** is the most frequently misused in this field, as it is currently used as a blanket term for the whole area of research. Therefore, during this thesis the correct definitions will be used. The **microbiome** encompasses the whole study environment,

the organisms, and their genetic content. Finally, the term **microflora** is often used. However, by definition it means 'small plants', which is an area not linked to microbiome research or this thesis (Marchesi and Ravel 2015).

There are other terms used during microbiome research that should also be included here such as, **richness**, **diversity**, and **evenness**. More commonly these are used to describe environmental habitats. However, they have been translated into the field of microbiome research to describe this unique environmental habitat. Firstly, there are 2 levels of diversity, **alpha-** and **beta-**. **Alpha diversity** refers to the diversity of organisms within a specific environment, such as the gut microbiome. In contrast, **beta diversity** refers to diversity between environments, such as the gut and lung microbiomes (Whittaker 1960). **Alpha diversity** is measured in terms of both **richness** and **evenness**. The term **richness** is defined as the number of species present in an ecosystem and does not consider the number of individuals from that species. Often **richness** is measured using the Chao 1 or Ace indices. In contrast, the term **evenness** takes into account the number of organisms in a species, thereby giving a calculation of equality within a community. Ultimately, **diversity** takes into account of the previous measurements and calculates the number of species present and the number of organisms present in each species. For instance, an environment with high diversity would have a large number of organisms from each species present. Again, **diversity** is often measured using the Simpson's or Shannon's diversity indices. Often diversity measurements include or account for the evenness within a system (Hill *et al.* 2003).

The results of microbiome research utilise taxonomic nomenclature. The most widely used levels of nomenclature used during microbiome research is **phylum** and **genus**. This is due to the limitations with the current power of next generation sequencing that **species** identity cannot be readily identified. For example, the bacterium *Escherichia coli* includes the phylum, *Proteobacteria*, genus, *Escherichia*, and the species, *coli*.

The key focus of this project concerns gut microbiome development in preterm infants. A preterm infant is defined as such if born before 37 weeks' gestation, including the following sub definitions. An **extreme** preterm is an infant delivered at less than 28 weeks' gestation. A **very** preterm infant is one born between 29 to 32 weeks' gestation. Finally, a **late** preterm is an infant born between 33 and 37 weeks' gestation (Goldenberg *et al.* 2008). Often these infants require hospitalization and receive high

levels of care in a specialized ward, the neonatal intensive care unit (NICU). This deviation from normal development is of great interest to the scientific community and has huge consequences for not only the general health of the infant, but the development of their gut microbiome.

1.1.2 Overview of the Topics Included in this Thesis

Research has shown that gut microbiome colonisation begins before birth, with subsequent changes throughout life. Current knowledge suggests that being born prematurely disrupts the normal colonisation of the gut microbiome, leading to acute and chronic disease (Pammi *et al.* 2017). As a result, there have been a number of studies into the acquisition of the gut microbiome in preterm infants, with the results indicating clear differences from their full-term counterparts (Zhou *et al.* 2015). However, additional areas, such as the effect of gender, of gut microbiome development in these infants have yet to be ascertained.

One of these areas is the contribution of proteases to the normal preterm gut microbiome environment. Research, discussed in more detail in Chapter 4, clearly demonstrates an association between proteases and gastrointestinal health in adults. Data from these studies suggests that both the immune system and intestinal epithelial barrier interact with proteases and, as a result, contribute to the pathology of disease (Gecse *et al.* 2008; Shulman *et al.* 2008). Therefore, it is not unreasonable to hypothesise that the same processes are occurring in the gastrointestinal (GI) system of preterm infants.

Currently, there has been little evidence of a detectable inflammatory response in the stool of preterm infants (Rougé *et al.* 2010). As yet, it is undetermined if this is simply a lack of immune response or because it is undetectable in the stool of preterm infants. Therefore, the effect the microorganisms are having on the microbiome in this patient group is unclear. Furthermore, there has been no data produced on the potential proteases in this system.

Lastly, an emerging niche of microbiome research has shown that bacteria and their metabolic products, can influence different areas of the human body. This is achieved through a breakdown of the epithelial barrier and subsequent translocation of

microbiome components into the bloodstream (Reddy *et al.* 2007). Studies have demonstrated a link between the translocation of gut bacteria from the GI tract to other body sites including the lungs, which are then able to exacerbate common diseases as they colonize these new areas (Dickson *et al.* 2016).

1.2 The Microbiome

This section will provide a background on the gut microbiome and its development. In addition, detail will be provided on the methods used to study the microbiome and how they have developed. Finally, how the gut microbiome changes during disease shall be presented.

1.2.1 Introduction

Twenty years ago, the concept of a microbiome was just that, a concept. It was radical thinking to suggest that human body was home to millions of organisms occupying several sites on the human body. Detailed microbiome research was not possible until the launch of the Human Microbiome Project (HMP) in 2007 (Peterson *et al.* 2009). Since then the research conducted has grown exponentially and continues to do so (Marchesi 2011).

From the HMP it has been discovered that the relationship with the microbiome is symbiotic and affects every aspect of human health (Turnbaugh *et al.* 2007). Gnotobiotic, meaning all organisms are absent from an individual, mice have been found to develop severe autism (Desbonnet *et al.* 2014), the scalp microbiome has been linked to excessive dandruff (Xu *et al.* 2016) and diseases such as inflammatory bowel disease (IBD) and Crohn's disease are known, in part, to be caused by an abnormal microbiome (Willing *et al.* 2010; Fujimoto *et al.* 2013).

Stemming from the HMP and other projects it was believed, up until recently, that humans were host to ten times more bacterial cells than constituent hot cells. However, a study by Sender *et al* in 2016, showed this value to be grossly over estimated, with the actual number more resembling a 1:1 ratio (Sender *et al.* 2016).

An interesting characteristic of the microbiome is that it is as unique as a fingerprint (Wilkins *et al.* 2017). Therefore, it is not surprising that research has shown there are

numerous contributing factors such as environment, genetics, diet and health, which influence the microbiome (Turnbaugh *et al.* 2010; Goodrich *et al.* 2014). It is the influence of all these factors, and many others, which constantly shape and change the diversity and interactions in these micro-ecosystems.

As mentioned previously in this section, there are numerous different microbiomes occupying several sites on and within the human body. However, this thesis is primarily focused on the gut and lung microbiomes. As a result, the following are discussed in more detail in Chapter 3 and Chapter 6.

1.2.2 The Gut Microbiome

1.2.2.1 Introduction and Definitions

As discussed in Section 1.1.1, the term microbiome has become ambiguous. Therefore, to avoid confusion throughout this thesis, it is useful to clarify what the gut microbiome is and its key components. Encompassing the definition proposed by Marchesi and Ravel, the gut microbiome comprises of the following components. Firstly, the microorganisms and the products they produce, predominantly this is bacteria, but also includes viruses, fungi, yeasts and other eukaryotes (Reyes *et al.* 2010; Nash *et al.* 2017). Secondly, there are the host contributions which includes, the physical intestinal lumen and the products it produces. Lastly, are the contents of the intestinal lumen, affected by diet, but necessary to sustain the vast quantity of life in this micro-ecosystem (Marchesi 2011; Marchesi and Ravel 2015). This concept is taken forward and expanded upon throughout this thesis.

In addition, it is necessary to define what is meant by the ‘gut’ in gut microbiome. Often the gut microbiome is considered to encompass the entire gastrointestinal tract. However, the oral and stomach microbiomes are very much separate and unique ecosystems compared to each other and the gut microbiome (Dewhirst *et al.* 2010; Klymiuk *et al.* 2017). As with any ecosystem the occupants have evolved to become adapted to a specific environment in order to exploit whatever resource is plentiful there. Therefore, in this thesis the ‘gut’ microbiome is considered to encompass the lower intestinal tract, from the small intestine to the colon, see Table 1 (Marchesi *et al.* 2015).

Table 1. The Components of the Gastrointestinal Tract. The gastrointestinal (GI) tract is split into 3 main sections, upper, mid and lower. Each of these sections contain numerous components that perform a specific function necessary for homeostasis. The gut microbiome, when analysed with a stool sample, gives an impression the environment in only the lower GI tract. This table was constructed using information from (Bailey and Keshav 2012).

Section of GI Tract	Components	Function
Upper	Mouth	Admits food into the gastrointestinal tract and is responsible for the initial breakdown of the food.
	Salivary Glands	Responsible for the lubrication of the mouth. Contains digestive and antimicrobial enzymes.
	Oesophagus	Carries food and liquid from the mouth to the stomach by peristalsis.
Mid	Liver	Performs metabolic, synthetic, secretory and excretory roles vital for life.
	Stomach	Performs the storage, churning and digestion of food. Contains numerous enzymes, such as pepsin, to aid this process.
	Duodenum	First part of the intestine and begins the adsorption and digestive process in the intestine. Bile, pancreatic juice and enteric secretions are added.
	Gall bladder	Part of the host defence system produces bile for the removal of toxins and metabolic waste.
	Pancreas	Gland responsible for producing the digestive enzymes in the intestine. Also produces insulin.
Lower	Jejunum	Part of the main absorptive surface of the gastrointestinal tract. Part of the small intestine.
	Ileum	Part of the main absorptive surface of the gastrointestinal tract. Precedes the caecum. Part of the small intestine.
	Colon	The predominant component of the large intestine. Main function is to reabsorb water.
	Appendix	Follows front the caecum and have no special function in humans.
	Caecum	The most proximal part of the large intestine.
	Rectum	Stores faeces before defecation.
	Anus	The most distal part of the gastrointestinal tract. Controls defecation.

The gut microbiome is known to contribute a vast amount of advantages to their hosts, such as aiding digestion (Cantarel *et al.* 2012), providing vitamins (Gill *et al.* 2006), preventing the overgrowth of pathogens (Momose *et al.* 2008), and regulating host metabolism and immune system (Tremaroli and Bäckhed 2012). All of these functions are necessary to maintain homeostasis, therefore a substantial body of research has been invested to determine if there is common core microbiome shared by all individuals (Turnbaugh and Gordon 2009; Huse *et al.* 2012; Li *et al.* 2013). However, studies have demonstrated the malleable nature of the human gut microbiome (Eckburg *et al.* 2005; Koren *et al.* 2012). For example, due to the seasonal variations in food abundance, a study by Davenport *et al* in 2014 discovered an isolated human population, whose gut microbiomes composition changes in response to seasonal changes (Davenport *et al.* 2014). Current research suggests that adults do not share a core microbiome, due to the influence of environmental factors (Lloyd-Price *et al.* 2016).

However, Turnbaugh *et al* in 2009 found that the gut microbiome is shared within families (Turnbaugh *et al.* 2009). This could be a result of a shared evolution. Studies have shown that the microbiome has evolved in parallel with the human population. For example, in modern history the microbiome has co-evolved to cope with urbanization (Winglee *et al.* 2017), highly processed diets (Mozaffarian *et al.* 2011), global travel (Nordahl Petersen *et al.* 2015), improved hygiene and medicine (Dethlefsen *et al.* 2007). On the other hand, there are disadvantages to the co-evolution of the microbiome and improved healthcare. Research has shown that the commensal microbiota are transferring antibiotic resistance genes to opportunistic pathogens (Salyers *et al.* 2004; Sommer *et al.* 2009), which has the potential to become a huge problem for health care.

As will be discussed in Section 1.2.4, the gut microbiome is acquired before and during birth, changes throughout life, and reduces during old age. One of the most dramatic changes in the transition from an infant microbiome to one that is considered more ‘adult-like’, at around 2 years of age (Adlerberth and Wold 2009). However, this may sound contradictory, as current research suggests there is not a core microbiome. Therefore, the ‘adult-like’ microbiome refers to a change from infant associated taxa to organisms more prevalent in adulthood. As a result of the early acquisition and constant companionship throughout life, the gut microbiome has significant functions in a healthy human host.

1.2.2.2 The Function of the Healthy Gut Microbiome

After detailing the concept, it would be appropriate to discuss the function of the gut microbiome. Studies discussed in Section 1.2.2.1, have shown that the gut microbiome has co-evolved with humans, adapting to our changing lifestyles. As a result, it can be inferred that these organisms have a defined role within their appropriate microbiome, similarly the host must provide several benefits to the microbiota in order for the two to coexist. Furthermore, there is a significant amount of variation in the human gut microbiome between individuals because of the influence of outside factors. Therefore, it is very difficult to define what constitutes a 'healthy' microbiome. Consequently, an alternative hypothesis proposed by Lloyd-Price *et al* is a logical explanation.

In their 2016 article, they propose a 'functional core', defined as:

> *"a complement of metabolic and other molecular functions that are performed by the microbiome within a particular habitat but are not necessarily provided by the same organisms in different people"*
> *(Lloyd-Price et al. 2016)*

Furthermore, they inferred this 'functional core' to include several components. Firstly, there needs to be an element of genetic potential, in other words enough genetic material to maintain a healthy population, and a lack of deleterious mutations. Secondly, there must be a set of house-keeping functions that enable the organisms present in the gut microbiome to fully exploit the environmental niché in which they have colonised. An example of this, is the metabolism of dietary components that would otherwise be wasted. Thirdly, this 'healthy' core must be able to resist changes that would drive the gut microbiome into dysbiosis, such as the use of antibiotics (Lloyd-Price *et al.* 2016). Dysbiosis refers to a change from the norm or 'healthy' microbiome community to one that can progress disease.

Taking all of the previous information into account, attempts were made to determine the bacterial content of the gut microbiome in healthy members of the population. The most well cited of these, the results from the HMP. The HMP concluded the dominant phyla in healthy adults to be *Bacteroidetes*, followed by a large proportion of *Firmicutes* and a small contribution of *Proteobacteria*. Consequently, the most dominant genus

present was *Bacteroides* (Consortium *et al.* 2012). A later study found the key components of the heathy gut microbiome to be *Bacteroides*, *Ruminococcaceae* (*Firmicutes*), *Clostridales* (*Firmicutes*), *Alistipes* (*Bacteroidetes*), and *Parabacteroides* (Li *et al.* 2013). A study investigating the enterotypes of the healthy human gut microbiome, found the 3 main genera to be *Bacteroides*, *Faecalbacterium* and *Bifidobacterium*. Ultimately form this research they proposed 3 enterotypes of the healthy human gut microbiome, *Bacteroides*, *Prevotella*, and *Ruminococcus* dominated (Arumugam *et al.* 2011). In summary, over 1000 gut microbial species have been identified (Rajilić-Stojanović *et al.* 2014), of which the most common in healthy individuals belong to the *Bacteroides* genus.

For an ecological community to be considered healthy it must be diverse and contain a significant number of inhabitants. Therefore, this is also true for the gut microbiome, as it is a micro-ecosystem. This theory of diversity equals health has been shown by investigations into diseases associated with the gut microbiome. Studies have demonstrated that gut microbiome diversity is decreased in obesity, inflammatory bowel disease (IBD) and diabetes (Manichanh *et al.* 2006; Turnbaugh *et al.* 2009; Giongo *et al.* 2010). This reduction in microbial diversity has been linked to the Western diet of high fat and sugar, accompanied by low fibre (Sonnenburg *et al.* 2016).

However, as with all biological systems a higher diversity does not imply health, as a high diversity in the vaginal microbiome has been associated with preterm birth (DiGiulio *et al.* 2015). Furthermore, it is unclear as to whether these changes are as a result of disease or if it is the microbiome changes driving disease progression, this will discussed further in 1.2.5.

A summary of the healthy human gut microbiome can be found in Figure 1. To summarize, the healthy gut microbiome must have a high diversity, a resistance to dysbiosis, the necessary house-keeping genes, a low amount of deleterious mutations and a high genetic potential. Furthermore, is most likely to be dominated by species from the *Bacteroidetes* and *Firmicutes* phyla.

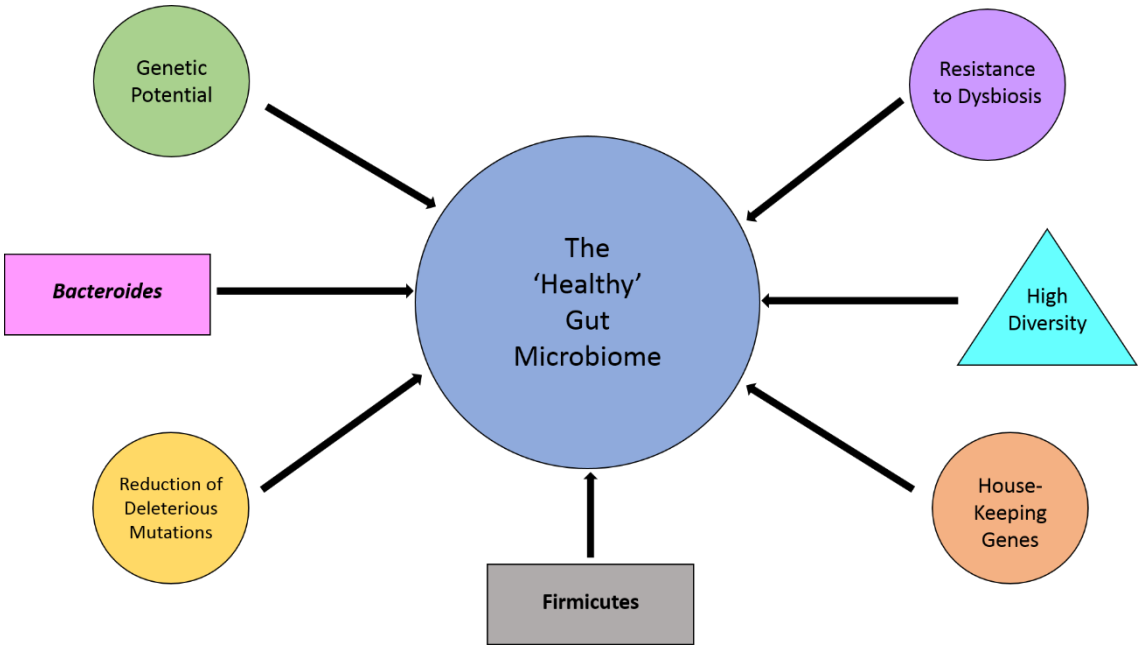

Figure 1. The Factors Contributing to a Healthy Gut Microbiome. A schematic representation of the factors contributing to a healthy gut microbiome.

1.2.2.3 A Lifelong Relationship

The gut microbiome is a constant companion throughout life and undergoes dramatic changes during this time. Colonisation begins before birth, detailed in 1.2.4, and continues throughout the first two years of life. From birth until 3 months of age, the infant microbiome is dominated by *Firmicutes*, after 3 months the community is dominated by *Actinobacteria*. This continues until 1 year of age when *Bacteroidetes* and *Firmicutes* are the main components (Koenig *et al.* 2011a; Azad *et al.* 2013). At two to three years of life, the infant microbiome begins to resemble that of an adult. This is because the abundance of *Clostridia* (*Firmicutes*) becomes predominant alongside the *Bacteroidia* (*Bacteroidetes*) (Avershina *et al.* 2016). As detailed in Section 1.2.2.2, the healthy adult microbiome is dominated by the *Bacteroidetes* and *Firmicutes* phyla, and remains relatively stable until old age. During old age *Bacteroidetes* increases, whilst the *Firmicutes* and *Proteobacteria* decrease (Claesson *et al.* 2011). Furthermore, in individuals over a hundred years old, the gut microbiome shifts to include more *Proteobacteria*, whilst the numbers of *Firmicutes* decrease (Odamaki *et al.* 2016). An earlier study showed that centenarians have a less diverse gut microbiome compared to adults and elderly individuals under a hundred years of age (Biagi *et al.* 2010). On a more individual basis the stability of an individual's gut microbiome was sampled over 46 years. The study discovered that host genetics were the main contributing factor

towards both changes and the maintenance of a core microbiome throughout a lifetime (Jayasinghe *et al.* 2017).

In summary, significant gut microbiome acquisition begins at birth and is subject to change during the first 3 years of life. Once this developmental period has past, the gut microbiome remains stable throughout adult life. As will be discussed in upcoming sections, the gut microbiome can undergo changes during this time as a result of medication or disease. At the transition into old age the gut microbiome undergoes changes to reflect the change in lifestyle the elderly enjoy. Finally, the gut microbiome is subject to change up until death as centenarians have a different community structure compared to younger individuals.

A key part of microbiome research is the way in which it is conducted. The current techniques use high-throughput sequencing technologies that can rapidly sequence numerous DNA sequences from numerous samples in a single experiment. It is often referred to as 16S rRNA gene sequence analysis.

1.2.3 16S rRNA Gene Sequencing for Gut Microbiome Analysis

For a century, thanks to the pioneer Robert Koch in 1881, the field of microbiology was entirely culture-dependent, and organisms could only be studied if it was able to grow in laboratory conditions. As a result, many of the inhabitants of the human gut could not be discovered or investigated due to the difficulty in culturing them. One hundred years after the development of plating techniques by Robert Koch, the discovery of DNA sequencing techniques was led by Fred Sanger. However, this technique remained out of the reach of mainstream research due to high time and monetary costs. It was not until the 1980s when DNA-based techniques became a mainstay of microbiology. One of the main draws to this technology was that a sample could be sequenced without the need for isolating and the culturing of pure cultures beforehand. The earliest of these techniques was fluorescent in situ hybridization (FISH), where nucleotide specific fluorescent markers are used to target genes of interest (Amann *et al.* 1995). Another popular technique at that time used the polymerase chain reaction (PCR) to amplify specific genes, which were then available for sequencing (Ward *et al.* 1990). It was this technique and the description of 16S rRNA gene fragments from a bacterial community, which led to the development of the application of 16S rRNA gene sequencing for

community analysis (Olsen *et al.* 1986; Giovannoni *et al*. 1990). This subunit is approximately 1542 nucleotides long and can form many bonds, creating a complex secondary structure, Figure 2. However, before the implementation of high throughput 16S rRNA technologies, 16S rRNA PCR-DGGE (denaturing gradient gel electrophoresis) methods were used to identify components of microbial communities. Using this method species are then identified by comparing the migration distance of band within the gel and comparing to a reference strain. DGGE gels are polyacrylamide containing a linear gradient of urea and formamide (Piterina and Pembroke 2013). As with previously described techniques the throughput for the DGGE method is significantly less than that for high throughput methods.

Ribosomal components are measured and identified in Svedburg units (S), based on the rate of sedimentation during centrifugations, the heavier more nucleotide rich molecules or subunits will sediment before the lighter, less nucleotide rich subunits. In bacteria, the three rRNA genes are organised into a single ribosomal operon. This entire operon is transcribed into a single 30S rRNA precursor, this is then subsequently cleaved by RNase III into the 16S, 23S, and 5S rRNA subunits, Figure 3 (Nikolaev *et al.* 1974). The unique characteristics of the ribosomal operon are size, sequence and secondary structures. These are all highly conserved within the bacterial kingdom (Maidak *et al.* 1997). Therefore, making them the ideal candidate for metataxonomic studies.

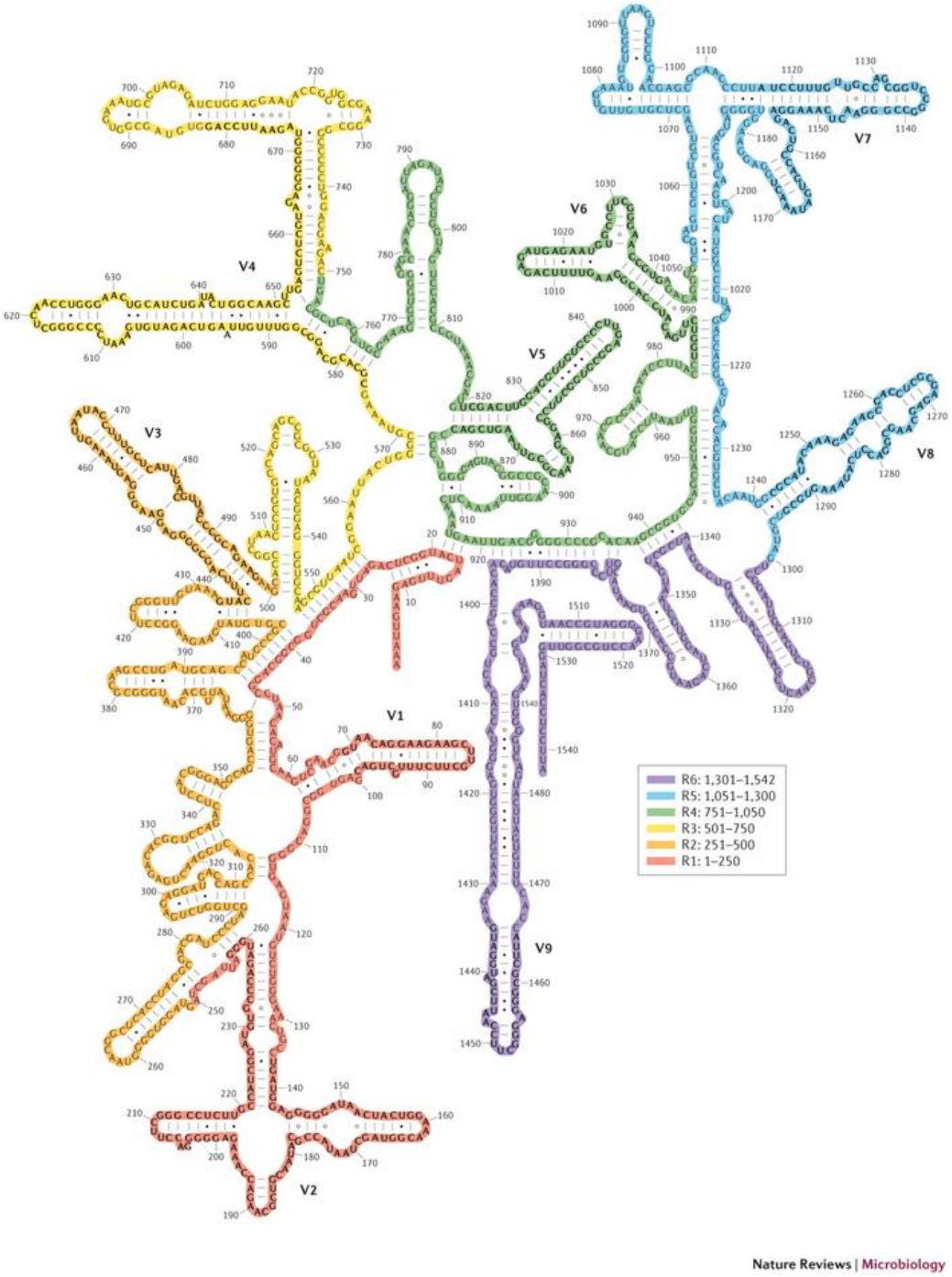

Figure 2. The 16S Ribosomal Subunit. The diagrammatic representation gives the full nucleotide sequence and secondary structures of the 16S subunit in *Escherichia coli*. Each region of the subunit is coded in a different colour and is approximately 250 nucleotides in length. Each one of the regions contains variable (V) regions. These variable regions are different in every species of bacteria and is the target area for sequencing in metataxonomic techniques. The R1, red, area contains V1 and V2; the orange area R2, contains V3; R3 in yellow contains region V4; the green area, R4, contains regions V5 and V6; R5, in blue, contains V7 and V8 regions; finally, the purple area, R6, contains the final V9 region. Permission given to reuse the image on 6.6.18, and is taken from (Yarza *et al.* 2014)

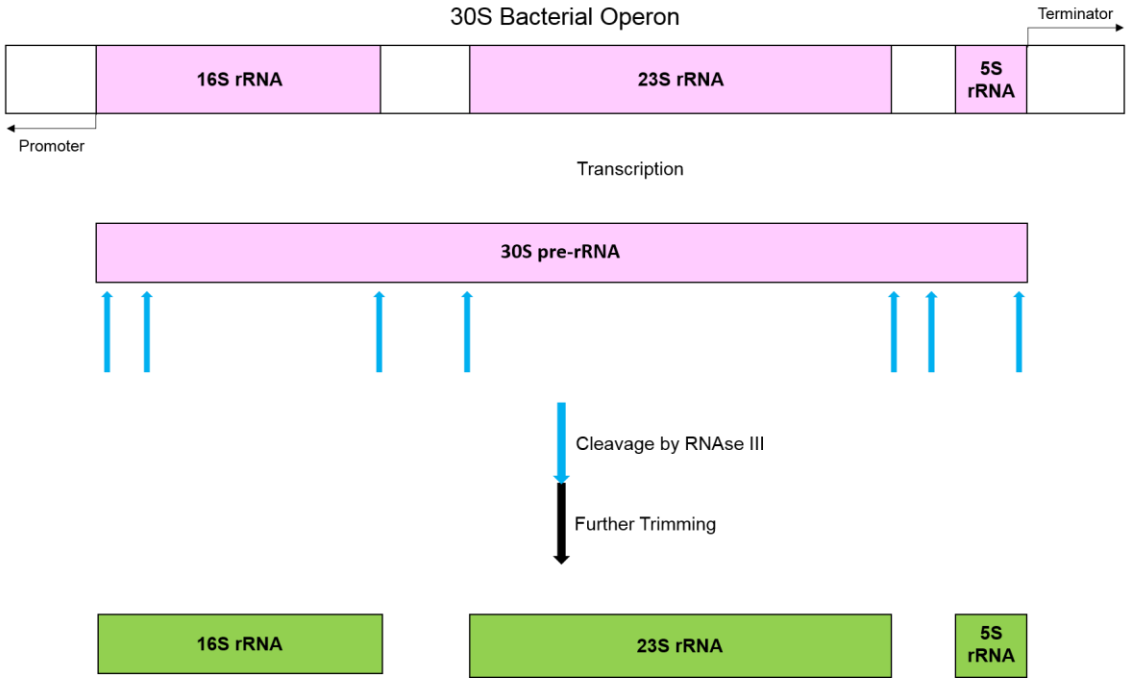

Figure 3. A Simplified Diagram of the Transcription and Processing of the 30S Bacterial Operon. The 30S bacterial operon contains an upstream promoter region and a downstream terminator region. In the middle of these regions is the DNA sequence for each subunit of the 30S bacterial ribosome. When the promoter region is open the operon is transcribed by an RNA polymerase. This 30S pre-rRNA sequence is then cleaved by RNase III to isolate the 16S, 23S and 5S rRNA subunits.

The first microbiologist to use this technique consistently was Carl Woese and colleagues in the late 1970s, as they were able to demonstrate that phylogenetic trees, by comparing taxa and their relationships, could be drawn by comparing relatively stable parts of the bacterial genome, the ribosomal operon (Woese and Fox 1977). Furthermore, the unique features of the ribosomal operon that make it extremely suitable for microbial identification is the presence of both conserved and hypervariable regions, Figure 2. These hypervariable regions are unique for every species of bacteria, therefore allowing for easy and reliable identification. In addition, this means that universal primers can be made that will bind to the conserved regions either side of one of these hypervariable regions, allowing for amplification. The bacterium are identified by aligning the sequencing results to a reference database and a match is considered if the percentage identity is above a certain threshold (Yarza *et al.* 2014). The first time this technique is demonstrated in microbial ecology was in the 1990s. It was a breakthrough, and led the way for future research, and ultimately microbiome research (Giovannoni *et al.* 1990).

Until the mid-2000s this was a highly used method for the area of microbial ecology. In 2005, the next step in sequencing technology was invented, high throughput sequencing, often referred to as ‘next generation’. As the discoveries of both Koch and Sanger had provided multiple advances, high throughput sequencing superseded both these techniques, such that an entire bacterial genome could be sequenced within hours (Metzker 2005). The previous drawbacks of older methods, such as incomplete community analysis, became possible with the advent of deep and high throughput sequencing methods. As a result, this is one of the most popular scientific techniques currently in use, as the entire genetic content of a sample can be analysed at relatively low economical and time cost.

Due to the financial viability of high throughput sequencing, large scale projects such as the Metagenomics of the Human Intestinal Tract (MetaHIT) and the Human Microbiome Project (HMP), became feasible. In 2008, the MetaHIT project was founded and aimed to elucidate the microbial genomes of the gut, using stool samples donated from IBD, obese and healthy participants (Qin *et al.* 2010; Arumugam *et al.* 2011; Le Chatelier *et al.* 2013; Li *et al.* 2014). This was a four and a half year project, 50% funded by the European Union, with a total cost of 22 million euros. In contrast, the HMP cost a total of 115 million dollars, ran between 2008 and 2013, and was funded by the US National Institute of Health. The aims of the HMP was to determine the diversity of the microbiome, sampled from numerous body sites such as, the gut, skin and vagina (Peterson *et al.* 2009; Consortium *et al.* 2012; Weinstock 2012).

To summarise, microbiology has come a long way since the culture dependent days of the late 1800s. The wealth of data that has that has been generated from high throughput methods is akin to the industrial revolution. However, our knowledge of bacterial communities would not have been possible without the discovery of hypervariable regions in the 16S ribosomal subunit. However, we are at a critical point in microbiome research where we know the microbial communities in various systems but cannot conclusively link them to disease or other functions. Therefore, the next breakthrough in microbiome research will be the discovery of a functional assay to assess the roles these organisms play in the microbiome.

In Section 1.2.2.3, the development of the gut microbiome was briefly explained in relationships to our lifelong relationship. However, as this thesis focuses on the

development of the gut microbiome, it is necessary to expand upon the research mentioned previously.

1.2.4 The Development of the Human Gut Microbiome

1.2.4.1 Prenatal Development of the Human Gut Microbiome

There is conflicting evidence debating the 'sterile womb' hypothesis, which will be discussed in this section. However, the current consensus within the scientific community is that in a healthy pregnancy it is normal for low levels of bacteria to be present in the placenta, amniotic fluid and meconium. To add, in a healthy pregnancy the bacteria present in the areas are of a low abundance, richness and diversity (Collado *et al.* 2015). However, infections such as chorioamnionitis do occur, and result in a very different prenatal environment.

Research in 2005 first identified the types of bacteria present in the umbilical cord of healthy neonates born by caesarean section. In 45% of the samples tested, bacterial cultures were grown and identified to contain species from the following genera, *Enterococcus*, *Streptococcus*, *Staphylococcus*, and *Propionibacterium* (Jiménez *et al.* 2005). These bacteria were described as commensal and the findings of Stout *et al* in 2013 reiterate this. They found that 27% of placentas examined were found to contain intracellular bacteria. Furthermore, they investigated if this was related to preterm birth and intrauterine infection, such as chorioamnionitis, and found no link between the bacteria present in the placenta and these adverse events (Stout *et al.* 2013).

To expand on the study by Collado *et al*, they found an array of bacterial diversity across several sites from the mother and new born infant. Firstly, they discovered a unique microbial community in the placenta and amniotic fluid, distinct from the maternal faecal microbiome. The most prominent phylum in these samples were the *Proteobacteria*. Furthermore, they observed that the bacterial communities at different sites and individuals were highly similar and consistent. Moreover, to solidify the presence of bacteria *in utero* they found that the majority of the microbial community observed in the placenta and amniotic fluid was shared with the meconium of the infant. This shared community included, *Streptococcus*, *Enterobacteriaceae*, *Propionibacterium*, *Lactobacillus*, and *Bacillales* (Collado *et al.* 2015). An earlier study

showed similar results. Aagaard *et al* in 2014 found the placental microbiota to contain the phyla, *Firmicutes*, *Tenericutes*, *Proteobacteria*, *Bacteroidetes* and *Fusobacteria*; two of which were also identified in Collado *et al*'s 2015 study. Of interest, this study also found the microbiota of the placenta was significantly associated with preterm birth. Furthermore, it was noted that the microbiome of the first stool passed by the infant most closely resembles that of the amniotic fluid. It is believed that this is a result of the swallowing of the amniotic fluid during the third trimester of pregnancy. This link between the *in utero* microbiome and infection has been further investigated (Aagaard *et al.* 2014).

Chorioamnionitis and intrauterine infections have been known to cause preterm birth for over twenty years, however it was only recently that the organisms responsible have been investigated (Seo *et al.* 1992). The microbiome of preterm infants, born to mothers with chorioamnionitis, were enriched with urogenital and oral commensal bacteria (Prince *et al.* 2016). Therefore, the *in utero* microbiome is subject to infective processes.

Notable research by Gosalbes *et al* in 2012, found a link between infant meconium and maternal factors. For example, they found that a maternal history of atopic eczema resulted in a reduced microbiome diversity in the meconium (Gosalbes *et al.* 2016). This and previously discussed research suggest the *in utero* environment to be non-sterile.

On the contrary, some studies have suggested that the womb is a sterile environment (DiGiulio 2012). Therefore, a consensus on the sterility of the womb and the *in utero* environment has not been established. For the purpose of this thesis, is it assumed that the *in utero* environment contains a unique microbiome that contributes to the developing GI microbiome of the new born infant.

1.2.4.2 Post Natal Development of the Human Gut Microbiome

Postnatal development of the infant gut microbiome is considered to begin during the birthing process and conclude at approximately two years of age. There are many factors that influence this process and are discussed in more detail in 1.2.4.3. However,

"healthy" postnatal infant development can be considered as a vaginal birth, breast feeding, and no health complications requiring medication such as antibiotics.

Studies from the era of culture-based techniques through to high throughput methods show that infants born via the birth canal and receive only breast milk have a microbiome dominated by the *Bifidobacterium* genus (Mackie *et al.* 1999; Favier *et al.* 2002). Once weaning commences the number of bacteria that can utilize oligosaccharides decreases and a community dominated by *Firmicutes* and *Bacteroides* begins to become established (Koenig *et al.* 2011a). However, before weaning commences around 6 months of age, there are temporal changes occurring in the infant gut. Firstly, the profile is dominated by intrauterine or vaginal associated taxa. Secondly, skin derived taxa become predominant such as *Streptococcus* and *Enterobacteriaceae*. Thirdly, a domination of *Bifidobacteriaceae* occurs in the faecal microbiota. Finally, adult-like taxa begin to appear such as *Blautia, Eggerthella* and *Clostridium* (Timmerman *et al.* 2017).

A recent metagenomic study found that at 6 weeks of life, the metagenomes of the stool begin to converge between individuals, suggesting the development of a core set of metabolic pathways (Chu *et al.* 2017). In addition, Del Cheirico *et al* in 2015 reported upon a "core microbiome" during the first 30 days of life (Del Chierico *et al.* 2015). In addition, to encompass all areas of the microbiome, recent research has shown a succession of archaea and microeukaryotes during the first year of life (Wampach *et al.* 2017).

1.2.4.3 Factors Affecting Gut Microbiome Development

1.2.4.3.1 The Effect of Mode of Birth on the Development of the Gut Microbiome

The effect of mode of birth on the gut microbiome colonization was published in 1999. Grönlund *et al* found that the *Bacteorides* were reduced in caesarean delivered infants (Grönlund *et al.* 1999). Since 1999, caesarean section (CS) has been found to seed the neonatal microbiome with opportunistic pathogens such as, *Haemophilus*, *Enterobacter*, *Veilonella* and *Staphylococcus* (Dominguez-Bello *et al.* 2010; Bäckhed *et al.* 2015). This predominantly from the mother's skin microbiome. Furthermore, these infants have a delayed colonisation and reduced diversity of the *Bacteroidetes* phylum

(Jakobsson *et al.* 2014). In a subset of the population this reduced colonisation by *Bacteroidetes* was also reported, alongside an increased incidence of *Bacilli* and *Clostridium* in the CS delivered infants. Interestingly, they observed that this difference in colonisation disappeared with age and the microbiota of both CS and vaginally delivered infants became similar (Lee *et al.* 2016). In addition, to changing the microbial content of the gut, CS infants have been reported to have a reduced diversity during the first years of life (Jakobsson *et al.* 2014; Lee *et al.* 2016). A recent systematic review provided further evidence to a reduced diversity in CS delivered infants, and showed that the significant differences observed in the infants during the first 6 months of life were removed after the first 6 months (Rutayisire *et al.* 2016).

Not only has CS delivery been associated with changes to the microbiota recently after birth, long term affects have been noted in a handful of studies. CS delivery has been associated with childhood obesity (Mueller *et al.* 2015). With the underlying causes still unknown, it was found that elective CS was associated with adult psychosis (O'Neill *et al.* 2016).

A vaginal delivery is currently accepted as the norm as it is believed to be more beneficial, as demonstrated in several studies. Infants delivered vaginally are enriched with *Escherichia-Shigella* and *Bacteroides* compared with infants delivered by CS (Azad *et al.* 2013). Furthermore, there is a high level of *Lactobacilli* in vaginally delivered infants (Dominguez-Bello *et al.* 2010; Aagaard *et al.* 2012; Avershina *et al.* 2014).

Conversely, a recent study by Chu *et al* found that body site had more of an effect on microbial reorganisation than the mode of delivery during the first 6 weeks of life (Chu *et al.* 2017). Furthermore, a phylogenetic study during the first 30 days of life found that there is a "core microbiome" irrespective of birth mode (Del Chierico *et al.* 2015).

In summary, it is currently accepted that a vaginal delivery seeds a more beneficial and "healthy" microbiome compared to a CS delivery. However, as detailed there is research to suggest that it has less of an effect. As a result, more longitudinal research needs to be done in order to provide a conclusive answer and discover if mode of delivery is responsible for disease etc., in later life. Finally, an infant's microbiome

reflects the vaginal microbiome when born naturally, however, the mother's skin microbiome predominates that of CS delivered infants.

1.2.4.3.2 The Effect of Diet on the Developing Gut Microbiome

As discussed previously, the gut microbiome encompasses not only the organisms that reside there but also host physiology and transient components. The food that we ingest as part of our diet is part of this transient community and, therefore, can have significant impact on the composition and interactions within this micro ecosystem. For the developing gut, a substantial body of research, has shown that breast milk is still considered "best" for the most beneficial outcomes. However, specific components and additives, such as probiotics, have been shown to have a significant effect on the developing gut microbiome.

A multi-centre study, published in 2016, reported that a formula based diet that included a *Bifidobacterium breve* probiotic, provided adequate nutrition for normal growth in healthy infants when compared to World Health Organization (WHO) guidelines (Abrahamse-Berkeveld *et al.* 2016). An earlier study discovered that infants who received a more breast milk like formula or a probiotic showed similar *Bifidobacteria* counts to breast milk fed individuals and higher counts than those on a readily available formula (Hascoët *et al.* 2011). This suggests that there is the potential for infant formula to mimic that of breast milk, which as these and other studies have shown breast milk to be more beneficial.

A unique study by Anvarian *et al* in 2016, investigated the bacteria present in a powdered infant formula production facility. The results of their study noted the following organisms colonising all areas of the production facility, *Actinobacter*, *Pseudomonas* and *Streptococcus* (Anvarian *et al.* 2016). This is an interesting finding, as will be discussed in following sections, these are bacteria often found in preterm infants, and are more likely to be formula fed.

In vivo studies into the effect of formula feeding on developing offspring have also been conducted, such as the following on a neonatal porcine model. The benefit of using animal models is the ability to study interventions and histological changes. This study demonstrated significant physiological changes in the formula-fed cohort such as, ileum

and jejunum villus length were increased along with the depth of Peyer's patches. In addition, lymphoid follicle size was decreased, indicating at a reduction in immune education. Immunological changes included an up-regulation in AMCFII, IL-8, IL-15, VEGFA, LIF, FASL, CACL11, CCL4, CCL25, and a down regulation in IL-6, IL-9, IL-10, IL-27, IFNA4 and CSF3 at transcript level. This study added further evidence to the effect of formula feeding on the bacterial content of the developing gut. There was a significant increase in *Lactobacillaceae* and *Clostridia* in the sow fed group compared to formula fed piglets who were found to have higher levels of *Enterobacteriaceae*; a pattern also seen in human infants (Yeruva *et al.* 2016). This data shows that not only can a formula dominated diet affect the bacterial content of the developing gut, but also host physiology and immune development.

1.2.4.3.3 The Effect of Antibiotics on the Development of the Gut Microbiome

The effect of antibiotics on the gut microbiome of preterm infants will be discussed in more detail in Chapter 6. In summary, research has shown that antibiotic use has been linked to adverse outcomes in the preterm population. However, any links between the gut microbiome, antibiotics and disease or adverse outcomes has yet to be investigated. Moreover, antibiotics are the most frequently administered drugs in the NICU.

1.2.5 Changes in the Gut Microbiome during Disease

The definition of disease within an environment, such as the gut microbiome, is referred to as a dysbiosis. Therefore, dysbiosis within the gut microbiome is considered to be a departure from a balanced ecological state by an environmental change (Petersen and Round 2014). There are a number of diseases affected by changes in the gut microbiome, however the most prevalent in Western society are inflammatory bowel diseases (IBD), such as Crohn's disease (CD) and ulcerative colitis (UC). IBD is used as an umbrella term for CD and UC, as both diseases are mediated by inflammatory responses (Stange *et al.* 2006; Mulder *et al.* 2014). Using adult intestinal diseases as a basis, the causes of similar neonatal disease can be investigated. For example, CD and IBD are discussed here, as both are inflammatory diseases, and have a similar pathology to necrotising enterocolitis (NEC). Therefore, in order to understand possible pathologies previously found in adults and their possible role in preterm infants, an understanding of these diseases is necessary.

1.2.5.1 Inflammatory Bowel Disease: Crohn's Disease and Ulcerative Colitis

Crohn's disease (CD) is defined as a transmural inflammatory disease of the mucosa that affects all parts of the gastrointestinal (GI) system. It appears episodically and progresses in severity with each occurrence. Symptoms of the disease include parts of the GI system shutting down, this then leads to complications such as strictures, fistulas and abscesses (Silverberg *et al.* 2005; Stange *et al.* 2008). Ulcerative colitis (UC) is defined as a non-transmural inflammatory disease, which includes the episodic progression seen in CD. However, this disease is limited to the colon, but the inflammation can spread to the terminal ileum, therefore making it difficult to distinguish from CD (Silverberg *et al.* 2005; Stange *et al.* 2008). IBD severely affects a patient's quality of life and is great cost to health care (Molodecky *et al.* 2012; Tóthová *et al.* 2014). As a result, a substantial amount of research has been committed to discovering the origins of this disease. To add, IBD is a multi-factorial disease continuing with the definition of the microbiome to be affect by physiology, microbiology, immunology and genetics. However, for the purpose of this thesis, the microbiological changes will be focused upon.

To begin, CD patients have a reduced Shannon diversity index compared to healthy controls. Furthermore, the overall microbial content of the gut is different from that of healthy counterparts (Fujimoto *et al.* 2013). This was also seen in an earlier study where the increased inflammatory response seen in IBD is a result of a loss of anaerobic bacteria in the colonic mucosa (Ott *et al.* 2004). In contrast, the microbial diversity in patients with CD is unaffected by their disease state (Seksik *et al.* 2003). On the other hand, this same study identified an increase in *Enterobacteriaceae* in the CD affected cohort. This decrease in diversity has been accredited to the overall loss of diversity within the *Firmicutes* phylum (Kang *et al.* 2010).

Specific microbial changes occurring in CD include a significant decrease of *Faecalbacterium prausnitzii* compared to healthy controls. In addition, *Bifidobacterium wadsworthia* was also decreased in CD patients. The ileum CD phenotype has been found to correlate with a dysbiosis of the ileum mucosa associated bacteria. As a result, it is plausible to suggest that a novel group of invasive *E. coli* is involved in the pathogenesis of CD (Baumgart *et al.* 2007). Furthermore, virulence and secretion metabolic pathways are enhanced during ileum CD (Morgan *et al.* 2012). The

Fusobacterium genus has also been shown to increase in the colon of UC patients (Toshifumi *et al.* 2002).

Furthermore, in keeping with the microbiome definition, Morgan *et al* also found that metabolic pathways were more affected by disease status than the taxa of the microbiome (Morgan *et al.* 2012). In addition, the fungi *Candida albicans* has been found to colonize all sections of the GI tract in patients with IBD (Trojanowska *et al.* 2010).

1.2.6 Summary

The human microbiome includes several different body sites and performs a plethora of functions to maintain health for both the host and its symbiotic community. Furthermore, the microbiome is a combination of organismal, host and transient components. The gut microbiome encompasses the GI tract from small intestine to colon; includes organisms such as bacteria, archaea and fungi; and can be changed due to diet, a transient component. Therefore, the combination of all this makes it an exciting and rigorously researched area.

In order to fully investigate the gut microbiome, technology has made significant advances to allow for research into this area to become what it is today. The field of microbiology has moved from culture-dependent to culture-independent methods, from petri-dishes to high throughput sequencing. This leap in technology has allowed researchers discover that gut microbiome colonisation begins before birth and continues to change and adapt throughout a lifetime. Furthermore, it has given insights into the role of the microbiome during diseases of the gut, such as IBD. However, it is difficult to draw a solid conclusion about the impact of the gut microbiome on disease, as they are multifactorial conditions, resulting in discrepancies between the studies discussed previously. Although, it can be said with a high degree of certainty that changes in the gut microbiome during IBD play a significant role in the pathophysiology of the condition. Furthermore, the role of proteases, discussed in Section 1.4, provides further evidence to the role of the gut microbiome in the development of intestinal diseases. Moreover, it will become clear the association between IBD, proteaes and NEC. Ultimately, this will lead to a greater understanding of the interactions between the human host and organismal content in preterm infants.

1.3 Premature Infants

1.3.1 Introduction

An infant is classed as premature when delivered before 37 weeks' gestation. This definition is further sub-divided into extreme preterms or extremely low birth weight (ELBW), less than 1 kg, born at less than 28 week's gestation; very preterm or very low birth weight (VLBW), less than 1.5 kg, delivered between 28 and 32 week's gestation; and finally late preterms or low birth weight (LBW), less than 2.5 kg, born between 32 and 37 week's gestation. Currently, preterm birth is the leading cause of neonatal morbidity and mortality worldwide (Lawn *et al.* 2005; Beck *et al.* 2010). The consequences of preterm birth are not only acute but can have chronic and dramatic consequences throughout life (Huddy *et al.* 2001; Wang *et al.* 2004; Goldenberg *et al.* 2008). Not only are there long term health issues for the preterm infant, the cost of healthcare during and long after birth, is a burden on the global economy (Petrou *et al.* 2003; Petrou 2005). As a result, research into the prevention of preterm birth and the treatment and care of preterm infants is of great importance.

The care of premature infants has developed extensively over the past 50 years to provide safe, prompt and effective care to the most highly dependent infants. The criteria for admission to the neonatal intensive care unit (NICU) is an extensive list but includes, for example: 1) gestational age of less than 34 weeks, 2) a birth weight less than 1.8 kg, 3) prolonged resuscitation, 4) severe congenital abnormalities. In general, an infant is admitted to the NICU who needs close observation, continuous monitoring or active management by a specialist neonatal team.

Upon admission to the NICU a number of checks and baseline measurements will be taken in order to draw up a care plan and ultimately to deliver appropriate care to see the infant through to health and discharge. In order to achieve this goal, a substantial number of treatments are administered.

1.3.1.1 Treatment Received in the Neonatal Intensive Care Unit

Preterm infants receive a plethora of treatments in the NICU (Hsieh *et al.* 2014). Antibiotics are the most common medication prescribed in the NICU, shown

specifically to be ampicillin in the Clark *et al* study (Clark *et al.* 2006; Patel *et al.* 2009; Schulman *et al.* 2015). For example, studies by Tripathi *et al* and Piantino *et al* in 2012 and 2013, respectively, show that 95% of the NICU patients, 65% of VLBW and 50% of LBW infants receive antibiotics for more than 3-5 days with negative culture results (Tripathi *et al.* 2012; Piantino *et al.* 2013). The reason for this systemic use of antibiotics is the extensive list of risk factors such as pre-eclampsia, chorioamnionitis, pneumonia, sepsis, low birth weight, prematurity, feeding and ventilation, to name a few (Cotten *et al.* 2009; Kuppala *et al.* 2011). In addition, they are often used prophylactically to cover GBS and *E. coli* infections. Furthermore, clinicians prefer to err on the side of caution when the above are present (Stoll *et al.* 2002; Stoll *et al.* 2011; Wirtschafter *et al.* 2011; Cantey *et al.* 2016). Therefore, the role of antibiotics on the development of the preterm infant will be investigated in Chapter 5.

However, this widespread use of antibiotics is not always associated with healthier patient outcomes. For instance, several studies have shown that late onset sepsis (LOS), NEC, fungal infections, NICU outbreaks, mortality, increased hospital stay and costs, are associated with blanket antibiotic prescriptions (Cotten *et al.* 2009; Kuppala *et al.* 2011; Afjeh *et al.* 2016). In addition, the risk of sepsis, especially from commensals and fungi, is increased in preterm infants who are exposed to broad spectrum antibiotics (Madan *et al.* 2012; Mai *et al.* 2013).

Not only are infections problematic in the NICU, there are a number of serious complications associated with preterm birth. The first is patent ductus arteriosus (PDA), where indomethacin is the conventional treatment, ranked the 8th most used treatment in US NICUs. The second is bronchopulmonary dysplasia (BPD), where oxygen is a commonly used treatment (Hsieh *et al.* 2014). As a result, ventilation is a common treatment given in the NICU. Moreover, respiratory distress is an admission criteria.

The preterm infant has complicated and demanding needs, therefore a substantial amount of research has been conducted into providing the best care possible to improve long term outcomes. As a result, guidelines on feeding and ventilation are constantly being updated taking into account the latest research. However, there are short term goals in mind that allow the medical team to determine the success of the treatment. For feeding the aim is to accelerate infant growth, meet the nutritional needs of the infant,

and prevent feeding related morbidities such as necrotizing enterocolitis (NEC) (Thomas 2016).

There are a number of key factors to consider when started feeding in the hospitalised infant. An infant should begin feeding within 24 hours of life unless medically unfit to do so. In this situation, an infant would be given intravenous parenteral nutrition. If parenteral feeding is the main source of nutrition for the infant than minimal enteral feeding should be considered to be delivered in parallel. The aim of this is to provide up to 1 ml/kg/day of milk to stimulate gut hormone production. The next factor to be considered is the rate at which to increase feeds of they are well tolerated by the infant.

Typical feeds given in the NICU range from breast to formula milk, due to several criteria that must be considered before the feed is chosen. Breast milk is considered the gold standard in the feeding of preterm infants, who make up the majority of admissions to the NICU. Breast milk fortification (BMF) can be given to infants of a certain weight to increase the protein intake in infants receiving breast milk. BMF is not needed if the infant is taking preterm formula. Donor breast milk (DBM) is another option if the infant's mother is unable to express milk herself. However, the supplies of are often inconsistent. Another obvious choice is the use of preterm formulas, if the infant is under 34 weeks' gestation, if greater than 34 weeks' gestation and term formula can administered. When the infant is well enough to be discharged from the NICU, nutrient enhanced post discharge formulas can be used until catch up growth has been achieved.

Currently the use of probiotic in the NICU is not standard practice, however some of the infants recruited to this study were administered this treatment. Probiotics will not be administered until the infant is receiving a given amount of enteral nutrition.

An infant would be considered for ventilation in the NICU under the following circumstances, the presence of a lung disease, having a poor respiratory drive, lung malformations and mechanical issues such as blockages. Similar to feeding, the overall goal of ventilation is to improve the long-term outcomes for the infant, but again there are acute benefits also. Firstly, adequate oxygenation and ventilation are key, followed by patient comfort and a decrease in ventilation associated lung injury.

1.3.1.2 Is the NICU a Breeding Ground for Bacteria?

As with any hospital environment, every care is taken to ensure a sterile and pathogen free level of care. However, this is an impossible target and hospital acquired infections are an everyday par of hospital life. There has been research showing the NICU is no exception. However, one could argue that the NICU is a particularly potent breeding ground. For instance, the infants are housed in incubators to mimic the *in utero* environment, this means they are warm and humid, and the perfect conditions for bacterial growth.

In a study into the bacterial diversity of hospital equipment in a NICU a high diversity of organisms was found, closely resembling that of general building surface and air samples. Many of the genera in these samples were opportunistic pathogens such as *Propionibacterium*. Furthermore, in one of the sites sampled there was a high number of faecal coliform bacteria, such as *Enterobacteriales*, in samples taken from surfaces (Hewitt *et al.* 2013). In addition, to a high diversity of bacteria within the NICU, the presence of specific pathogens has also been tracked. In one NICU in Zurich, there was three consecutive outbreaks of *Serratia marcescens*. Numerous steps were taken to isolate the source of contamination and prevent further outbreaks, however these were proven to be ineffective (Fleisch *et al.*).

In contrast to this, an interesting study investigated the effect of cleaning on the NICU environment. They found that intensive cleaning of surfaces in contact with the neonates did decrease the overall bacterial diversity and the levels of *Streptococcus* and *Staphylococcus* species. This is encouraging as these bacteria are responsible for numerous nosocomial infections. Finally, all surfaces tested had a low level of bacterial diversity irrespective of cleaning (Bokulich *et al.* 2013).

One of the huge problems facing research into hospital acquired infections is the proliferation of antibiotic-resistant strains of pathogens. One study showed that vancomycin resistant *E. coli* was present in 12% of infants on a NICU in Australia (Flokas *et al.* 2017). Furthermore, a study in two NICUs in the Philippines showed that nearly half if all infants admitted to the NICU were found to be colonized by drug-resistant bacteria. More importantly, the rate of colonisation with these drug-resistant bacteria did not change after infection control procedures were successfully introduced

(Gill *et al.* 2009). In contrast, an increase in antibiotic resistant infections by the pathogen *Klebsiella pneumonia* was due to a breach in infection control procedures by NICU staff (Fabbri *et al.* 2013).

In summary, numerous incentives are in place to reduce and prevent potential pathogenic outbreaks in the NICUs. However, it appears that even with all of these fail safes in place, opportunistic pathogens are still able to exert their effects. Perhaps this is due to the overuse of antibiotics and cleaning within the NICUs. Overall, much more research is needed to determine the most effective protocol for infection control in the NICU. Moreover, this suggests that the NICU has the potential to influence the acquisition of the preterm infant gut microbiome.

1.3.2 Overall Normal Infant Development and Specific Gastrointestinal Development

1.3.2.1 Prenatal Development

The definition of prenatal development is the process of growth and development within the womb from fertilization to birth (Dean and Grizzle 2011). There are many systems that begin to develop to maturity during this time. The focus in this thesis will be the prenatal development of the gastrointestinal system.

Embryonic gastrointestinal development is formed from the endoderm, bending from the head to the tail of the embryo to form a tubular gut, where the yolk sac attaches in the middle. This tubular gut comprises of three regions, see Table 2. During foetal development the foregut gives rise to the pharynx, oesophagus, stomach, cranial halves of the liver, gallbladder, bile duct and pancreas. The intestine is formed from a mix of the mid and hind gut to form the duodenum, jejunum, ileum, cecum, appendix, ascending colon, transverse colon, sigmoid colon and rectum. The developing gastrointestinal tract begins to form distinct histology through a process called recanalization. The intestinal villi begin to form around 11 weeks post-conception along with enterocyte differentiation and goblet cells. At 13 weeks the stomach, duodenum and small intestine are now fully grown, prior to birth. By 16 weeks the whole length of the intestine has villi. Finally, at 20 weeks, Peyer’s patches begin to appear and all major components of the mature gut are present (Harding and Bocking 2001).

Table 2. The Formation of the Tubular Gut. A table describing each section of the tubular gut and how it is formed during prenatal development.

Tubular Gut	Formation
Foregut	Formed by a lateral and head fold of the endoderm.
Midgut	Comprises of the yolk sack.
Hindgut	Formed by a lateral and tail fold of the endoderm.

Akin to the foetal period of embryonic growth there are a plethora of structural, functional and molecular changes occurring throughout the human body postnatally. As a result, the focus of this section will be to detail the postnatal maturation of the intestinal system.

The most prominent structural change after birth is the gradual elongation of the small intestine throughout the first twenty years of life. Other than this structural changes during the maturation of the intestinal system after birth occurs by interactions with the developing microbiome and the education of the immune system. The development of gut immunity will be discussed in a later section and the microbial colonization patterns of the newborn infant have already been detailed. Therefore, this section has been rather short.

1.3.3 The Development of the Preterm Infant with a Focus on Gastrointestinal Development

A large field of research has detailed the changes in growth and development in infants born before 37 weeks' gestation. Overall, the picture is of underdevelopment and is referred to as catch up growth. There are numerous diseases and disabilities linked to preterm birth such as asthma, behavioural issues, autism-spectrum, blindness etc. The long-term outcomes for infants born at less than 37 weeks' gestation have dramatically improved over recent decades. As previously mentioned the structure of the gastrointestinal system is developed by 20 week's gestation, therefore the preterm infant has the necessary structural components but lacks the immune system and brain development to fully exploit these structures (Knight *et al.* 2014).

One of the most problematic areas of intestinal function for premature infants is motor function. For example, the suck swallow ability is not developed until approximately 34 weeks' gestation, the motility of the intestine is underdeveloped, and the stomach shows delayed emptying. This is due to the immaturity of the enteric nervous system. This delay in passage can cause bacterial overgrowth and distension, along with immune immaturity can lead to the development of a common preterm morbidity, NEC (Neu 2007a). However, the risk factors and the pathology of NEC will be discussed in a later section.

1.3.4 The Gut Microbiome of Preterm Infants

Due to numerous factors, previously discussed in detail, the gut microbiome of preterm infants is remarkably different to that of full-term infants. One of the main reasons for this is the difference in care environments, such as the NICU. Other reasons include mode of delivery etc. Examples of this includes, limited exposure to parental skin, type of feed received, environmental surfaces, health care workers skin and antibiotic use. Research currently dictates that full-term infants have a diverse and rich gut microbiome dominated by *Firmicutes* and *Bifidobacteria*. This is in stark contrast to that of preterm infants.

Firstly, it is believed that the cumulative exposure of antibiotics in the NICU, results in a significant reduction of species richness and diversity. In addition, in a majority of infants sampled, bacteria of the gut microbiome contained plasmids encoded antibiotic resistance genes for more than six antibiotic classes (Gibson *et al.* 2016; Ward *et al.* 2016). Not only do treatments received in the NICU contribute to colonization patterns of the gut microbiome, the abnormal environment has been found to play a significant role.

As previously described the NICU is a breeding ground for bacteria. Therefore, it is interesting for these environmental bacteria to be seen in the guts of hospitalized infants. One study found that dominant gut taxa such as *Staphylococcus epidermidis*, *Klebsiella pneumoniae*, *Bacteroides fragilis*, and *Escherichia coli* were also found in over half of samples taken from the NICU environment (Brooks *et al.* 2014).

Several studies have shown that the initial colonization of the preterm infant gut begins with Gram-positive cocci, such as the *Bacilli* family, from the *Firmicutes* phylum. These initial colonizers are then overtaken by facultative anaerobes within the *Gammaproteobacteria* class. This then leads to a final strictly anaerobic state (Jacquot *et al.* 2011; Normann *et al.* 2013; Torrazza *et al.* 2013; La Rosa *et al.* 2014; Sim *et al.* 2015; Zhou *et al.* 2015). Due to the lack of diversity and richness seen in these infants, greater than 90% of the organisms predominate the microbiome of preterm infants (La Rosa *et al.* 2014). Furthermore, the *Gammaproteobacteria* class are proportionally over represented in preterm infants, often comprising greater than 50% relative abundance. This contrasts with less than 20% seen in full term infants (La Rosa *et al.* 2014; Ward *et al.* 2016).

1.3.5 Necrotizing Enterocolitis

Necrotizing enterocolitis, or NEC, is the most common gastrointestinal disease seen in the preterm infant population. It is an extremely serious disease with mortality rates as high as 30% in VLBW infants (Kosloske 1994; Holman *et al.* 2006; Hunter *et al.* 2008a; Fitzgibbons *et al.* 2009; Horbar *et al.* 2012). In addition to the high mortality rates, infants who survive are left with significant morbidities. This is a result of invasive surgery to remove the necrotized parts of the bowel, leading to short bowel syndrome (Salhab *et al.* 2004; Blakely *et al.* 2005; Schulzke *et al.* 2007; Wadhawan *et al.* 2013). Therefore, a vast amount of research time and money has been invested to investigate the cause of this disease. This research has shown that NEC is a multi-factorial disease, as shown in Figure 4. As a result, research into potential treatments and therapies has been more problematic.

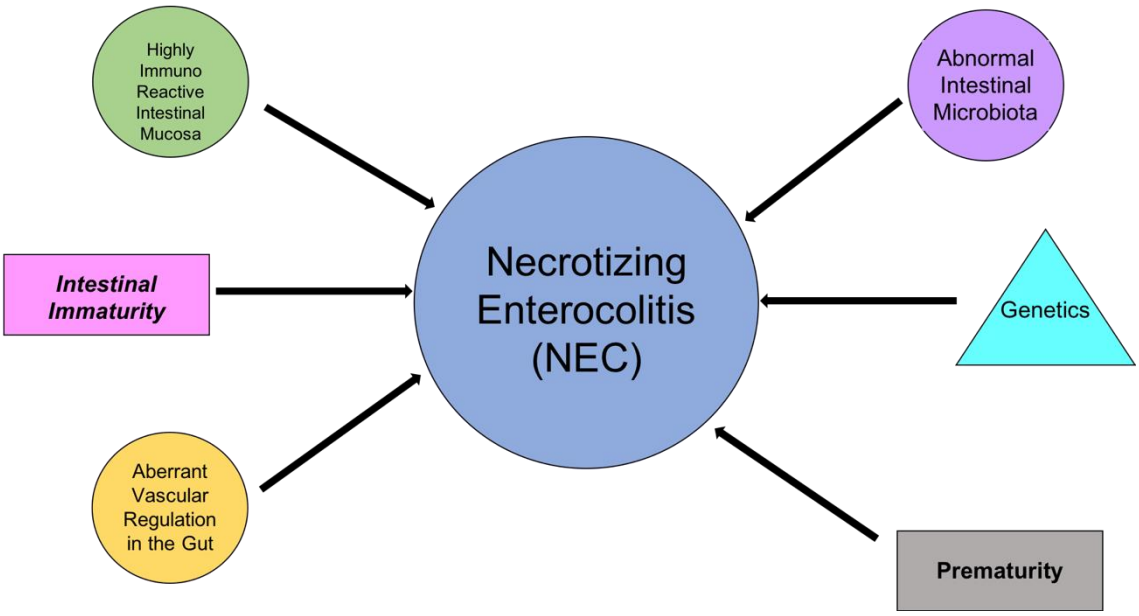

Figure 4. Necrotizing Enterocolitis is a multifactorial disease. Adapted from (Haque 2016).

The disease is characterized by increasing levels of damage to the intestinal tract, beginning with mucosal injury to full thickness necrosis and subsequent perforation. The first symptoms of NEC include vomiting, diarrhoea, a delay in gastric emptying, abdominal distension and or tenderness, decreased bowel sounds, abdominal wall erythema. Along with these observational symptoms, there are others that can also indicate the development of NEC, such as apnoea, lethargy, decreased peripheral perfusion, shock, cardiovascular collapse and bleeding diathesis. If NEC is suspected a number of clinical tests can be performed in order to determine a diagnosis, such as white blood cell and platelet counts, blood loss, blood culture, plasma sodium and bicarbonate levels, arterial blood gas levels and abdominal radiography. To note, all of these symptoms will not be present in the infant, as there is a definite progression of the disease. Finally, not examined during clinical investigation but has been discovered during research is that patients with NEC have a severely damaged gut barrier (Martin and Walker 2006).

In 1978, Dr Martin Bell proposed the Bell staging criteria for the diagnosis of NEC and this is still used today. Stage I, often called suspected NEC, refers to the patient displaying the mildest symptoms of NEC. Typical symptoms would include, temperature instability, lethargy, apnoea, and bradycardia. Furthermore, the infant may feed poorly, vomit, present a mildly distended abdomen, or pass stool with blood. Stage II infants display the classical signs of pneumatosis intestinalis and are proven NEC

cases. Typical symptoms seen in these patients are marked abdominal distension and persistent blood in the stool. The final criteria is Stage III, and is classified by showing most or all symptoms from the previous two stages. These infants deteriorate quickly, with a reduction of vital signs, septic shock, and gastrointestinal bleeding. Ultimately, all of these issues lead to necrosis of the bowel giving the disease it's name. Once necrosis has occurred surgical intervention is the only option for these infants (Gregory *et al.* 2011). A flow diagram of the treatments for NEC is shown in Figure 5.

Typically, the disease develops during the first two weeks of life, but still results in significant mortalities and morbidities. The causative factors, shown in Figure 4, provide some insight into the pathology of this neonatal disease.

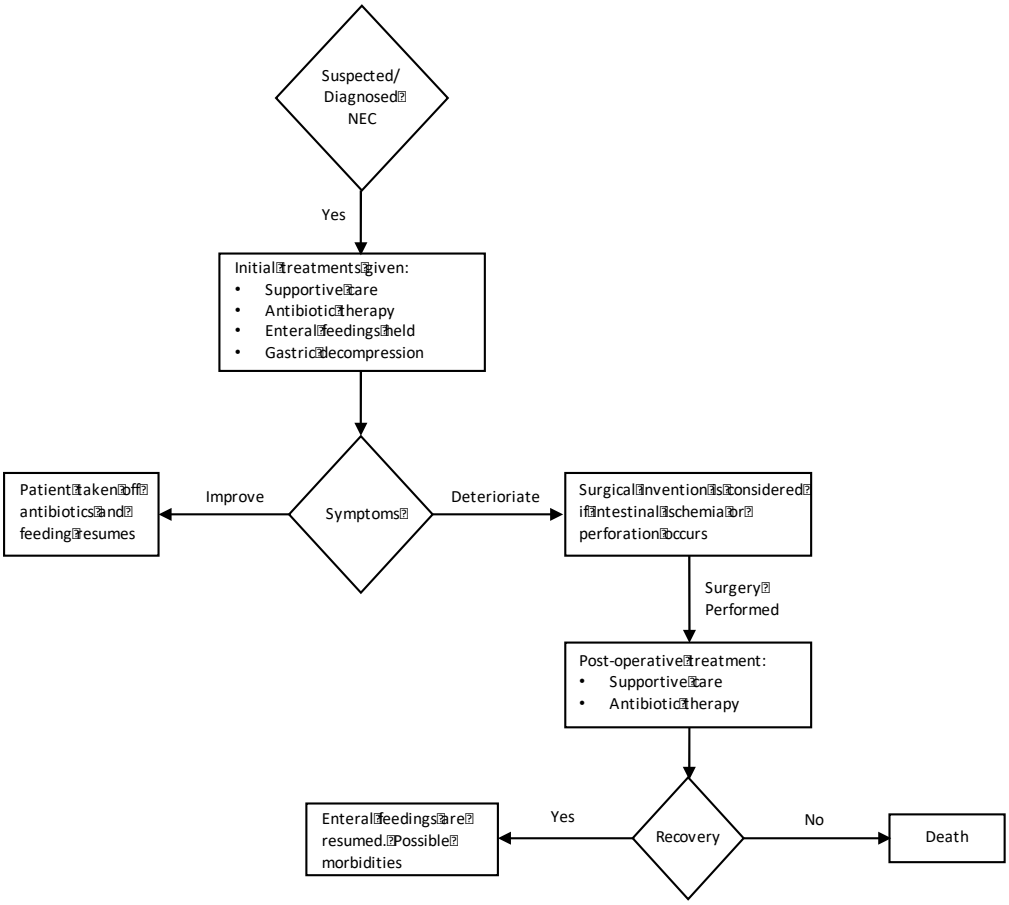

Figure 5. A flow diagram to illustrate the treatment path of necrotising enterocolitis (NEC). The above figure visualizes the clinical treatment for NEC.

1.3.5.1 Risk Factors for NEC

A number of risk factors have been identified for the development of NEC. The first and primary cause of NEC is prematurity, this then results in other risk factors such as low-

birth weight, enteral feeding, blood transfusion, and sepsis. In addition, research indicates that breast milk and probiotics may be protective against this disease, therefore removing these from the care of the infant can act as a risk factor (Lu *et al.* 2017).

Since the early 90s it has been shown that formula feeding is a risk factor for the development of NEC, more specifically cases of the disease were increased six fold in infants exclusively formula fed (Lucas and Cole 1990). Since then, a substantial amount of research has been targeted at this area, as it is a relatively cost-effective way of reducing the incidence of NEC.

In a randomized control trial of 243 infants it was found that the feeding of maternal milk resulted in fewer episodes of late onset sepsis (LOS) and or NEC. Furthermore, with these infants spending less time hospitalized, this resulted in an overall reduction in infection. However, this study found that there was no decrease in NEC between formula feeding and donor breast milk (Schanler *et al.* 2005).

More recently in 2017, Autran *et al* found that infants exposed to breast milk deficient in select oligosaccharides were more likely to develop NEC at Bell's stages II or III. Furthermore, they found this correlation to be stronger when one specific oligosaccharide was missing from the breast milk (Autran *et al.* 2017). In this study, rats fed by their mothers on formulas containing human milk oligosaccharides (HMO) displayed the lowest pathology scores, compared to exclusively formula fed where the pathology scores were the highest (Autran *et al.* 2016).

Moreover, it was found that this "protective effect" of human breast milk could work in a dose dependent fashion. In detail, the likelihood of NEC or death, as a result of NEC, was decreased by a factor of 0.83 for every 10% increase in the total proportion of human breast milk received (Meinzen-Derr *et al.* 2009).

1.3.5.2 Treatments for NEC

Due to a lack of a specific causal factor and a lack of conclusive data from studies into prospective treatments, there is a whole spectrum of treatments that are used across different countries and NICUs (Liem *et al.* 2010; Wójkowska-Mach *et al.* 2014). More specifically, several studies have shown that there is no consensus on the types,

combinations and duration of antibiotics given both pre- and post-operatively to NEC patients (Downard *et al.* 2012; Shah and Sinn 2012; Blackwood *et al.* 2017).

A systematic review by Downward *et al* in 2012 found a number of interesting outcomes for treatments of NEC. Firstly, probiotics were found to significantly reduce the incidence of severe NEC and NEC associated mortality was significantly lower in the groups receiving probiotics. Secondly, formula feeding was associated with significant growth in preterm infants, however, it was associated with a higher incidence of NEC and feeding intolerance. Thirdly, a delay in enteral feeding was found to have no significant reduction in the incidence of NEC. Finally, there was a lack of evidence to infer if antibiotics given after surgical intervention reduced the rate of reoccurrence (Downard *et al.* 2012).

Despite this lack of consensus on the efficacy of treatments, there is a standard procedure used when NEC is present, shown in Figure 5. For Stage I NEC, when the disease is suspected, enteral feeds are stopped and parenteral feeds initiated, nasogastric decompression performed, and broad-spectrum antibiotics administered in the short term. During Stage II, when the disease is confirmed, total parenteral nutrition is continued, antibiotics are now administered for longer periods, and surgical consultation begins. Finally, during the advanced stages, Stage III, total parenteral nutrition is continued, fluid resuscitation is given, inotropic and ventilator support, surgery is confirmed and performed.

1.3.6 Changes in the Preterm Gut Microbiome as a Result of NEC

Links to the bacterial content of the gut and neonatal diseases, such as NEC, have been mentioned in previous sections. However, in this section more detail will be discussed on the role of the gut microbiome in the development and progression of NEC.

A retrospective cohort analysis by Cotton *et al* 2009 showed that empirical antibiotic therapy is associated with an increased risk of NEC and or death. In addition, incidence of Bell's Stage III NEC, or surgical NEC, was higher (54%) than medical NEC (46%) in infants treated with antibiotics in the first three days of life. The majority (83%) of their cohort were prescribed a combination of ampicillin and gentamicin (Cotten *et al.* 2009). In addition, a study by Greenwood *et al* in 2014 showed that the use of early

antibiotics in preterm infants increased the incidence of NEC, sepsis, and or death (Greenwood *et al.* 2014). However, it must be considered that the infants diagnosed with the more severe cases of NEC will have had a greater exposure to antibiotics. Therefore, a strict cause and effect relationship cannot be determined.

Culture-dependent techniques have shown a marked difference between the microbial communities of preterm infants with and without NEC. The NEC infants contained more coagulase negative *Staphylococci* and less *Enterococcus faecalis* (Stewart *et al.* 2012). A systematic review showed that NEC is preceded by an increase in the relative abundance of *Proteobacteria* and decreased relative abundance of *Firmicutes* and *Bacteroides* (Pammi *et al.* 2017).

Despite research indicating a link between the microbial content of the gut and NEC, there has been research demonstrating a lack of difference between control and affected individuals. For instance, the bacterial load of stool taken from NEC patients was not significantly different from that of preterm controls (Abdulkadir *et al.* 2016b). Furthermore, the gut microbiome prior to NEC did not influence the severity of disease progression (Barron *et al.* 2017).

This and other studies further demonstrate the multifactorial nature of the disease. Therefore, further research in this area is needed in order to develop effective treatments. However, the gut microbiome of the preterm infant has been linked to other diseases associated with this patient group.

1.3.7 The Role of the Gut Microbiome in Diseases Presented in Preterm Infants

The gut microbiome of preterm infants is not only associated with NEC, but it also has been linked to other common preterm diseases. The first of these being sepsis, comprising of both late onset sepsis (LOS) and early onset (EOS). LOS is defined as such for occurring at greater than 3 days of life and a positive blood culture indicating infection by a pathogenic organism is required (Stoll *et al.* 2002). Similar to NEC, sepsis is one of the most common causes of neonatal morbidity and mortality in preterms (Stoll *et al.* 2004).

The bacteria most associated with LOS are Gram-negative organisms. In contrast, Gram-positive organisms acquired during the first days of life and coagulase negative *Staphylococcus* are the most prominent genera in infants who contract a Gram-positive LOS infection (Stoll *et al.* 2002). Furthermore, *Enterobacter* and *Staphylococcus* species have been associated with NEC and sepsis, respectively (Stewart *et al.* 2012). Later research found that in VLBW infants with sepsis had higher levels of the *Enterobacteriaceae* family, *Proteobacteria* phylum, and lower levels of *Bifidobacterium* species. Furthermore, principal coordinate analysis (PCoA) showed differences between the microbiome of infants with sepsis and unaffected infants (Collado *et al.* 2015).

On the other hand, other research has shown that it is an overall dysbiosis in the microbiome of preterm infants that is responsible for sepsis rather than individual organisms (Mai *et al.* 2013). In addition, an overall reduction in bacterial diversity has been linked to sepsis (Madan *et al.* 2012).

1.3.8 Summary

Necrotising enterocolitis is primarily a disease of prematurity, which still remains a significant cause of mortality and morbidity in the infant community. A significant amount of research has begun to unravel the risk factors, causes and treatments for this disease. However, there is still much more research needed until the disease can be controlled and cured effectively.

Gastrointestinal diseases are not isolated to the preterm infant population. As previously discussed, IBD is a common disease in adults and is known to be linked to changes in the gut microbial community. Moreover, like preterm gastrointestinal diseases, several factors have been found to influence the development of the disease. A key example of this is gut proteases. However, the total protease activity of the preterm gut has yet to be determined.

1.4 Proteases

Proteases are critical for homeostasis (Clausen *et al.* 2011) and make up approximately 2% of the entire human genome (Turk 2006). For example, proteases are involved in both protective and regulatory functions. For instance, they are protective when they degrade potentially problematic polypeptides and are regulatory when they activate other proteins via cleavage (Page and Di Cera 2008). These processes include development, coagulation, cell death, inflammation and immunity (Turk 2006). Proteases are also referred to in the literature as proteinases, as they facilitate the cleavage of peptide bonds during the breakdown of proteins. Despite this universal function, proteases have evolved into several different families as a result of specific target sites. The families of proteases are as follows: threonine, aspartate, serine, cysteine and metalloproteases (Puente *et al.* 2005). As the name suggests threonine, aspartate, serine and cysteine proteases contain these amino acids in their active sites. However, metalloproteases have a metal ion in their active site.

It was the research of Linderstrom-Lang that changed the historical view of proteases from solely degradative enzymes, only present in the breakdown and removal of proteins from cellular systems, to the concept of 'limited proteolysis'. This theory demonstrated that proteases are key components of pathways such as intestinal epithelial cell signalling. As a result, it was shown that proteases can have specificity in addition to non-specific degradative activities (Schellman and Schellman 1997).

The majority of proteases are synthesized in an inactive form, referred to as a zymogen. In order for them to become active, they need to undergo proteolytic cleavage; this is one of the key mechanisms by which protease activity is controlled. A key example is serine proteases, the simplest zymogen cascade involves two consecutive proteolysis reactions, with the zymogen being the substrate for an already active protease (Neurath and Walsh 1976). This process is known as a protease cascade, typified by a zymogen being converted to an active protease by another specific protease, this process is irreversible and results in an increase in protease potential and signal activation. Nutrient digestion relies on a protease cascade. Pancreatic trypsinogen (zymogen), is released into the small intestinal lumen and is activated to the serine protease trypsin by enteropeptidase, another serine protease spanning the membrane of intestinal epithelial

cells (IECs) in the brush border (Hermon-Taylor *et al.* 1977). As a result, trypsin activates pancreatic chymotrysinogen, procarboxypeptidases, proelastases, and prolipases (Neurath and Walsh 1976). If these protease cascades are prematurely activated it can lead to numerous diseases. For example, if trypsin becomes active in the pancreas or pancreatic ducts, this can leads to uncontrolled proteolysis contributing to diseases such as pancreatitis and cystic fibrosis (Truninger *et al.* 2001). Often protease cascades do not act alone and several protease cascades overlap to form protease networks, such as during cell migration and extra cellular matrix (ECM) degradation (Turk 2006). During ECM degradation the matrix metalloprotease (MMP) zymogens are activated by serine proteases from the plasminogen cascade (Netzel-Arnett *et al.* 2002), as demonstrated in Figure 6.

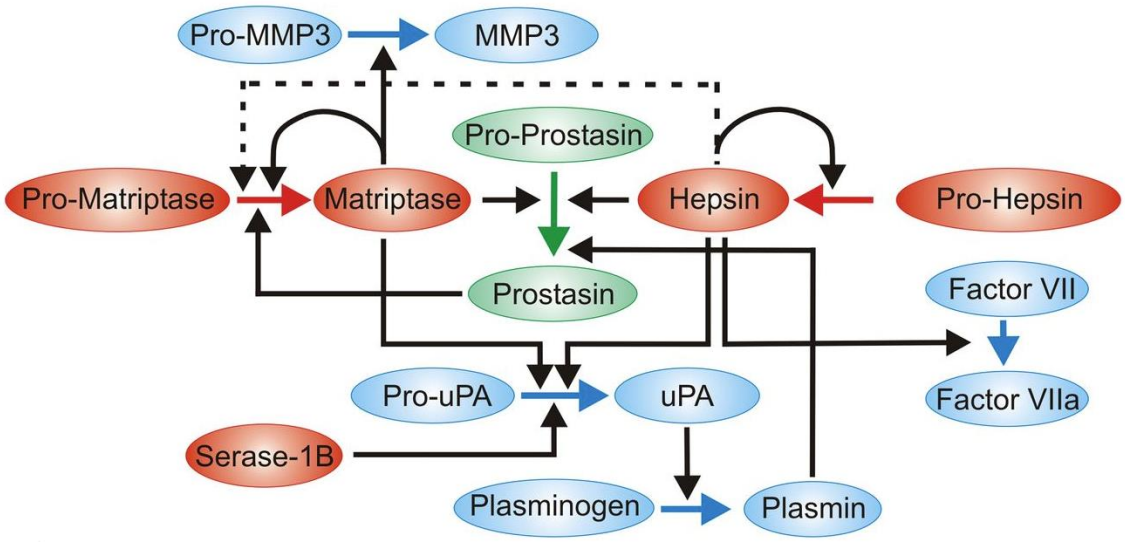

Figure 6. An Example of Proteases and their Inactive Zymogens. Figure was reproduced from (Antalis *et al.* 2010). Pathways in which membrane-anchored serine proteases have been shown to activate, or be activated by, serine proteases *in vitro* and *in vivo*. Proteases are colour coded according to membrane localization sequences: red, type II transmembrane serine proteases (TTSPs); green, GPI-anchored proteases; blue, other secreted proteases. Lines indicate activation cleavages and loops indicate auto-activation. The broken line indicates that hepsin is a weak activator of the matriptase zymogen. Membrane-anchored serine proteases intersect the coagulation cascade (Factor VII activation), fibrinolysis (pro-uPA activation) and metalloprotease pathways (pro-MMP-3 activation). Not shown is the activation of trypsinogen by enteropeptidase.

A widely studied family of serine proteases are the high temperature requirement A (HTRA) proteases. They are widely conserved in both single and multicellular organisms, for example *E. coli* are known to express three HTRAs and humans four (Clausen *et al.* 2002; Page and Di Cera 2008; Huesgen *et al.* 2009). In addition, they are known to participate in a wide variety of cellular processes such as bacterial virulence,

maintenance of photosynthesis components, proliferation, cell migration and fate (Clausen *et al.* 2002; Chien *et al.* 2009; Huesgen *et al.* 2009). The most well-known of the human HTRAs are, HTRA1 and HTRA2. It has been shown that concentrations of HTRA1 are increased in diseases such as arthritis, which could be contributing towards the degradation of cartilage as well as inflammation (Milner *et al.* 2008).

Faecal supernatants from healthy patients contain a limited quantity or serine proteases, originating from different sources such as digestive enzymes, inflammatory cells and microbiota (Róka *et al.* 2007b).The area of the human body known to contain the highest levels of proteases, both endogenous and exogenous, is the GI tract (Antalis *et al.* 2007). As mentioned previously, proteases are critical for the breakdown of peptide bonds during food digestion and other metabolic processes. However, it has also been shown that proteases are involved in much more subtle processes of regulation, such as blood clotting (Macfarlane 1964). Fifty decades on from that research we now know proteases to be involved in an array of vital functions such as, cell-cycle progression, cell proliferation and cell death, DNA replication, tissue remodelling, coagulation, wound healing and the immune response (Turk 2006). As will be discussed in more detail in another section, the activity of these proteases is tightly regulated to prevent disastrous consequences.

Matrix metalloproteases (MMPs) belong to the endopeptidases group, specifically they are responsible for the breaking of peptide bonds within the protein rather than amino acids near the terminal ends. One of their main functions is to breakdown the ECM (Baugh *et al.* 1999; Stallmach *et al.* 2000; Pender *et al.* 2003; Lubetzky *et al.* 2010). The main inhibitors of these proteases are the tissue inhibitors of metalloproteinases or TIMPs. Moreover, they are produced by the same cell types who produce MMPs and their main role is to regulate or maintain their proteolytic function (Visse and Nagase 2003).

There have been many studies demonstrating a role for MMPs in disease. For example, Medina *et al* in 2006 found that MMPs are upregulated in NEC (Medina and Radomski 2006). Prior to this Bister *et al* demonstrated that MMP-1 was found in the epithelial cells of NEC samples, and showed that alongside MMP-1, other MMPs played key roles in the tissue destruction in NEC (Bister *et al.* 2005). More recently, it has been found that under stressed conditions Caco-2 cells are known to express the active form

of MMP-2, further implicating the role of MMPs in the pathogenesis of NEC (Bein *et al.* 2015). Interestingly, it has been found that MMPs, specifically MMP-2 and MMP-9, are a component of human breast milk (Lubetzky *et al.* 2010).

1.4.1 Proteases in the Gastrointestinal Tract

There are a number of proteases present in the GI tract, they can be luminal, circulating, secreted, intracellular, intramembrane and pericellular. For example, GI function is dependent on pericellular proteolysis, this is a result of the continual exposure of intestinal epithelial cells (IECs), both at the apical and basolateral surface, to proteases from different sources (Medina and Radomski 2006). In contrast, proteases secreted by IECs are known to regulate their environment, such as remodelling the extracellular matrix (ECM) (Medina and Radomski 2006). In a later section, the role of proteases in GI diseases will be discussed, here the types and functions will be detailed.

One of the main functions of proteases in the GI tract is as signalling molecules, acting on the autocrine, paracrine, and endocrine systems, in multicomponent pathways. Protease signalling is a highly effective form of signalling as it is, in most cases, irreversible. Examples of these pathways include growth factor activation, proprotein maturation, enzyme activation, shedding of cell surface receptors, and ECM degradation and turnover.

Gastrointestinal physiology relies heavily on proteolytic substrates including: epidermal growth factors (EGFs), transforming growth factor β (TGF-β), insulin-like growth factors (IGFs), fibroblast growth factors (FGFs), trefoil factors (TFF), colony-stimulating factors (CSFs), hepatocyte growth factor (HGF), platelet-derived growth factor (PDGF), tumour necrosis factor (TNF) family members, ILs and interferons. An example of this is proteases belonging to the ADAM (a disintegrin and metalloproteinase domain) family of proteases. Firstly, the ADAMs are activated by G-protein coupled receptors, who in turn cleave membrane-tethered growth factors, which then activate the target, epidermal growth factor receptor (Gschwind *et al.* 2001; Ohtsu *et al.* 2006).

The regulation of many GI processes are controlled by G-protein coupled receptors, known as proteinase activated receptors (PARs) (Vergnolle 2000; Kawabata 2003;

MacNaughton 2005; Steinhoff *et al.* 2005). PARs become active as a result of proteolytic cleavage, as dictated by being protease activated, of the amino-terminal domains. As a result, a new amino terminus becomes exposed and acts as a tethered ligand that is able to bind and therefore activate the receptor (Coughlin 1999). Serine proteases are known to activate PARs and are found in large quantities in the intestinal lumen as a result of secretion by different mucosal cell types (MacNaughton 2005). PAR1 and PAR3 are activated by thrombin (Nakanishi-Matsui *et al.* 2000). PAR2 is activated by numerous serine proteases such as trypsin and mast cell trypstase (Cottrell *et al.* 2003; Cottrell *et al.* 2004). Other proteases act upon PARs in a different manner. For example, elastase and chymase cleave downstream of the amino-terminal placed ligand, therefore rendering the PAR signalling inactive (Déry *et al.* 1998; Dulon *et al.* 2003; Dulon *et al.* 2005).

PAR1 and PAR2 activation leads to alterations in the functions of smooth muscle in the colon and small intestines. As yet functions of PAR3 and PAR4 have yet to be elucidated in the gut (Vergnolle 2005). Finally, PARs are known to have multiple implications in intestinal pathologies, but these will be discussed in a later section.

An integral part of the GI tract is the mucosal barrier and research has shown that key functions of this vital component are regulated by proteases. Investigations in this area has found that there is an increase in epithelial membrane permeability as a consequence of trypsin injections, a known activator of PAR2 (Cenac *et al.* 2002). Furthermore, tight junction permeability has been found to be altered by the serine protease zonulin (Fasano *et al.* 2000; Wang *et al.* 2000). It is believed that this increase in intestinal zonulin is a result of bacterial contamination within the GI tract and luminal exposure to gliadin (El Asmar *et al.* 2002; Clemente *et al.* 2003; Drago *et al.* 2006). Therefore, regulation of proteases in the GI tract is critical in maintaining health and homeostasis.

1.4.1.1 Protease Regulation in the Gastrointestinal Tract

Proteases system wide and in the GI tract are regulated by numerous methods. As mentioned previously, the targeted activation of zymogens is one method, but there are also two others, compartmentalization and termination by inhibitors. Enterocytes display both spatial and temporal compartmentalization. Proteases produced by mast

cells, such as trypstase, chymase and granzyme B, are sequestered within intracellular granules and are released in response to inflammation (Jacob *et al.* 2005). This is an example of spatial compartmentalization, whereas temporal compartmentalization has already been discussed via the activation of proteases during cascades. Proteases can become localised in the pericellular space by direct and indirect mechanisms. An example of direct localisation is the tethering of MMP-14, matriptase, and dipeptidyl peptidase 4, to the cell plasma membrane by a membrane spanning domain (Hooper *et al.* 2001; Lorey *et al.* 2002; Netzel-Arnett *et al.* 2003; Medina and Radomski 2006). In contrast, indirect localisation would be the sequestering of proteases in granules or vesicles, as previously detailed.

The activity of many of the proteases in the GI tract are controlled by specific inhibitors, several examples are detailed as follows. As mentioned previously MMPs are regulated by TIMPs. Secondly, there are serpins, a large family of serine protease inhibitors that target different stages of protease cascades causing irreversible inhibition (Antalis and Lawrence 2004). Furthermore, in certain cascades each step can be inhibited by specific substrates. Inhibition of protease cascades is not a simple process as serpins, serine protease inhibitors, can have their inhibitory function removed by MMPs. After the proteases have been neutralized by their specific inhibitors, the resulting complex is rapidly cleared by specific receptors, acting as scavengers. An example of this receptor is the family of low density lipoprotein (LDL) receptors (Herz and Strickland 2001).

1.4.2 Bacterial Proteases

As discussed previously, host proteases contribute a significant proportion of proteolytic activity that occurs in the GI tract. However, combined with the knowledge of the microbial content of the GI tract, it is not surprising that proteases of bacterial origin also contribute to the total proteolytic activity of the intestines. These bacteria are known to contribute serine, cysteine and matrix metalloproteases to the “proteolytic broth” (Macfarlane *et al.* 1988; Gibson *et al.* 1989; Róka *et al.* 2007b).

Bacterial proteases can exist in many forms. They can be excreted by the bacterium, remain attached to the cell surface, or they can be embedded within the bacterial membrane. Furthermore, they take part in numerous biological processes, such as post-translational regulation of gene expression. Specifically, in the processing and

maturation of various surface-associated proteins (Laskowska *et al.* 1996; Gottesman *et al.* 1997). For example, HtrA proteases have been shown to have housekeeping functions such chaperoning proteins and degrading misfolded proteins (Spiess *et al.* 1999). In addition, the HtrA serine-like proteases help bacteria to survive environmental stresses such as, elevated temperature, oxidative and osmotic stresses (Pallen and Wren 1997; Clausen *et al.* 2002).

One role of bacterial proteases is to interact with integral and peripheral proteins in the intestinal cell wall, this normally results in inflammation and cytotoxicity (Coleman *et al.* 2013; Sumitomo *et al.* 2013; Golovkine *et al.* 2014). Another role for bacterial proteases is in quality control, for example *E. coli* DegP, DegQ and DegS enzymes, located on the cell envelope (Spiess *et al.* 1999). It is believed that this quality control mechanism is related to bacterial pathogenesis (Ingmer and Brøndsted 2009; Huston 2010). HTR proteins can also be found in other bacterial species such as *Streptococcus mutans*, such as HTRA. These proteases are important for biofilm formation, as they are needed during the biogenesis of extracellular proteins, therefore it can be said that HtrA is involved in bacterial pathogenesis by modulating biofilm formation (Biswas and Biswas 2005).

One of the main groups of bacterial proteases are members of the omptin family. They are outer membrane proteases known to direct the pathogenicity of Gram negative organisms such as *E. coli* (OmpT) (Grodberg and Dunn 1988). These proteases are most often encoded on plasmids or prophages and so are likely to be transferred via horizontal gene transfer (Hritonenko and Stathopoulos 2007). Furthermore, it is known that omptins interact with LPS as part of their proteolytic activity (Kramer *et al.* 2002; Kukkonen and Korhonen 2004). One of the key reasons omptins are so unique, is their catalytic activity. They are technically classed as aspartate proteases by the MEROPs database but also share a similar catalytic structure to that of serine proteases (Vandeputte-Rutten *et al.* 2001). As a result of the unique catalytic site, OmpT is not inhibited by any of the normal protease inhibitors, such as PMSF a known serine protease inhibitor. However, a recent study by Brannon *et al* in 2015 found that Aprotinin, typically a serine protease inhibitor such as trypsin, was able to reduce the proteolytic activity of OmpT (Brannon *et al.* 2015).

E. coli is not the only pathogen shown to release proteases that aid in their pathogenicity. *Campylobacter jejuni* (*C. jejuni*) has been a key player in the protease assisted virulence. Firstly, it is known to invade human intestinal epithelial cells, this occurs either via a transcellular or paracellular pathway (Boehm *et al.* 2012; O Cróinín and Backert 2012). More specifically, colonization by *C. jejuni* begins via a paracellular breach of the intestinal cell wall and binding to fibronectin on the basolateral surface of the membrane (Backert and Hofreuter 2013). In addition, it can occur via the breakdown of E-cadherin, present in adheren junctions, by proteases (Boehm *et al.* 2012).

Furthermore, this method of invasion has been shown by another pathogenic species, *Helicobacter pylori*. This bacterium releases proteases that compromise the integrity of adheren junctions, therefore allowing *H. pylori* access to the basolateral side of the gastric membrane (Hoy *et al.* 2010). A more recent study by Elmi *et al* found an outer membrane vesicle (OMV) of *C. jejuni* to contain three serine-like proteases. In addition, these proteases increased the release of lactate dehydrogenase, a measure of cytotoxicity, in T84 cells; caused the breakdown of key tight junction proteins; and enhanced the invasion mechanism of *C. jejuni* into T84 cells (Elmi *et al.* 2015).

In addition, *Chlamydia* species have been shown to secrete proteases to degrade intracellular membranes, which creates the necessary compartment for bacterial survival and growth (Derré 2015). Furthermore, there are other species of bacteria showing protease activity. Proteases from *Bacteroides fragilis* and *Porphyromonas gingivalis* have been shown to breakdown E-cadherin, an integral protein of intestinal wall adheren junctions (Wu *et al.* 1998; Katz *et al.* 2000; Remacle *et al.* 2014).

It has been shown that samples rich in *Clostridium* species have increased proteolytic activity in human faecal samples (Woodmansey *et al.* 2004). A mouse model has shown that a reduction in microbial density and protease activity occurs after the administration of oral antibiotics (Róka *et al.* 2007a). Interestingly, it is worth mentioning that proteolytic activity from bacteria is independent of inflammation in the intestine (Pruteanu *et al.* 2011).

1.4.3 Protease Activity and Disease

There have been several studies demonstrating elevated levels of protease activity in disease, such as inflammatory bowel disease (IBD) and irritable bowel syndrome (IBS) (Bustos *et al.* 1998; Dunlop *et al.* 2006; Cenac *et al.* 2007; Róka *et al.* 2007b; Gecse *et al.* 2008; Shulman *et al.* 2008; Annaházi *et al.* 2009), Figure 7. The pathology behind these diseases occurs as a result of excessive proteolysis by proteases. Research is currently undergoing as to whether this is a result of faulty host proteolytic systems or of a bacterial origin as virulence factors or symbiosis. In addition, bacteria in the GI tract could also contribute to disease via "pathogen host mimicry".

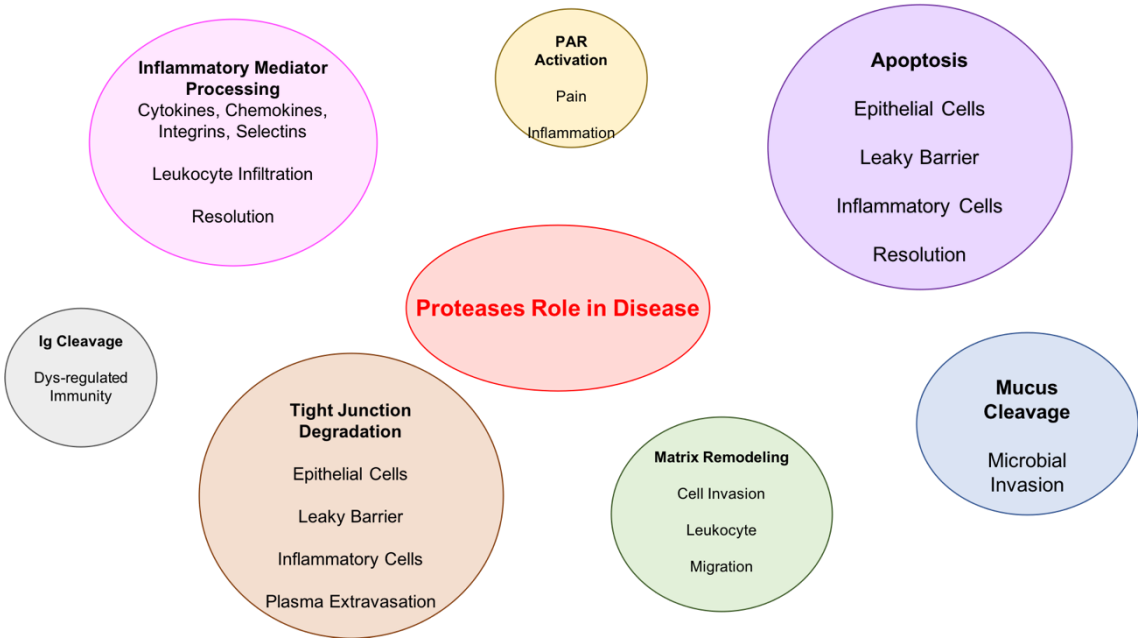

Figure 7. Proteases and their role in disease. Each bubble contains a different way protease exert their effect in the development and progression of GI diseases. The bold text refers to the biological function affected by the proteases and the other text are the affect components and the results. Adapted from Vergnolle (Vergnolle 2000).

For example, high levels of neutrophil mediators such as elastase and calprotectin were found in the faecal samples of ulcerative colitis (UC) patients (Gecse *et al.* 2008). To add, cathepsin-G (Cat-G), a serine protease produced by neutrophils, binds structures on intestinal epithelial cells. These structures are called proteinase-activated receptors (PARs). PARs are a family of transmembrane G-protein-coupled receptors, which are activated by serine proteases cleaving their N-terminal domain (Hollenberg and Compton 2002; Cenac *et al.* 2007). Specifically, Cat-G binds to PAR4 activating

important inflammatory processes and neutrophil functions (Sabri *et al.* 2003; Shimoda *et al.* 2007).

A key study by Carroll *et al* in 2013 clearly demonstrated a distinct difference, detailed next, between bacterial communities in samples demonstrating high and low faecal protease activity. In addition, a wide variety of protease activity was seen in the samples, all patients with IBS. Furthermore, a significant decrease in the number of observed Operational Taxonomic Units (OTUs) and the Shannon index of diversity was found in the samples with high faecal protease activity. *Lactobacillales*, *Lachnospiracae* and *Streptococcaceae* groups were positively associated with faecal protease activity. On the other hand, the *Ruminococcaceae* family was negatively correlated with faecal protease activity. Lastly, they found a significant correlation of *F. prausnitzii* with faecal protease activity (Carroll *et al.* 2013). From this research, it can be thought that individuals with a high faecal protease activity have lower numbers and diversity of bacterial species in their intestines and harbour a microbiome that is distinct from individuals with low faecal protease activity.

Disease is not solely caused by bacterial or host proteases alone, there is a certain degree of interplay between the proteases of different origins. As will be mentioned throughout this thesis, there is little evidence in this area of protease research, due to the difficulties in determining the origin of the proteases present in the sample. However, there has been interesting research indicating at complex and detailed interactions between these cross-kingdom proteases. MMP-7, known to be excreted by Paneth cells, catalyse the activation of cryptdins, a family of antimicrobial peptides (Weeks *et al.* 2006). It was found that in MMP-7 deficient mice, enteric pathogens such as *E. coli* and *Salmonella typhimurium* colonise in greater numbers in these individuals, demonstrating the role of MMP-7 in the activation of antimicrobial peptides (Wilson *et al.* 1999). Also, in mice, *Citrobacter rodentium*, a relative of the human pathogenic *E. coli*, increases the release of serine proteases and granzyme A. However, in PAR2 deficient mice the pathogenic effect of *C. rodentium* is greatly reduced (Hansen *et al.* 2005).

It is known that bacteria, specifically enteric pathogens, secrete specific proteases leading to either increased gut stimulation and inflammation, or by inhibiting host immunity. A key example of this is *Bacteroides fragilis*, known to produce the

virulence factor, zinc-dependent metalloproteinase enterotoxin. This enterotoxin causes a number of changes in the host, such as rapid release of IEC proteins, cleavage of tight junction protein E-cadherin and reduction in colonic permeability, colonic epithelial cell changes, release of pro-inflammatory cytokines and colonic cell proliferation (Sears 2001). It is believed that these changes contribute to *B. fragilis* associated diarrheal disease, IBD and colorectal cancer. *B. fragilis* is not the only pathogen known to use proteases as a virulence factor. *Vibrio cholera* secrets a zinc-dependent protease with mucinase-like activity, thought to stimulate the cleavage of occludin. In contrast, the enterotoxic protease of *C. perfringens* is thought to cleave claudins, therefore affecting intestinal barrier integrity (Sears 2000).

1.4.4 The Role of Proteases in the Infant and Preterm Infant Gastrointestinal Tract.

1.4.4.1 Introduction

Protein is an essential nutrient for growth and development of infants. However, preterm infants are born requiring a greater amount of protein in order to achieve adequate or catch up growth. As a result, preterm feeding is often supplemented (Arslanoglu *et al.* 2010). Little is known about protein digestion in both full preterm infants, thus making it an important area of research in order to provide the best care for a vulnerable population.

1.4.4.2 The Source of Proteases in the Infant GI Tract.

The primary source of protein for both term and preterm infants is the mother's breast milk. Research has shown that breast milk from mothers of preterm infants contains a higher protein content, which decreases over the first 8 weeks of lactation (Klein 2002). Furthermore, a variety of proteases and their inhibitors are present in breast milk. This was shown by the presence of over 100 unique casein fragments in milk from both term and preterm mothers (Ferranti *et al.* 2004; Armaforte *et al.* 2010).

Interestingly, data has suggested that preterm milk undergoes more proteolysis than term milk (Armaforte *et al.* 2010). This study indicates at the presence of a higher level proteases in the gut of preterm infants, the origin of which could not be host or bacterial.

Several proteases have been discovered in breast milk of term infants such as, trypsin, elastase, plasmin, cathepsin D and kallikrein (Astrup and Sterndorff 1953; Fox 1981; Vetvicka *et al.* 1993; Palmer *et al.* 2006; Christensen *et al.* 2010). However, research has not yet investigated the concentrations and activities of proteases in the breast milk of mother's to preterm infants. This protease activity decreases over time in term milk (Heyndrickx 1963; Tor *et al.* 1982). The proteases present in breast milk are currently believed to initiate the digestion of protein for the infant, as the decrease of proteases in breast milk coincides with the increase of the infant's own degradative capacity.

As mentioned previously protease inhibitors are also present in human breast milk, and similarly, to other systems the inhibitors are present in order to regulate the activity of proteases. Specifically, to protect the human milk proteins from digestion. This inhibitory action was detected in term and preterm milk samples from birth up to 160 days postpartum (Tor *et al.* 1982). Interestingly for this thesis, α1-antitrypsin has been found in the faeces of term breastfed infants. Therefore, if antiproteases are present in the stool it is possible for proteases to also survive the gastrointestinal tract and be detectable in faeces (Davidson and Lönnerdal 1987). However, the persistence of proteases and or antiproteases have not been reported in the faeces of preterm infants.

1.4.4.3 The Proteases Present in the GI Tract of Infants.

Overall, key luminal proteases in the adult intestine such as trypsin, chymotrypsin, elastase, enterokinase and carboxypeptidase B; are also present in both term and preterm infants. However, they have been found at concentrations and activities lower than those in adults.

Enterokinase (also called enteropeptidase) is a protease secreted from intestinal epithelial cells in response to food stimulation (Neu 2007b). This protease is essential for proteolysis in the intestine as it is leads to the activation of trypsinogen to trypsin (Britton and Koldovsky 1989). Trypsin is then available to initiate the conversion of chymotrypsinogen to chymotrypsin, proteoelastase to elastase and procarboxypeptidase to carboxypeptidase (Dallas *et al.* 2012). Very few studies have shown that enterokinase is present at birth in both term and preterm infants. Furthermore, the enzyme was detected in the duodenal mucosa by 24 to 26 weeks gestation (Antonowicz and Lebenthal 1977). Moreover, this enzyme to be active in both preterm and full term

infants (Dallas *et al.* 2012). However, compared to that of older children, enterokinase activity was 6% and 20% in 26 to 30 week gestational age preterm infants and term infants, respectively (Antonowicz and Lebenthal 1977).

Trypsin is a ubiquitous human protease, targeted to cleave peptides at the carboxyl side of lysine and arginine (Leiros 2004). In the duodenum trypsin concentrations are less than those of adults in both full and preterm infants (BorgstrÖM *et al.* 1960). However, trypsin concentrations in the duodenum of preterm infants was less than term infants, until 2 to 4 weeks after birth (BorgstrÖM *et al.* 1960). Ultimately, a month after birth both term and preterm infants display concentrations and activity of trypsin similar to those of adults (Lebenthal and Lee 1980).

Chymotrypsin is another luminal protease known to cleave the carboxyl side of tyrosine, tryptophan and phenylalanine (Appel 1986). The concentration of chymotrypsin remains relatively stable during the first month of life in both full and preterm infants, but only reach 10 to 60% of the concentrations seen in adults (Lebenthal and Lee 1980; Kolacek *et al.* 1990). Of interest, during the first 30 days after birth there was no difference in the activity of chymotrypsin between term and preterm infants (Kolacek *et al.* 1990). Furthermore, chymotrypsin was present in the faeces of both term and preterm infants in similar values (Vendrell *et al.* 2000).

Carboxypeptidase B is a protease that complements the activity of trypsin as it cleaves arginine and lysine from the carboxy terminus and peptides (Kim *et al.* 1972). Carboxypeptidase B is also present in similar concentrations and activities in both term and preterm infant duodenal fluids at birth and 30 days of age. In addition, concentrations and activities were 10 to 25% of those of a 2 year old (Lebenthal and Lee 1980).

In summary, the only protease discussed to have less activity in preterm infants is enterokinase, the others had similar concentrations and activities. However, this was not achieved until 30 days after birth, the most critical time for growth and development. Therefore, preterm infants are likely to be less capable of digesting proteins.

1.4.4.4 Bacterial Protease Degradation in the Infant Gut.

As previously detailed, bacteria of the gut microbiome are known to produce proteases and contribute to the degradation of dietary proteins. Several key components of the gut microbiome are known to breakdown protein such as *Bacteroides*, *Propionibacterium*, *Streptococcus*, *Clostridium*, *Bacillus* and *Staphylococcus* (Macfarlane and Allison 1986). For example, intestinal bacteria degrade casein and bovine serum albumin via cell bound and extracellular proteases (Smith and Macfarlane 1997b). These proteins are first broken into peptides and then into volatile fatty acids, ammonia, dicarboxylic acids and various phenolic compounds (Smith and Macfarlane 1997b). A key observation is that the amino acids do not accumulate when these bacteria degrade protein, suggests the amino acids are quickly metabolized by the intestinal microbial community. This could be a result of some of the gut microbial community being unable to break down peptides directly and utilizes free amino acids (Whiteley 1957). Moreover, a wide variety anaerobes can ferment amino acids, such as species from the following genera *Peptostreptococcus*, *Campylobacter*, *Acidaminococcus*, *Acidaminobacter*, *Fusobacterium* and *Eubacterium* (Cato *et al.* 1983; Dürre *et al.* 1983; Stams and Hansen 1984; Buckel 1986; Nanninga *et al.* 1986; Zindel *et al.* 1988; Rogers *et al.* 1992; Smith and Macfarlane 1997a). Some bacteria can utilize both carbohydrates and proteins as an energy source, whereas others are obligate amino acid fermenters (Liepke *et al.* 2002).

Researchers have not yet determined the amount of bacterial protein degradation in the intestinal tract and colon of term and preterm infants. However, studies have provided an insight to the possible processes. For instance, the observation that *Bifidobacterium longum* subspecies *infantis*, a bacterial strain common in the intestinal tract of breast-fed infants, grows on culture medium made of pepsin-digested human milk Lf (lactoferrin) and sIgA suggests that bacterial fermentation of dietary proteins is common in breast fed infants (Brock *et al.* 1976). Moreover, a synthesized peptide called prebiotic lactoferrin-derived peptide-I (PREP-I) that is based on these peptides stimulated growth of *B. infantis* at a concentration of 1 to 10 µM, but did not stimulate four pathogenic bacterial strains (Brock *et al.* 1976). The observation that Lf and sIgA can survive intact in stools of term and preterm infants, suggests that such stimulatory peptide fragments could survive to support the growth of *B. infantis* in the colon, but also that even after

exposure to bacteria in the large intestine, some milk proteins resist degradation (Richard *et al.* 1986; Davidson and Lönnerdal 1987).

A comprehensive comparison of the protein content of ileostomy fluid with that of faeces has not been made, so it is not possible to comment further on protein degradation that occurs in the colon. Any proteolysis in the colon would likely be primarily that result of bacterial proteases. As protein-degrading bacteria are present in the colon (Windey *et al.* 2012).

1.4.4.5 Summary

In summary, there is clear evidence for a detectable protease activity level in the stool of preterm infants. However, data is needed to determine if this is the case. Yet to be discussed is the role these proteases may play in disease, specifically the common gastrointestinal disease in preterm infants, NEC. As detailed previously, excessive protease activity has been shown to occur in individuals with IBD. However, no research to date has found any link between preterm faecal protease activity and disease. But proteases are responsible for the activation of pro-inflammatory signals in the gut, a key pathology in the development of NEC is an exacerbated inflammatory response. Therefore, pro-inflammatory cytokines and protease activity could be closely linked in the developing preterm infant gut.

1.5 The Intestinal Epithelial Barrier

1.5.1 The Function of the Intestinal Epithelial Barrier

Regulation of the intestinal epithelial barrier is a result of the action of tight junctions (TJ), which are present between the cells of the single cell epithelial layer. It has been shown that they can be activated by myosin light chain (MLC) phosphorylation, causing the actinomyosin ring to contract and open the tight junction (Turner *et al.* 1997). The main function of the intestinal epithelial barrier is to regulate the passage of nutrients and ions from the intestinal lumen to the blood stream and vice versa, as shown in Figure 8.

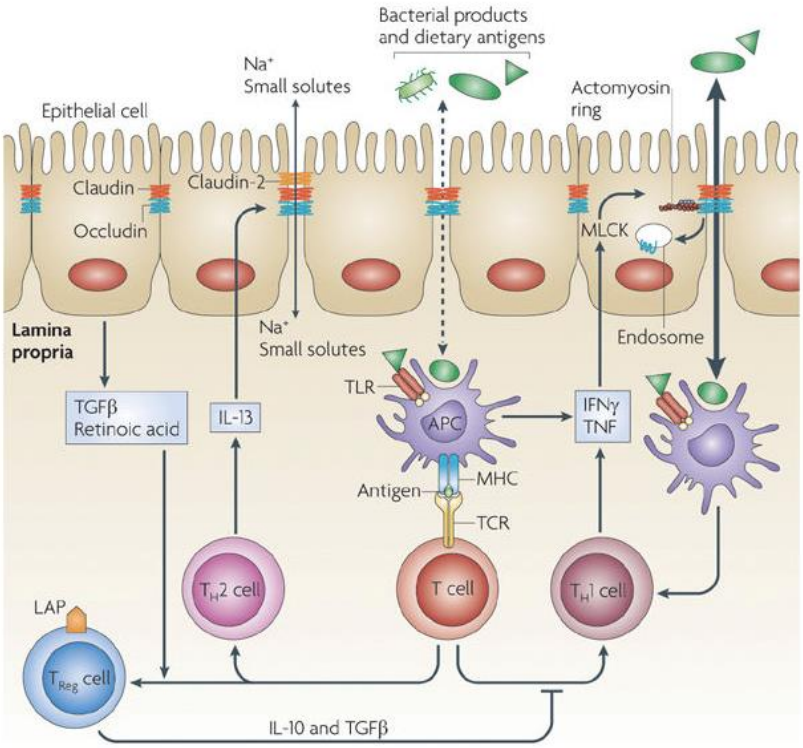

Figure 8. The Function of the Intestinal Epithelial Barrier. Image reproduced with permission from Turner 2009 (Turner 2009), permission granted on 25.2.19. Minor barrier defects allow bacterial products and dietary antigens to cross the epithelium and enter the lamina propria. This can lead to disease or homeostasis. If the foreign materials are taken up by antigen-presenting cells (APCs), such as dendritic cells, that direct the differentiation of T helper 1 (Th1) or Th2 cells, disease can develop. In this process, APCs and Th1 cells can release tumour necrosis factor (TNF) and interferon-γ (IFNγ), which signal to epithelial cells to increase flux across the tight junction leak pathway, thereby allowing further leakage of bacterial products and dietary antigens from the lumen into the lamina propria and amplifying the cycle of inflammation. This may, ultimately, culminate in established disease. Alternatively, interleukin-13 (IL-13) released by Th2 cells increases flux across small cation-selective pores, potentially contributing to ongoing disease. Conversely, homeostasis may dominate if APCs promote regulatory T (Treg) cell differentiation, which can be enhanced by epithelial cell derived transforming growth factor β (TGFβ)

and retinoic acid. The Treg cells display latency associated peptide (LAP) on their surfaces and may secrete IL-10 and TGFβ to prevent disease. Abbreviations as follows: MLCK, myosin light chain kinase; TLR, Toll-like receptor; TCR, T cell receptor.

In the intestine, the epithelia are arranged into crypts and villi. The crypts contain stem cells which regenerate every five days and the villi increase the surface area of the barrier to encourage the passage of nutrients etc. Furthermore, the epithelial barrier has a protective system that also separates it from the intestinal lumen, a thick mucus layer. This mucus layer contains mucins, glycoproteins produced from the goblet cells of the epithelial layer. The function of this barrier is to prevent direct host-microbial binding, reduce aggregation of adherent bacteria, and increase bacterial removal (Halpern and Denning 2015).

The integrity of this barrier is further increased by the apical junction complex (AJC), these are intercellular contacts consisting of membrane proteins and cytoskeletal anchor proteins, resulting in tight and adheren junctions. TJs consist of over 40 transmembrane proteins including occludin, claudins and junctional adhesion molecules. The adheren junctions include the proteins E-cadherin, and α- and β-catenin. The whole AJC is linked to the cytoskeletal scaffolding to create an F-actin ring. The AJCs "zipper" the intestinal epithelial cells together to regulate the passage of ions and small molecules. The AJC and the cytoskeletal connections can be regulated by physiologic and pathologic signals (Halpern and Denning 2015).

The key role of the intestinal epithelial barrier is to prevent bacteria and other toxins from entering the blood stream. One of the ways it does this is through the action of Toll-like receptor 11 (TLR11). TLR11 has been shown to "block" the passage of *Salmonella* through the Peyer's patches, this is in contrast to TLR5, which is referred to as a "carrier" of *Salmonella* (Shi *et al.* 2012). The role of bacteria in the control of the intestinal epithelial barrier will be discussed in a later section.

1.5.2 Disruption of the Epithelial Barrier and Disease

There are numerous factors that contribute to the disruption of the intestinal epithelial barrier. It has been shown that proinflammatory cytokines such as IFN-γ can increase intestinal epithelial permeability. Research suggests that IFN-γ can induce endocytosis of TJ proteins, therefore increasing paracellular permeability (Utech *et al.* 2005). An

earlier study showed that IFN-γ alongside tumour necrosis factor-α (TNFα) can downregulate the TJ protein occludin, as a result increasing TJ permeability (Mankertz *et al.* 2000). The action of these proinflammatory cytokines by inducing endocytosis of TJ proteins acts independently and primarily over the apoptotic effect, previously known (Bruewer *et al.* 2003).

Inflammatory bowel disease is a severe intestinal disorder affecting a significant number of people and a large amount of research into the pathophysiology of this disease has been done and is discussed in a previous section. It has been suggested that epithelial barrier dysfunction has an important role in the progression of this disease. For example, relapse in Crohn's disease patients is preceded by an increase in intestinal epithelial barrier integrity. Furthermore, a disruption on barrier integrity has been seen in patients with celiac disease (Clayburgh *et al.* 2004).

One of the key changes to epithelial integrity during disease is via the action of bacteria. There are numerous pathogens with a vast array of methods to invade the intestinal barrier.

1.5.3 The Effect of Bacteria on the Intestinal-Epithelial Barrier

The disease effect of Enterohaemorrhagic *E. coli* (EHEC) infection was unknown till it was discovered that infection causes an increase in intestinal barrier permeability. It was found that infection with this pathogen led to activation of intracellular transduction pathways, which resulted in epithelial permeability (Philpott *et al.* 1998). An *in vivo* murine model of EHEC infection was developed and provided more detail on the effect of this pathogen. This model showed that EHEC led to a redistribution of TJ proteins as a result of altered transcription of key TJ proteins such as occludin and claudin (Roxas *et al.* 2010). As previously identified TJs are opened by the phosphorylation of the MLC, therefore it would be feasible to hypothesize that bacteria may act upon the TJ in a similar way in order to enter the bloodstream. Simonovic *et al* in 2000 found that EHEC infection caused the dephosphorylation of occludin, which led to the same effect as MLC phosphorylation, by increasing TJ permeability (Simonovic *et al.* 2000).

It is known that pathologies such as Crohn's disease are linked to changes in the gut, but it has been shown that pathogens such as adherent-invasive *E. coli* (AIEC) can

contribute to the progression of this disease. Research has shown that adherent-invasive *E. coli* produce a long polar finbrae that interacts with Peyer's patch cells, increasing the effectiveness of M cell translocation, therefore contributing the lesions in Peyer's patches (Chassaing *et al.* 2011). Furthermore, AIEC cells are known to interact with other intestinal cell types, as a result increasing claudin-2 expression and TJ permeability (Denizot *et al.* 2011).

Another bacterium known to contribute to epithelial barrier breakdown is *Salmonella typhimurium*. It has been found that *S. typhimurium* causes a rapid decrease in transepithelial electrical resistance (TEER) and a rearrangement of key TJ proteins (Jepson *et al.* 1995). Later research has elicited part of the mechanism by which *S. typhimurium* causes epithelial barrier disruption. In order to interact with the barrier *S. typhimurium* extends fibrils, similar to AEIC, these protrusions activate TLR2 causing a decrease in TEER and an increase in translocation across the epithelium (Oppong *et al.* 2013).

One of the most well-known and possible researched gastrointestinal pathogens in *Campylobacter jejuni*, known to attack the gut intestinal epithelial barrier as part of it's pathogenesis. Research has shown that *C. jejuni* can inhibit absorptive cell function and alter TEER through a rearrangement of TJ proteins such as occludin (MacCallum *et al.* 2005). Further research showed that epithelial cell invasion by *C. jejuni* preceded epithelial barrier decrease, as a result *C. jejuni* barrier disruption is mediated by invasion (Wine *et al.* 2008). A very interesting piece of research showed that not only does *C. jejuni* enable translocation of itself across the intestinal epithelial barrier, but it can also enable commensal gut bacteria to travel outside of the intestinal lumen via a lipid-raft translocation mechanism (Kalischuk *et al.* 2009).

A known pathogen responsible to many cases of gastroenteritis is *Yersinia enterocolitica* and as with many other pathogens previously described elicits a diarrheal effect by intestinal barrier dysfunction. It was found that *Y. enterocolitica* caused a decrease in TEER only in the presence of live bacteria and necrosis. Furthermore, this decrease in TEER was a result of TJ protein rearrangement (Hering *et al.* 2017). Another bacterium in this genus, *Y. pseudotuberculosis* has been found to alter intestinal barrier function by disrupting the interplay between immune and epithelial cells via TLR2 stimulation (Jung *et al.* 2012). This mode of action has been discovered in

another bacterium *S. typhimirium*. The research by Jung *et al* was built upon by Meinzer *et al* who found that *Y. pseudotuberculosis* effector YopJ by disrupting the immune receptor Nod2, instead of acting as an innate immune receptor Nod2 activates caspase-1, thereby increasing levels of IL-β, leading to intestinal barrier disruption (Meinzer *et al.* 2012). Not only does *Y. pseudotuberculosis* share TLR2 disruption with *S. typhimirium*, this bacterium can also favour the translocation of commensal gut bacterium (Ragnarsson *et al.* 2008).

There are a substantial number of bacteria who can translocate the epithelial barrier as has already been discussed, but as yet to be mentioned is the subtle differences between there invasion tactics. So far, we have seen TJ protein rearrangement, membrane interactions, cell signalling events, phosphorylation methods and the ability to allow commensal bacteria to "tag along". Next there is *Shigella flexneri* who cause disease by translocating the epithelial barrier, by releasing Ipa proteins (invasins) that cause rearrangement of the cytoskeleton therefore allowing access for the bacteria to cross the epithelial barrier (Sansonetti *et al.* 1999). However, earlier research showed that *Shigella flexneri* translocated through the epithelial barrier via the basolateral surface (Mounier *et al.* 1992). This conflict shows that there is much more research needed to elucidate the exact mechanisms by which bacteria, pathogens and their toxins are able to move from the gut to the wider host environment.

1.5.4 Host Proteases and the Intestinal Epithelial Barrier

It is known that bacterial proteases are not the only proteases present in the intestinal lumen, as a significant contributor to total protease activity is host derived proteases. There are a number of host proteases known to contribute to the function of the intestinal epithelial barrier. Firstly, host matriptase, a serine protease, increases the production of clausin-2 a tight junctional protein (Buzza *et al.* 2010). Proteases released by mast cells have been found to increase intestinal epithelial barrier permeability (Overman *et al.* 2012). Furthermore, trypstase is released from mast cells during times of stress and inflammation (Camilleri *et al.* 2012).

1.5.5 The Development of Intestinal Permeability in the Premature Infant

The development of the intestine *in utero* and in premature infants has been discussed previously, 1.3.3, however, there are differences associated with the development of the epithelial barrier, seen in Table 1, that are different from general intestinal development. Firstly, development of the intestinal barrier occurs both *in utero* and postnatally. The barrier of the human intestinal monolayer forms during the first trimester. Furthermore, epithelial architecture begins to form at 8 weeks of gestation to form the crypts and villi. Two weeks later the AJC develops and tight junctions can be detected. This development occurs relatively early in gestation, however, the intestinal barrier is not considered fully mature until term, as growth factors are not present until 26 weeks of gestation.

In a continuation of the theme of this thesis, the preterm infant differs dramatically in terms of development from that of full-term infants. Studies have shown that Paneth cells are developmentally deficient in number and function in an infant born at 24 weeks. Furthermore, T-cells are shown to be recruited early to the premature gut.

One of the most prominent defects in preterm infants is an increased intestinal permeability. As a result, this leads to an increased invasion of bacteria and toxins causing inflammation and potentially intestinal injury. Studies have shown that commensal bacteria and probiotics decreases this intestinal permeability, progressing the preterm gut towards maturity (Halpern and Denning 2015).

Table 3. The Effects of Immaturity of the Intestinal Epithelial Barrier in Preterm Infants.
Reproduced from Halpern and Denning 2015 with permission (Halpern and Denning 2015).

Intestinal Barrier Component	Time of Maturation	Effect of Immaturity in the Preterm Infant
Epithelial Apical Junctional Complex (AJC)	Mature structure at 12 wks gestation (in utero)Mature function at term	Increased intestinal permeabilityImmature absorptive capability Immature secretory capability
Paneth Cells	Detectable at 12 wks gestation Secretory capability at 13–20 wks gestation	Decreased number Decreased secretory capability (Lack of antimicrobial peptides required to regulate intestinal colonization)
Mucin (Goblet Cells)	Term	Immature mucus layer allows bacteria to contact intestinal epithelia(Lack of physical protective barrier)
Intestinal epithelial lymphyocyte (IEL)	TCRγδ subset recruited early (24 wks gestation)	Early recruitment may be important to promote immature barrier function (Important to prevent bacterial translocation, promote TJ function, regulate inflammation, and promote epithelial repair)

1.5.6 Summary

The intestinal epithelial barrier is integral in maintaining gut homeostasis. This is achieved by maintaining a selectively permeable barrier. Proteases, bacteria and inflammation have been shown to affect the permeability of this barrier. However, their effect on the gut barrier permeability in preterm infants have yet to be investigated. Therefore, by analysing the levels of proteases and inflammation in the gut of preterm infant's possible barrier degradation can be suggested. Moreover, by comparing the bacteria present in different sites within the infant the degradation of the intestinal epithelial barrier can be investigated.

1.6 The Immune System of the Gut

1.6.1 The Healthy Gut Immune System

The immune system of the gut is unique as it needs to avoid triggering a large immune response to food antigens and commensal microbiota, whilst still being able to detect and remove pathogenic bacteria. This level of specificity is known as oral tolerance. The gut immune system has evolved to be able to determine the difference between antigens derived from food and those from other foreign sources.

The main component of the gut immune system is the mucosal immune system, concentrated at the Peyer's patches. Also located at the Peyer's patches are the microfold or M cells. These cells actually lack mucosal layer and can therefore interact directly with the luminal contents. The M cells take part in transcytosis, taking up luminal contents and passing them through the cell to the basal layer where they can be used for antigen presenting during the adaptive immune response. As a result, M cells are often targeted by pathogens. In addition to the mucosal immunity, lymphocytes and plasma cells are spread along the lamina propia of the gut wall. Once the lymphocytes have become activated by luminal contents, they drain out of the intestine into the bloodstream (Charles A Janeway *et al.* 2001).

1.6.2 Changes during Disease and Prematurity

As mentioned previously, IBD and IBS and other gut diseases are multifactorial with numerous pathways contributing to pathophysiology. In this section, detail on how the immune system changes during well-known gut disorders in both adults and preterm infants will be discussed.

Studies in patients with IBS have shown that the numbers of immunocompetent cells are increased in the intestinal mucosa. For example, there are increased numbers of T cells, intraepithelial lymphocytes, and IL-2 expressing receptor cells. Furthermore, in 50% of patients the numbers of neutrophil and mucosal mast cells were increased. Patients experiencing Crohn's disease have been found to harbour abnormal natural killer T cells. Genetic defects in IBD patients have been found to affect Paneth cell function. In addition, it has been found that a loss of tolerance to the commensal

microbial community is seen in IBD pathogenesis. There are numerous other studies that demonstrate the aberrant immune system in IBD and other gut diseases, again adding further evidence to the multifactorial nature of these diseases (Cader and Kaser 2013).

The immune cell composition of the full term infant gut, and as a result the preterm infant gut, is not fully understood, as the intestinal tissue is difficult to obtain for research (Battersby and Gibbons 2013). As previously described, the adult gut immunological tissue is distributed into three main compartments: the epithelium, lamina propria, and within the Peyer's patches. Development of the gut immune system is thought to begin with the appearance of T cells in the human foetal intestine at 11 weeks of gestation. Moreover, Peyer's patches and B cells begin to form by 16 weeks of gestation (Braegger *et al.* 1996).

To summarise, previous research has shown that the development of preterm infants is very different from full term infants, and this includes the immune system of the gut. The infant, even born before 37 weeks of gestation, contains all the physical gut components necessary to develop and program the local immune system. Therefore, it is the environment these infants encounter after birth that causes changes in the development of the immune system. In the womb, the infant would have been exposed to a small volume of bacteria that would provide the immune system with a baseline in which to program tolerance. However, a preterm infant is born early and exposed to a significant increase in the number of bacteria. This overwhelms the immune system, leading to increased permeability and disease. Furthermore, a preterm infant is often unable to consume food orally and therefore the gut immune system is delayed in recognising orally derived antigens. In addition, the infant is given antibiotics, this will reduce the number of commensal bacteria leading to a reduction in tolerance programming and possibly allowing for the colonisation by pathogenic bacteria (Melville and Moss 2013).

1.6.3 The Role of IL-6 and IL-8 in Adult and Preterm Gut Disease

Levels of IL-6 and IL-8 have been found to be increased in the gut disease UC. It is believed that the increased amount on IL-6 results in excessive inflammatory response progressing the disease. Moreover, in CD increased levels of IL-6 have been associated

with relapse rates. In contrast, increased IL-8 levels in gut diseases has been linked to activation and migration of neutrophils (Műzes *et al.* 2012).

In preterm infants the levels of IL-6 and IL-8 have been found to be increased in the plasma of infants with NEC. Both molecules are believed to progress NEC through excessive inflammation and necrosis (De Plaen 2013).

1.6.4 Proteases and the Gut Immune System

Controlled proteolytic activity is crucial for the maintenance of gut immune homeostasis. However, it has been found that in inflammatory diseases, such as IBD, pro-inflammatory cytokines induce the up regulation of proteases. These proteases then destroy the mucosal layer. This then allows access to the epithelial membrane, where the up-regulated proteases breakdown the tight junctions and apical junctional complexes. Consequently, bacteria, their toxins and other antigens can cross the intestinal epithelial barrier to activate the adaptive immune system and therefore sustaining this inflammatory process (Biancheri *et al.* 2013). Furthermore, serine proteases from bacterial and host sources can activate PARs, this causes increased epithelial barrier permeability via the contraction of the myosin light chain (Bueno and Fioramonti 2008). Therefore, proteases have a huge effect of the gut immune system. However, this is not in a direct way but as a secondary outcome of intestinal epithelial breakdown.

1.6.5 Summary

Inflammation has been shown to have a dramatic impact on the development and the homeostasis of the gut immune system. Moreover, aberrant inflammation has been shown to develop and progress both adult and preterm gut diseases. However, this has only been demonstrated systemically and the levels of localised inflammation, especially in preterm infants, have yet to be reported.

1.7 The Gut Lung Axis

1.7.1 Introduction

The gut lung axis is only beginning to become understood, however the research provided so far indicates a reservoir of potential for the manipulation of the gut microbial community in the treatment of lung diseases. More specifically, the gut-lung axis comprises of the following components, as shown in Figure 9: the lung and gut environment, the microbial community of these environments, the immune system and outside environmental stimuli.

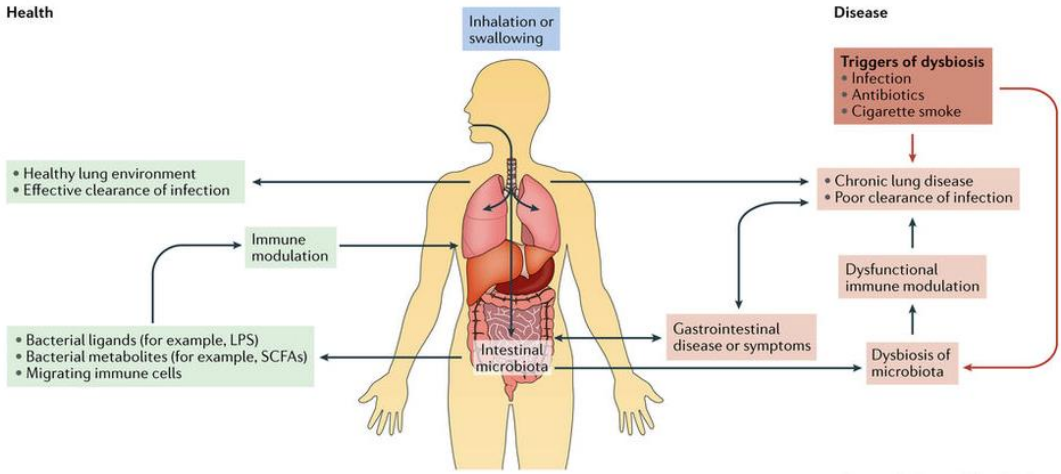

Figure 9. The Gut-Lung Axis in Health and Disease. Reproduced from Budden *et al* 2017, permission granted on 28.2.19. This figure demonstrates the normal contribution of the gut microbial community to lung health and the changes that occur during disease.

It has been found that chronic lung diseases, such as asthma and chronic obstructive pulmonary disease (COPD), occur together with gastrointestinal diseases, such as IBD (Roussos *et al.* 2003; Rutten *et al.* 2014). For instance, up to 50% of adults with IBD have some form for pulmonary involvement (Yazar *et al.* 2001). Furthermore, patients with COPD are 2-3 times more likely to be diagnosed with IBD (Keely *et al.* 2012). Functional structural alterations are more likely to occur in the intestinal mucosa in individuals with asthma (Vieira and Pretorius 2010). Therefore, the gut lung axis is an interesting area to research as this research clearly indicates at a level of cross talk between these two body sites. Due to the naivety of this field no research has been conducted into this cross talk in preterm infants.

1.7.2 The Interactions between the Gut and the Lungs

The epithelial surfaces of the gastrointestinal and respiratory tract are exposed to similar organisms as they can access both sites orally, via ingestion and micro aspiration. Furthermore, both epithelial surfaces are similar in structure and provide a physical barrier from the epithelium to the bacteria. This barrier also acts as the main line of immune defence in both organs, and commensal bacteria have been found to stimulate the immune system via this pathway (Buffie and Pamer 2013). Therefore, it is not unreasonable to hypothesize that these organs could support the same species of bacteria.

Further interesting studies have shown that gastrointestinal bacteria can have an effect on lung function. Oral gavage of faecal suspensions in *S. pneumoniae* infection model mice who were given antibiotics, showed an improvement in symptoms after the gavage (Schuijt *et al.* 2016). Furthermore, gut colonisation by beneficial bacteria such as *B. longum* has been found to reduce the incidence of asthma (Akay *et al.* 2014).

There has been considerable evidence to suggest that host epithelial immune cells assimilate information directly from microorganisms and from concomitant local cytokine response to adjust inflammatory responses. This has then been found to shape immune responses at distal sites, such as the lungs (Trompette *et al.* 2014; Marsland *et al.* 2015). However, there has been less evidence of direct transfer of microorganisms between the sites, but the translocation of bacteria has been seen in cases of sepsis and acute respiratory distress syndrome. This is when gut barrier integrity is compromised (Dickson *et al.* 2016).

The crucial role for commensal microbial community in health and disease has been proven by numerous studies in germ-free mice, whereby their susceptibility to allergic airway disease and some acute infections, is increased in these animals (Fagundes *et al.* 2012; Olszak *et al.* 2012). Current research has been diverging into assessing the effect of gut microbiota on systemic immunity, shown in Figure 10, as well as the effectiveness of pro and prebiotics on acute and chronic pulmonary disease. For example, certain bacteria in the gut have been found to stimulate pulmonary T helper cell responses. This was also shown for *S. pneumoniae* infection (Gauguet *et al.* 2015).

There is considerable evidence to show how the gut microbial community can affect immunomodulatory signals (Budden *et al.* 2017).

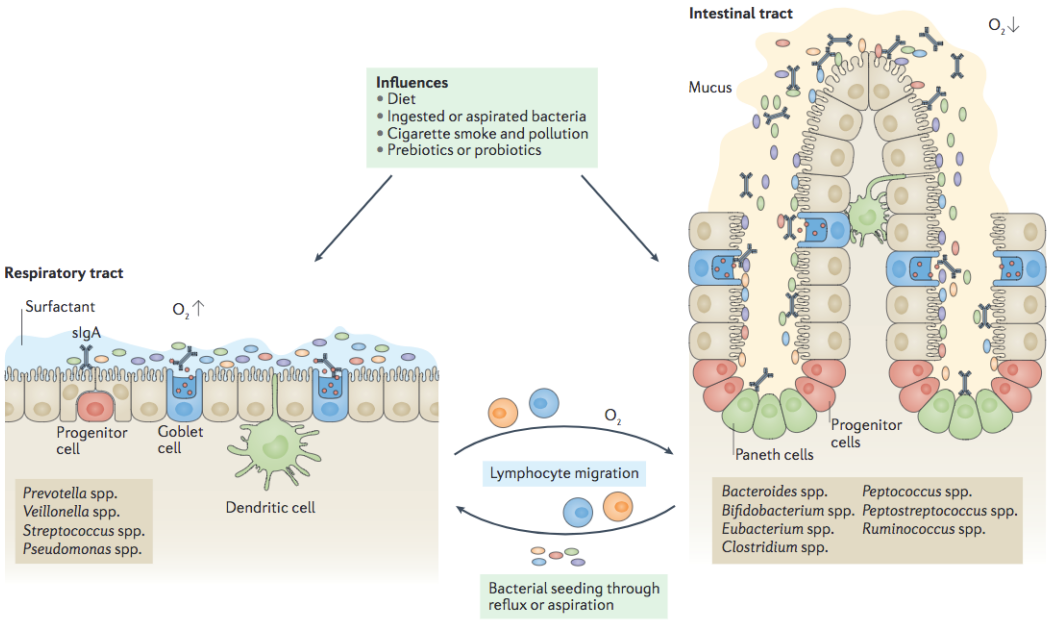

Figure 10. The Systemic Immunity Effect of Commensal Bacteria. Reproduced from Budden *et al* with permission, obtained on 25.2.19 (Budden *et al.* 2017). The diagram demonstrates the interplay between the commensal bacteria in the lungs and gut.

1.7.3 Gut Microbiota and Lung Diseases

An increased risk of asthma has been connected to the disruption of the gut microbiota in early life. The overall composition of the gut microbial community is not altered in infants at risk of asthma. However, it has been found that select changes in certain taxa can be detected in the first few months of life (Abrahamsson *et al.* 2014; Arrieta *et al.* 2015). For example, an increase in *B. fragilis* and total anaerobes in early life has been associated with an increased risk of asthma (Vael *et al.* 2008). In addition, decreases in diversity, *E. coli* and the relative abundance of *Faecalbacterium*, *Lachnospira*, *Rothia* and *Veillonella* species (Abrahamsson *et al.* 2014; Arrieta *et al.* 2015; Orivuori *et al.* 2015). This is similar in adults, with the overall diversity not differing between affected and healthy individuals but with species specific changes (Hua *et al.* 2015; Hevia *et al.* 2016).

There have been several proposed mechanisms by which the microbiota exerts it effect on the risk of asthma. Infants who were at risk of developing asthma had less LPS in

their stools (Arrieta *et al.* 2015). Polysaccharide A (PSA) from *B. fragilis* induced IL-10 in T cells protected against the development of asthma (Johnson *et al.* 2015). *H. pylori* has been found to alleviate allergic disease in mice in several different ways, by modulating the immune system (Koch *et al.* 2015; Sehrawat *et al.* 2015). Commensal bacteria can also have also have an effect on the development of asthma by the production and secretion of metabolites (Budden *et al.* 2017). Examples of these effects can be seen in Figure 11.

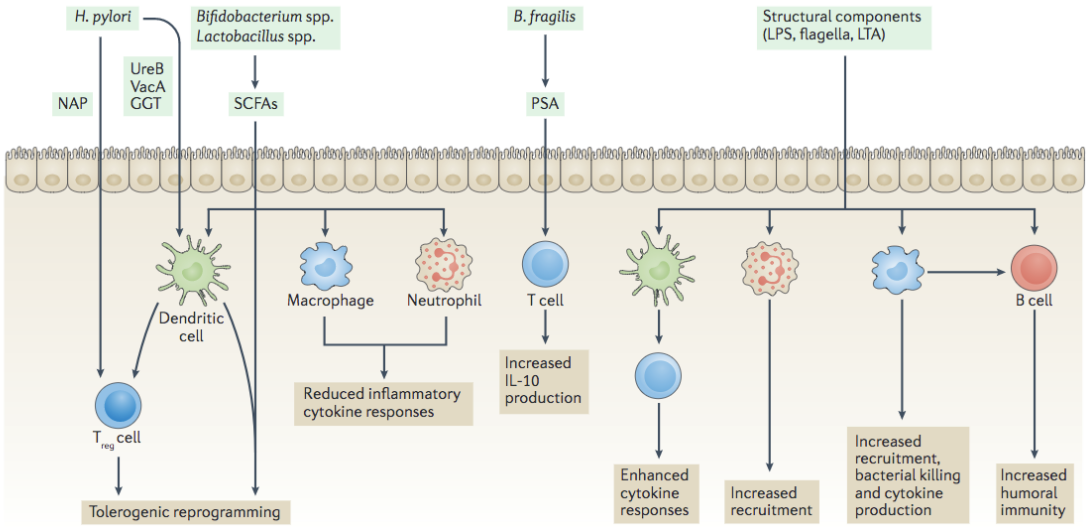

Figure 11. The Microbiotic Programming of the Immune System. Reproduced from Budden *et al* with permission, permission granted on 25.2.19. This diagram shows how the microbial community of the gut can program the local immune system.

Previously, respiratory microbiota research has focused on changes due to smoking, of which is a major risk factor for COPD. So far, it has shown that the lung microbial community is similar between smokers and non-smokers, but the oral microbiome differs significantly (Morris *et al.* 2013). It is believed that enrichment of the lung with the oral taxa of smokers causes increased inflammation in the lung (Segal *et al.* 2013). Then it is a combination of the increased inflammation and the inability to remove these bacteria contribute to COPD development, in only a subset of smoking populations. Furthermore, there are substantial differences between the lung microbiome of "healthy" smokers and those with COPD (Pragman *et al.* 2012; Sze *et al.* 2012). Interestingly, smokers have a decreased abundance of *Bifidobacterium* species (Zeitz *et al.* 2014).

The gut microbial community is broadly protective against respiratory infection, as a reduction in microbial content increased the virulence of infections (Chen *et al.* 2011a; Ichinohe *et al.* 2011; Fagundes *et al.* 2012; Wu *et al.* 2013; Schuijt *et al.* 2016). Administration of certain bacteria is protective against *S. aureus* pneumonia and *Bifidobacterium* species protected against pulmonary infection in mice (Wu *et al.* 2013; Gauguet *et al.* 2015; Kawahara *et al.* 2015; Vieira *et al.* 2016).

Several important mechanisms by which the gut microbiota promotes the clearance of pathogens have been identified. Innate immune responses to bacteria in the lungs are greatly enhanced by exposure to NOD-like receptor and TLR agonists in the gastrointestinal tract, which included peptidoglycan, LPS, lipoteichoic acid and CpG DNA (Chen *et al.* 2011a; Fagundes *et al.* 2012; Clarke 2014). Similarly, stimulation of TLRs by cell wall components and flagella of gut bacteria is necessary for effective adaptive immune responses to influenza (Ichinohe *et al.* 2011; Oh *et al.* 2014).

1.7.4 Summary

Current research suggests that the gut lung axis is a programming of the immune system by the microbial community of the body. However, this can be protective and detrimental in different situations. It has been shown that the translocation of bacteria from the gut to lungs and vice versa has less of an effect that immune system programming. Currently, there is no research on these interactions in preterm infants, but it would not be unreasonable to theorise that similar changes occur in these infants.

1.8 Thesis Aims

The overall aims of this PhD are to provide a detailed picture on the colonization and subsequent effects of the gut of the hospitalized preterm infant. The introduction to this thesis has shown that although there is a substantial body of knowledge on the preterm gut microbiome there are still significant gaps. As a result, this thesis will assist in bridging these gaps through the following hypotheses and aims.

The main aims and objectives of this thesis are as follows:-

1. **To investigate the development of the preterm gut microbiome development during the first 30 days of life.** (Chapter 3). This will be achieved using 16S rRNA gene metataxonomics and clinical data collected during sample collection.

2. **To determine the total faecal protease activity of preterm infant stool and attempt to make associations with the gut microbial community**. (Chapter 4). This will be achieved using a protease activity and inhibition assays.

3. **To conduct an in-depth analysis of the effect of antibiotics on the development of the preterm gut microbiome**. (Chapter 5). Previous research has shown that antibiotics can increase the incidence of disease. However, in-depth investigations into the bacterial community changes as a result of antibiotic administration have yet to be conducted. Therefore, I hypothesise that antibiotics will result in significant microbiome community changes.

4. **To analyse the gut and lung microbiomes of preterm infants to investigate if there is any translocation of bacteria**. (Chapter 6). I hypothesise that the development of the gut and lung microbiomes in these infants will be similar, potentially demonstrating a translocation of bacteria. Specifically, this will be achieved through the combination of the data from this thesis with that of Dr David Gallacher.

Chapter 2

Materials and Methods

Chapter 2. Materials and Methods

2.1 Ethical Approval and Consent Forms

Patient recruitment began in November 2014 at the University Hospital Wales (UHW) by Clinical Research Fellow Dr David Gallacher. This was later extended to North Bristol NHS Trust (NBT) in June 2015. Samples from NBT were collected by Dr Richard Wach and colleagues. Ethical permission for the study was granted by the Wales Research Ethics Committee 2 (Study No: 14/WA/0190). Other permissions were granted by the research and development committees of the participating NHS trusts. Ethical permission was obtained by Dr David Gallacher. Copies of the ethical approval alongside research and development approval can be found in Appendix 1. Participants were recruited using the following criteria by Clinical Research Fellow, Dr David Gallacher:

1. Preterm infants, ventilated. These were infants born at ≤ 32 weeks gestational age and required ventilation within the first 24 hours of life for respiratory distress syndrome.
2. Term infants, ventilated. These were infants born at ≥ 37 weeks gestational age, who were admitted to the neonatal unit and ventilated within 24 hours of age for a non- respiratory condition, e.g. abdominal surgical complications.
3. Term infants not ventilated. These were infants born at ≥ 37 weeks gestational age, with no known pathology, and who were considered well and residing on the post-natal ward.

There were several samples collected from these infants, described in more detail in 2.2, including, Bronchoalveolar Lavage (BAL), Tracheal Aspirate (TA), Nasopharyngeal Aspirate (NPA), blood, and stool. Samples were taken at every available opportunity over the first 30 days of life. All samples were collected by Dr David Gallacher.

Infants were not eligible for recruitment into the study if any of the following applied at birth. Infants with respiratory pathology, such as congenital adenomatous lesions or diaphragmatic hernias. Furthermore, incidences of meconium aspiration and ischaemic encephalopathy, were also excluded from recruitment. Finally, infants with undiagnosed chromosomal abnormalities or who were not expected to survive, as determined by the treating clinician, were also not recruited.

During participant recruitment, parents were approached to give consent for their child to be included in the study, either whilst preterm delivery was a threat or shortly after delivery. Upon consultation, an information sheet was provided, included in 0, and written informed consent was obtained from the parent or parents before the child was entered into the study. To protect the identity of the individuals included in the study, all participants were assigned a study number and patient identifiable information was inaccessible.

An amendment, mentioned previously, to the study protocol was given ethical approval. Permissions sought by Dr David Gallacher. This amendment allowed for the addition of Southmead Hospital, part of North Bristol NHS Trust, as a recruitment centre from May 2015 to September 2016. Samples were collected from Southmead Hospital by Dr Richard Wach's team. In addition, a second amendment was granted ethical approval in August 2016 to include the collection of stools from infants admitted to the post-natal ward. Permissions sought predominantly by Dr David Gallacher, assisted by me.

2.2 Sample Collection

2.2.1 Bronchoalveolar Lavage

Bronchoalveolar lavage samples were collected following the guidelines set out by the European Respiratory Society (Blic et al. 2000). In addition, an individual clinician, Dr David Gallacher at University Hospital Wales, performed the procedure on all recruited infants to reduce variability. Once the samples were collected, they were immediately placed on ice and transported to the laboratory for processing.

Intubated infants on the neonatal unit at the University Hospital Wales, routinely receive surfactant therapy at birth and if necessary, again at 12 hours of age. As a result, in order to avoid interference with the primary care of the infant, initial bronchoalveolar lavage (BAL) was performed at either 12 or 24 hours of age.

The procedure used during this study, established by our team, has been widely used in previous neonatal research. Firstly, the infant is moved to a supine position with the head to the left side. The ventilator was then briefly disconnected from the endotracheal tube. Following disconnection, a size 6Fr nasogastric feeding tube (Intervene, Chesterfield, UK) was then passed through the endotracheal tube until resistance was met. 1 ml/kg of sterile 0.9 % sodium chloride (B.Braun Medical Ltd, Sheffield, UK),

routinely used for injection, was then instilled through the catheter. After a brief pause a suction pressure of 5-7 kPa is applied to the nasogastric tube. Finally, the tube was slowly withdrawn, and the aspirated material collected in a mucous specimen trap (Pennine Healthcare, Derby, UK). Following the first procedure, the infant is given time to recover, and a second procedure is carried out and the resulting samples pooled. In an effort to reduce variability, the same clinician performed the procedure on all infants recruited to the study. The sample was then immediately placed on ice and transported to the laboratory for processing.

2.2.2 Tracheal Aspirate

Tracheal Aspirate (TAF) samples from University Hospital Wales were collected as part of routine clinical care by the nursing staff present on the neonatal ward, this was in accordance with local departmental policy. The endotracheal attachment was removed from the ventilator and 1 ml/kg of sterile 0.9 % sodium chloride (B.Braun Medical Ltd., Sheffield, UK) was instilled in to the endotracheal (ET) tube. A sterile suction catheter was then inserted to a depth equal to the length of the ET tube and a suction pressure of 5-7 kPa is applied as the catheter is slowly withdrawn. Finally, the infant is reconnected to the ventilator, after allowing the infant to recover the procedure is repeated, and the samples pooled. The sample was then immediately placed on ice and transported to the laboratory for processing.

Similar to Cardiff, samples from North Bristol Trust were collected as part of routine clinical care. However, the local policy utilises an in-line suction device (Halyard UK, Surrey, UK) that is changed every 24 hours as part of routine clinical care. Of interest, this allows the infant to remain attached to the ventilator during the sampling procedure, therefore reducing stress in the infant. Samples collected for the study were only done immediately after the device had been changed, this greatly reduced the risk of contamination of samples from previous aspirations. Firstly, 1 ml/kg of saline was instilled to the ET tube via a port on the in-line suction device. The catheter was then advanced to a depth equal to the length of the ET tube. Whilst the catheter is slowly withdrawn, a suction pressure of 5-7 kPa was applied to collect the sample. Once the infant has been reconnected to the ventilator and recovered from the previous procedure, it is repeated, and the samples pooled. Samples were immediately frozen at -20 °C, prior to transfer to the laboratory.

2.2.3 Nasopharyngeal Aspirate

At the University Hospital Wales site nasopharyngeal aspirates (NPAs) were collected alongside the BAL samples, by Dr David Gallacher. A size 6 Fr nasogastric feeding tube (Intervene, Chesterfield, UK) was gently inserted via the nostril to the nasopharynx. A suction pressure of 5-7 kPa was applied as the tube was slowly withdrawn. Furthermore, each nostril is suctioned to provide and inclusive sample. 2 ml of sterile sodium chloride (B.Braun Medical Ltd, Sheffield, UK), routinely used for injection, was then aspirated through the nasogastric tube in order to flush any mucous within the catheter into a mucous specimen trap (Pennine Healthcare, Derby, UK). The sample was then placed on ice and immediately transported to the laboratory for processing.

In contrast, NPA samples collected in Bristol were collected according to the local departmental guidelines. 2 ml/kg of sterile saline, routinely used for injection (B.Braun Medical Ltd, Sheffield, UK), was instilled directly through the nostrils. A sterile suction catheter was then used to aspirate the fluid from the nasopharynx under a suction pressure of 5-7 kPa. Samples were then immediately frozen at -18 °C, prior to transportation to the laboratory.

2.2.4 Stool

Stool samples were collected by nursing staff from within the infant's nappy using a universal container incorporating a sterile spatula (Thermoscientific, Leicestershire, UK). This was achieved as part of the infant's routine care. Samples were stored at 4 °C for up to 24 hours prior to transport to the laboratory and processing.

2.3 Total Samples Collected

As mentioned previously all samples were collected by Dr David Gallacher, Dr Richard Wach and his team at NBT or the nursing staff at UHW. Table 4, gives the demographics for the preterm infants recruited by site. This table was provided by Dr David Gallacher. In addition, Table 5 provides information on the types and number of samples collected from each preterm infant.

Table 4. Demographics of recruited preterm infants from two sites. Medians ± interquartile ranges are shown unless otherwise stated. UHW – University Hospital Wales, NBT – North Bristol Trust. Necrotizing Enterocolitis Grades were categorized according to Bell's staging criteria and CLD severity**.**

	All infants	Infants recruited from UHW	Infants recruited from NBT	p-value (UHW vs NBT
Number of infants	55	20 (36%)	35 (64%)	-
Number of samples				
Nasopharyngeal	539	145	394	-
aspirates	276	62	214	-
Tracheal aspirate fluid	89	89	0	-
Bronchoalveolar lavage	198	64	134	-
fluid	1102	360	742	-
Stool				
Total				
Sex (male)	36 (65%)	12 (60%)	24 (69%)	0.73
Gestation (weeks)	26.0 (24.7-27.5)	26.8 (25.3-29.4)	25.9 (24.7-26.6)	0.07
Birth weight (g)	764 (680-918)	835(695-1082)	746 (677-880)	0.16
Antenatal steroids	51/55 (93%)	18/20 (90%)	33/35 (94%)	0.18
Delivery mode (percent vaginal delivery)	29 (53%)	10 (50%)	19 (54%)	0.98
Multiple births	17 (31%)	6 (30%)	11 (31%)	1
Maternal antibiotic in labour	14 (25%)	4 (20%)	10 (29%)	0.70
Surfactant administration	55/55 (100%)	20/20 (100%)	35/35 (100%)	0.27
Chronic lung disease severity	5 (9%)	4 (20%)	1 (3%)	0.02
None	4 (7%)	3 (15%)	1 (3%)	
Mild	18 (33%)	3 (15%)	15 (43%)	
Moderate	28 (51%)	10 (50%)	18 (51%)	
Severe/Died				
Survival to discharge	47 (85%)	17 (85%)	30 (86%)	1
Ventilation days	17 (4-32.5)	25 (3.5-37.8)	17 (5-28.5)	0.55
Non-invasive ventilation days	47 (24.5-64.5)	38(19.8-53.5)	55(31.5-66)	0.14
Length of hospital stay (days)	93 (69.5-130)	104.5 (52.8-136.5)	93 (80-122)	0.95
Necrotising enterocolitis				
Grade 1	7 (13%)	4 (20%)	3 (9%)	1
Grade 2	1 (2%)	0 (0%)	1 (3%)	
Grade 3	6 (11%)	1 (5%)	5 (14%)	
Patent ductus arteriosus	40 (73%)	11 (55%)	29 (83%)	0.06
Initial breast milk	55 (100%)	20 (100%)	35 (100%)	1
Discharge breast milk	17 (31%)	6 (35%)	11 (37%)	1

Table 5. Demographics on the number and type of samples collected from each preterm infant.

Infant	NPA	TAF	BAL	Stool	Infant	NPA	TAF	BAL	Stool
1	11	0	7	5	**35**	11	0	1	5
2	8	0	3	3	**36**	4	3	0	1
3	4	1	4	1	**37**	13	12	0	5
4	9	0	8	3	**38**	13	2	0	4
5	13	0	13	6	**39**	10	1	0	5
6	4	0	1	1	**40**	13	8	0	5
7	2	0	2	0	**41**	7	7	0	1
8	9	6	9	4	**42**	13	7	0	4
9	7	0	1	5	**43**	13	1	0	4
10	6	9	7	4	**44**	12	7	0	3
11	13	1	0	5	**45**	12	1	0	5
12	12	6	0	5	**46**	12	12	0	4
13	9	0	0	4	**47**	12	11	0	4
14	9	4	0	3	**48**	12	9	0	5
15	13	12	0	5	**49**	8	2	2	4
16	10	4	0	4	**50**	6	5	1	4
17	9	4	6	4	**51**	7	7	0	0
18	2	2	2	0	**52**	13	13	0	4
19	10	9	0	4	**53**	13	4	0	3
20	12	6	0	5	**54**	13	4	0	5
21	10	1	0	4	**55**	13	13	0	1
22	13	2	0	4	**Total**	**535**	**276**	**87**	**181**
23	13	13	0	3					
24	7	6	7	5					
25	13	2	0	6					
26	12	12	0	3					
27	13	8	0	5					
28	6	5	0	1					
29	13	5	0	7					
30	13	2	0	5					
31	2	6	2	1					
32	8	9	8	4					
33	5	1	0	2					
34	5	11	3	1					

2.4 Sample Processing

2.4.1 Cardiff

2.4.1.1 Bronchoalveolar Lavage

This was performed by Dr David Gallacher. Samples were transported on ice immediately following collection, to the laboratory. On arrival, samples were then weighed, and the empty sample traps re-weighed after transferring samples to a microcentrifuge tube (Eppendorf, Stevenage, UK). This allowed for accurate weight calculation of the sample. The volume was taken to be 1 ml is equivalent to 1 g.

The BAL fluid was then centrifuged at 10,000 x g for 10 minutes at 4 °C, this produced a cell pellet and cell-free supernatant. The supernatant was aspirated and stored in 25 µl aliquots. The cell pellet and supernatant aliquots were labelled appropriately and then stored at -80 °C until further processing. Overall, processing of the raw sample was completed within 1 hour after collection.

2.4.1.2 Nasopharyngeal Aspirate

This was performed by Dr David Gallacher. Samples were transported on ice to the laboratory. Upon arrival, the NPA fluid was centrifuged at 10,000 x g for 10 minutes at 4 °C, this leaves a cell pellet and a cell free supernatant. The supernatant was aspirated and stored in 25 µl and 100 µl aliquots. Both the cell pellet and cell free supernatant were stored at -80 °C until further processing.

2.4.1.3 Tracheal Aspirate

This was performed by Dr David Gallacher. The tracheal aspirate samples taken by the nursing staff were immediately placed at 4 °C until collection. Furthermore, all samples were collected form the neonatal unit within 2 hours of sampling. Samples were then transported to the laboratory on ice and immediately stored at - 80 °C until further processing.

2.4.1.4 Stool

This was performed by Dr David Gallacher. Stool samples were weighed on arrival to the laboratory. Once the sample had been processed the empty container was weighed in

order to calculate the weight of stool obtained. The stool was divided into aliquots of 250 mg in sterile microcentrifuge tubes (Eppendorf UK, Stevenage, UK). The aliquots were stored at -80 °C until further processing.

2.4.2 Bristol

This was performed by the team at NBT. All samples taken in Bristol, including TA, NPA and stool samples, were immediately frozen at -18 °C. At several times during the study, samples were transferred from Bristol to the laboratory at University Hospital Wales. The samples remained frozen during the transportation, as they were transported in a heat resistant box containing freezer packs pre-frozen to -80 °C. The TA and NPA samples were stored at -80 °C upon arrival. However, the stool samples were defrosted and weighed on arrival to Cardiff, they were then divided into 250 mg aliquots in sterile microcentrifuge tubes (Eppendorf UK, Stevenage, UK) and stored at -80 °C until further processing.

2.4.2.1 NPA and TAF Samples

This was performed by Dr David Gallacher. At the point of DNA extraction, NPA and TAF samples were defrosted, on ice, and the samples weighed. The samples were then transferred to sterile microcentrifuge tubes (Eppendorf UK, Stevenage, UK). The samples were then centrifuged at 10,000 x g for 10 minutes at 4 °C to leave a cell pellet and a cell-free supernatant. The empty universal container was re-weighed to calculate sample weight and the volume taken as 1 mg is equal to 1 ml. the supernatant was removed and stored at -80 °C. the cell pellet was then used immediately used for DNA extraction.

2.5 Buffer Preparation

2.5.1 Buffers for Storage Assay

Six buffers designed to prolong the activity of extracted proteases were prepared as shown in Table 6, and stored at 4 ^{o}C until used.

Table 6. The number and composition of the protein storage buffers using during the storage assay.

Buffer	1 x PBS	25 % (v/v) Glycerol	0.05 % Sodium Azide	1 mM or 5 mM DTT
1	Yes	No	Yes	No
2	Yes	Yes	Yes	No
3	Yes	No	Yes	1 mM
4	Yes	No	Yes	5 mM
5	Yes	Yes	Yes	1 mM
6	Yes	Yes	Yes	5 mM

2.5.2 DNA Extraction Buffer (Maxwell DNA Extraction Protocol)

A DNA extraction buffer was prepared, by Dr David Gallacher, according to the protocol for DNA extraction from soil supplied by Promega. A buffer containing 5 M guanidine thiocyanate (Fisher Scientific, Leicestershire, UK), 1 % Na-Lauroylsarcosine (Sigma-Aldrich, Dorset, UK), 100 mM EDTA (Fisher Scientific, Leicestershire, UK) and 1 % Polyvinylpyrrolidone K30 (Sigma-Aldrich, Dorset, UK) was made using 150 mM sodium phosphate buffer. The solution was stored at 4 ˚C until needed.

2.5.3 1 x TAE Buffer

All working solutions of TAE solutions were made from a 50 x stock solution. Therefore, to make a 1 x solution the concentrated solution is diluted 50 x in sterile deionized water. To prepare the 50 x stock solution is as follows. Firstly, 242 g of Tris base was dissolved in 750 ml of sterile deionized water. Next 57.1 ml of glacial acid and 100 ml of 0.5 M EDTA was added to the Tris base. Finally, the solution was made up to a final volume of 1 l.

2.5.4 1 x PBS Buffer

All PBS solutions once made were autoclaved for 15 minutes at 121 ^{o}C to ensure sterility. The following protocol was used to make a 1 l stock of 1 x PBS. Begin by dissolving 8 g of NaCl in 800 ml of deionized water. To this solution add 0.2 g of KCl, 1.44g of Na_2HPO_4, and 0.24 g of KH_2PO_4. Next adjust the pH to 7.4 using HCl. Finally, add distilled water to make a final volume of 1 l.

2.5.5 IL-6 Wash Buffer

The IL-6 wash buffer was made up as follows, according to manufacturer's instructions (R&D Biosystems). A 0.05 % Tween20 solution was made in 1 x PBS, with a pH of 7.2 to 7.4. A 1 L solution was made, 5 ml of Tween20 was added to 700 ml of 1 x PBS and mixed until evenly distributed. The solution was then made up to 1 L. To ensure the sterility of the solution, it was passed through a 0.45 μm PES membrane vacuum filtration unit (VWR, Pennsylvania, USA). The container was then sealed and only opened in a Class 2 hood.

2.5.6 IL-6 Reagent Diluent

The Il-6 reagent diluent was made as follows, according to manufacturer's instructions (R&D Biosystems). A 1 x PBS was made as per Section 2.5.4, and then filtered through a 0.2 μm filter, to ensure sterility. For a full 96-well plate experiment, 52 ml of reagent diluent would be needed. Therefore, 5.2 ml of the 10% reagent diluent solution, provided in kit, would be diluted in 46.8 ml of the filtered 1 x PBS. A new reagent diluent solution was prepared fresh for each experiment.

2.5.7 IL-8 Wash Buffer

The Il-8 wash buffer was made exactly as per Section 2.5.5.

2.5.8 IL-8 Reagent Diluent

The IL-8 reagent diluent was prepared as follows, according to manufacturer's instructions (R&D Biosystems). A Tris-buffered saline (TBS) solution was prepared at concentration of 20mM Trizma base and 150mM NaCl. This solution was then filtered through a 0.2 μm filter to ensure sterility. Using the TBS a 0.01% BSA (provided in kit) and 0.05% Tween20 solution was prepared, this is the resulting reagent diluent.

2.6 Faecal Sample Processing

2.6.1 10% Faecal Slurry Preparation

Prior to faecal slurry preparation the faecal samples are defrosted on ice for 1 hour, furthermore all reagents were kept on ice during this process. This protocol was carried out in a Class II cabinet to prevent contamination by background bacteria. Firstly, an

empty sterile 2 ml tube (Eppendorf, UK) was weighed. Using a sterile swab 200 mg of faeces was transferred to the pre-weighed 2 ml tube. Next, 2 ml of sterile 1 x PBS (Gibco, Thermo Fisher Scientific, UK) was added to the faeces. The sample was then vortexed until a homogenous solution was made, approximately 5 minutes. The resulting homogenate was then centrifuged at 4 ^{o}C at 20,000 x g for 30 minutes. The supernatant was then aspirated and set aside to Section 81. The pellet was then taken forward to DNA extraction, Section 2.8.2. To add, if a sample was less than 200 mg in weight the volume of 1 x PBS added was adjusted accordingly to produce a 10% faecal slurry. This method was adapted from Morris *et al* 2012.

2.6.2 BCA Assay for Total Protein Concentration

The Bicinchononic Acid assay (Pierce, Thermo-Fisher, Loughborough, UK) was performed as per manufacturer's instructions, detail as follows. A set of Bovine Serum Albumin (BSA) standards were prepared according to manufacturer's instructions, where 1 x PBS was the diluent. The working reagent was prepared at the appropriate volumes to accommodate the number of samples tested. Due to the high protein concertation of the samples used, all samples were tested at a 1 in 10 and 1 in 100 dilutions. Negative controls included the working reagent and the 1 x PBS used as the diluent. Once the protein concentration has been determined, the sample is taken forward for normalisation to the desired concentration.

2.6.3 Normalisation to 1mg/ml Total Extracellular Protein

The total extracellular extract from Section 2.6.1 was the starting product for this protocol. Again, the supernatant and buffer were kept on ice during this protocol. Firstly, the total protein content of the faecal extract was tested according to Section 2.6.2. Once the total protein content was confirmed the faecal supernatant was normalised to 1 mg/ml total protein in a protein storage buffer. As shown in Section 2.5.1, the buffer containing 1 x PBS, 0.05% NaN_3 (Sodium Azide) and 5mM DTT (Sigma), was the most effective at preserving the protease activity in a total protein extract. All 1 mg/ml aliquots were 1 ml in size and stored at – 20 ^{o}C until further processing. Normalisation was performed in order to use 1mg/ml of Trypsin as a standard for detecting protease activity.

2.7 DNA Quantification using the QuBit Fluorometer

DNA concentration was quantified using the QuBit fluorometer from Life Technologies, UK. Firstly, a mastermix was prepared in a 1.5 ml tube (Eppendorf, UK), as follows: 0.5 µl of Quantiflour dye (Promega, Wisconsin, USA) per reaction with 99.5 µl of 1 x TE buffer (Promega, Wisconsin, USA) per reaction. Due to the measurement procedure of the machine, thin walled 0.5 ml micro centrifuge tubes were used. To these tubes 100 µl of 1 x TE buffer was added alongside 98 µl of the mastermix and 2 µl of the unknown sample. This gave a final reaction volume of 200 µl. In order for the machine to calculate an unknown DNA concentration, standards were also needed. These were prepared as follows: 100 µl of 1 x TE buffer, 99 µl of mastermix and 1 µl of the λ DNA solution provided (Progmega, Wisconsin, USA). There were 2 standards a high concentration and a low concentration. Firstly, the machine was calibrated using the 2 standards, the unknown samples could be tested. For samples that were below the limit of detection the protocol could be adjusted to use 5 µl of the unknown sample. This was performed as per manufacturer's instructions.

2.8 DNA Extraction Techniques

A comparison between these two kits was performed as part of this thesis to determine which would be most appropriate for this study. It was determined that the Qiagen kit gave a greater DNA quantity and was therefore used for the stool samples.

2.8.1 Maxwell DNA Extraction Process for Faecal Samples

The resulting stool pellet from preparation of total extracellular protein extract, Section 2.6, is the starting material for DNA extraction. The pellet was first resuspended in 500 µl DNA extraction buffer, Section 2.5.2, the sample was vortexed until the pellet was fully homogenised within the lysis buffer. The emulsion was then transferred to Lysing Tubes E (MP Biomedicals, UK). The samples were then homogenised using the FastPrep 24 device (MP Biomedicals, UK) for 30 seconds at 5.0 m/s. This was repeated 3 times with an incubation of 5 minutes at 4 °C in between each homogenisation. Samples were then centrifuged at full speed for 1 minute to displace the beads to the bottom of the tube. The resulting supernatants were aspirated into the Maxwell cartridges, ready for insertion into the Maxwell 16 automated DNA extraction machine (Promega UK, Southampton, UK). Furthermore, 300 µl of supplied elution buffer was

added to the appropriate chamber of the cartridge. The run would then be initiated. The eluted DNA was then aliquoted into sterile microcentrifuge tubes (Eppendorf UK, Stevenage, UK) in 50 µl volumes and stored at -20 °C until further use. This was done according to a method supplied by Dr David Gallacher.

2.8.2 Qiagen DNA Extraction Protocol for Faecal Samples

Prior to this protocol the total protein extract will have been removed and set aside, ready for further processing as detailed in Section 2.6. Firstly, 2 ml of InhibitEX buffer (Qiagen, Germany) was added to the faecal pellet from Section 2.6 and vortexed continuously for 1 minute, or until the sample was fully homogenized. 1.5 ml of this solution was transferred into an MPBiomedicals Lysing Matrix tube E (MPBio, Germany). The tubes were then homogenised for 60 seconds at 5.0 m/s, this process was repeated 3 times with 5 minute incubations on ice between each homogenisation. All samples were then heated for 5 minutes at 70 ^{o}C in a heat block or water bath. The heated samples were then vortexed for 15 seconds, before being centrifuged at 20,000 x g for 1 minute. After centrifugation 200 µl of the supernatant was transferred into a 1.5 ml tube (Eppendorf, UK) containing 15 µl of Proteinase K (Qiagen, Germany). Next 200 µl of buffer AL (Qiagen, Germany) was added and the tube vortexed for 15 seconds. The samples were then incubated at 70 ^{o}C for 10 minutes. After incubation, 200 µl of 96 – 100 % ethanol was added to the lysate and mixed by vortexing. Into the supplied QIAmp (Qiagen, Germany) spin coloumns, 600 µl of the lysate was added and centrifuged at full speed for 1 minute. The column was then placed into a new 2 ml collection tube (Qiagen, Germany) and 500 µl of buffer AW1 added to the column. The column was then centrifuged at full speed for 1 minute. The column was once again placed into a new 2 ml collection tube and 500 µl of buffer AW2 (Qiagen, Germany), was added to the column. The column was then centrifuged for 3 minutes at full speed. Again, the column was placed into a new 2 ml collection tube and centrifuged at full speed for 3 minutes. Finally, transfer the column into a clean 2 ml collection and 200 µl of buffer ATE was added and incubated for 1 minute at room temperature before being centrifuged for 1 minutes to elute the DNA. The eluate was taken as the faecal DNA extract and stored at -20 ^{o}C in 50 µl aliquots. This was as per manufacturer's instructions after homogenisation.

2.9 16S rRNA Gene Sequencing - Illumina MiSeq

Preparation of the extracted DNA and sequencing of the bacterial 16S rRNA gene for the BAL, TAF, and Stool samples was performed on my behalf by Prof Nigel Klein's team at the Institute of Child Health, in the laboratories of University College London. Particular thanks to Dr Dagmar Alber and Dr Grace Logan who performed the amplification and sequencing. The preparation prior to sequencing of the NPA samples was performed by Dr David Gallacher at the Institute of Child Health within University College London. All preparation and sequencing process followed the same protocol detailed in this Section, as performed by Dr Dagmar Alber and Dr Grace Logan. For the origin of primers used and other methodological techniques please refer to Gallacher *et al* 2020.

2.9.1 qPCR Inhibition Check

This was performed by Dr Dagmar Alber and colleagues. The qPCR inhibition assay was performed on a selection of samples. This was to ensure the DNA in the samples would amplify and there was no restrictions due to inhibition. A PowerSYBR Green PCR master mix (ThermoFisher, Leicestershire, UK), Table 7, was used with serial dilutions (neat, 1:10, 1:100, and 1:1000) of the sample DNA. All samples were run in duplicate. The sequence of the primers used during this assay are as follows: Forward primer TCCTACGGGAGGCAGCAGT, and the Reverse primer GGACTACCAGGGTATCTAATCCTGTT.

Table 7. The reagents and volumes used during a qPCR inhibition assay. The total volume of the reaction 20 µl. All reagents were stored at -20 ˚C and kept on ice during preparation of the experiment.

Volume	Reagent
10 µl	Master Mix (ThermoFisher, Leicestershire, UK)
6 µl	Nuclease Free H_2O (Bioline, London, UK)
2 µl	Sample DNA
1 µl	Forward Primer (Sigma-Aldrich, Dorset, UK)
1 µl	Reverse Primer (Sigma-Aldrich, Dorset, UK)

The thermal cycling conditions are as detailed below, and all assays were carried out on the Biorad T100:

- 10 minutes at 95 ˚C

- Followed by 45 cycles of:
    - 15 seconds at 95 ˚C
    - 1 minute at 60 ˚C

The resulting Ct values for each dilution were plotted against a dilution factor. A straight line demonstrated no inhibition to the PCR reaction.

2.9.2 Amplification of the 16S rRNA Gene Using Barcoded Primers and Adaptors

This was performed by Dr Dagmar Alber and colleagues. The 16S ribosomal RNA gene is amplified using a PCR, prior to the addition of specific primers and sequencing. To allow for the multiplexing of samples and the binding of amplicons to the flow cells of the MiSeq device, specific primers are designed, and incorporate barcodes and adapters. The barcodes were a unique sequence of bases used to identify the sample during the sequencing procedure. The adaptors introduce a 5' overhang to the sample DNA sequence, this allows for the binding of the target sequence to a complementary strand of DNA on the flow cell of the MiSeq instrument. Also included in these primers are the primer pad, link sequence and the gene specific sequence. The primer pad allows for adjustment of the melting temperature of the primer. Finally, the gene specific sequence is the traditional component of a primer and for this protocol is the V3-V4 region of the bacterial 16S gene. The sequences of the 16S rRNA gene specific part of the primer are 314F: CCTACGGGNGGCWGCAG and 805R: GACTACHVGGGTATCTAATCC. Originally, these sequences were designed to study marine bacterial communities (Herlemann et al. 2011). However, they have been used extensively in research utilising next generation sequencing, and have been found to capture an accurate reflection of the bacterial community, as they have a broad taxonomic range (Klindworth *et al.* 2013). Figure 12 below shows in detail and exemplar primer used during this protocol, a full list of all the primers can be found in Appendix 2.

Figure 12. An exemplar structure of the primers used during the 16S rRNA Gene Sequencing Illumina MiSeq sequencing. The full length of the primer is given below the detailed breakdown.

The PCR using the primers detailed in Figure 12 was performed using a Taq PCR Core Kit (Qiagen, Hilden, Germany), the master mix for the reaction is detailed in Table 8.

Table 8. The volume and reagents used during the amplification of the 16S rRNA gene with barcoded primers and adaptors. Preparation of this master mix and subsequent experiment was carried out in a fume hood to prevent contamination.

Volume	Reagent
0.25 µl	Taq Polymerase (Qiagen, Hilden, Germany)
1 µl	$MgCl_2$ Solution (25mM) (Qiagen, Hilden, Germany)
1 µl	dNTP Mix (10mM of each nucleotide) (Qiagen, Hilden, Germany)
10 µl	Q Solution (Qiagen, Hilden, Germany)
5 µl	10 x Buffer (Qiagen, Hilden, Germany)
25.25 µl	Nuclease Free H_2O (Bioline, London, UK)

Once the master mix had been made as described above, 42.5 µl was aliquoted into the appropriate wells of a 96 well PCR plate (Elkay, Hampshire, UK). This was followed by the addition of 1.25 µl of the relevant primer forward and reverse primer (Sigma-Aldrich, Dorset, UK) at a concentration of 20 pmol was added to each well as shown in Figure 13. Each sample for a total of two 96 well plates, was given a unique set of barcoded primers, as 2 sets of reverse primers were available. Amplification of samples were performed in batches of between 24-36 samples, in each batch a DNA extraction control and a negative control were run alongside the samples. Any batches that resulted in the production of a positive negative control were discarded and the experiment repeated.

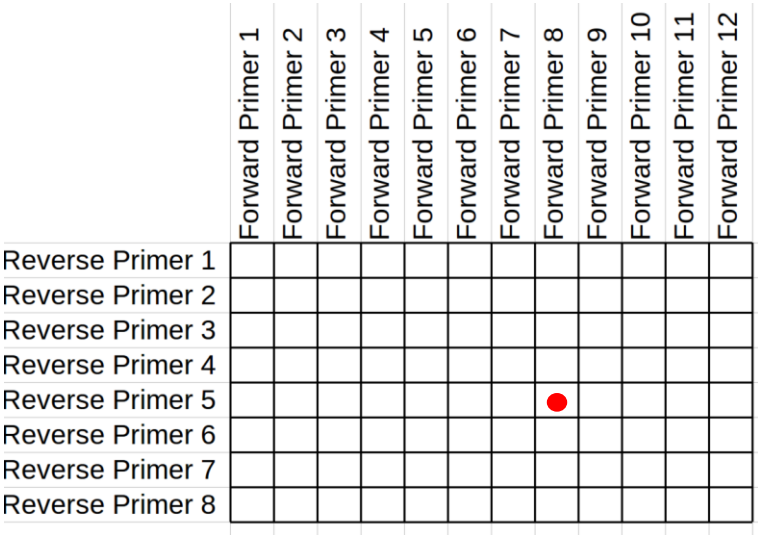

Figure 13. A diagrammatic layout of the primers included in each well of the amplification PCR. Therefore, each well contains a forward primer dictated by the columns in the diagram and a reverse

primer shown by the rows. For example, the well labelled with a red circle will include forward primer 8 and reverse primer 5.

Once the reaction was set up as detailed earlier, the plate was sealed and run on a thermal cycler under the conditions detailed in Table 9.

Table 9. Thermal cycling conditions used during the amplification of the 16S RNA gene and addition of barcoded primers.

Step	Temperature	Time	Cycles
Initial Denaturation	95 °C	3 minutes	1
Denaturation	95 °C	30 seconds	
Annealing	54 °C	30 seconds	x 30
Extension	72 °C	1 minute	
Final Extension	72 °C	10 minutes	1
Hold	12 °C	∞	1

2.9.3 Purification of PCR Products

This was performed by Dr Dagmar Alber and colleagues. Purification of the PCR products was achieved using AMPure Beads (Beckman Coulter (UK), High Wycombe, UK). Before purification could begin the AMPure Beads must be brought to room temperature and vortexed to ensure even distribution of the beads, before the addition of PCR products. As the V3-V4 primers for the 16S rRNA gene were used during this study, 35 µl of the AMPure Beads was added to each well containing PCR products. To ensure adequate mixing of the sample and beads pipetting up and down in each well at least 10 times was performed. The resulting solution was then incubated at room temperature for 5 minutes, to ensure complete binding of DNA to the beads. The plate was then placed on a magnetic stand for 2 minutes until the supernatant was clear and all beads were clustered at the bottom of the well. Whilst the plate remained on the stand the supernatant was carefully removed and discarded, 200 µl of 80% ethanol was added to each well and incubated for 30 seconds. The supernatant was then removed, and the process repeated, all ethanol was removed prior to the next step. With the plate *in situ* on the magnetic stand it was left to air dry for 15 minutes. The plate is then removed from the magnetic stand. 50 µl of AE buffer was added to each well and mixed thoroughly, then left to incubate at room temperature for 2 minutes. The plate was then placed on the magnetic stand and left for 2 minutes until the supernatant was clear. The supernatant was then aspirated and stored in a new 96 well PCR plate for up to 1 week at -20 °C.

2.9.4 Post PCR Quantification and Sample Pooling

This was performed by Dr Dagmar Alber and colleagues. Each sample was quantified using the Qubit HS assay for dsDNA on the Qubit analyser (Thermo Fisher, Leicestershire, UK), as previously described in Section 2.7. Furthermore, 2 µl of the sample DNA to be quantified was used per sample. Only samples with DNA concentrations greater than 0.5 ng/µl were taken forward to be sequenced, as they were considered to have amplified sufficiently for pooling. However, due to difficulties in amplifying the BAL samples, the threshold was lowered to 0.1 ng/µl. If a sample was above the threshold it was diluted appropriately to 0.5 ng/µl, if sample were 0.5 ng/µl they were used neat. Finally, 10 µl of each sample was combined into a single solution to produce the amplicon library. This was then aliquoted into smaller volumes and stored at -20 °C.

2.9.5 Pooled Library Quantification and Quality Check

This was performed by Dr Dagmar Alber and colleagues. The pooled library was quality checked and quantified using the TapeStation dsDNA assay (Agilent Technologies, California, USA). All samples were allowed to equilibrate to room temperature for 30 minutes before use. To the first well of a 96 well plate (Elkay, London, UK) 2 µl of ladder and 2 µl of sample buffer was added. To all other wells 2 µl of sample buffer and 2 µl of sample was added. The plate was then covered and vortexed (IKA) for 1 minute, each well was then pooled by centrifugation. Samples were then loaded into the Agilent 42000 TapeStation (Agilent Technologies, California, USA), and the assay started. The resulting data was quantified and displayed in the following formats. Firstly, an image similar to that of a gel electrophoresis is produced, Figure 14. The results should show one clean band at the expected size, this was determined by the ladder run alongside each sample. Finally, a ‘peak table’ is produced, which provided the concentration of each peak at pg/µl. These results were used during the dilution of the library.

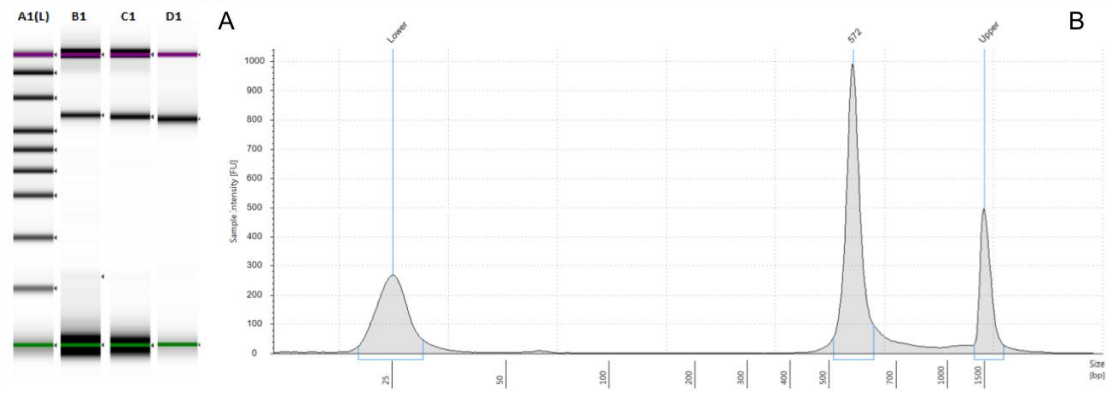

Figure 14. An example of the data and results output from the TapeStation. (A) Is a gel image produced by the TapeStation. Also shown on the image is an upper marker highlighted in purple and a lower marker highlighted in green. **(B)** Is a line graph produced by the TapeStation, that shows the same data as the gel image. This graph also includes in the upper and lower markers.

The final step is to quantify the library and to confirm the presence of Illumina primers on the 16S rRNA gene amplicons. This was achieved using the Next Library QuantKit for Illumina (New England Biolabs, Massachusetts, USA). The assay used a qPCR reaction containing 4 known standards to accurately quantify the concentration of DNA containing Illumina primers, within the pooled library. Prior to preparation of the library dilutions certain components of the QuantKit were prepared according to manufacturer's instructions. The library dilutions were prepared as follows: 1:100, 1:1000, 1: 10,000 and 1:100,000 with 1 x Dilution buffer prepared previously. Following this, the qPCR assays were prepared and run triplicate. 16 µl of the Master mix, prepared previously, was combined with 4 µl of each library dilution and 4 µl of the QuantKit standards (supplied in the kit ready for use). In addition, a no template control using only the dilution buffer was prepared. The assay was then run on the Biorad T100 machine using the FAM/SYBR channel, the thermal cycling conditions are shown in Table 10.

Table 10. Thermal cycler conditions for quantification of pooled DNA library.

Step	Temperature	Time	Cycles
Initial Denaturation	95 °C	1 minute	1
Denaturation	95 °C	15 seconds	x 35 Cycles
Extension	63 °C	45 seconds	

From the results the DNA concentration of the library was ascertained. The supplied DNA standards represents the following concentrations: 10 pM, 1pM, 0.1 pM, and 0.01 pM. Using the standards and the calculation below, the library concentration can be adjusted:

Adjusted Concentration = $\frac{\text{Calculated Concentration x 399}}{\text{Library Size (bp)}}$

2.9.6 Denaturing, Dilution and Loading of the Library into the MiSeq

This was performed by Dr Dagmar Alber and colleagues. Sequencing was performed on a MiSeq instrument (Illumina UK, Cambridge, UK). Prior to loading of the samples into the cartridge, the library is denatured by diluting with an equal volume of 0.2 N Sodium Hydroxide and Tris-HCL. The library was then further diluted with the hybridization HT1 buffer (Illumina UK, Cambridge, UK) to a concentration of 4 pM. Following denaturation and dilution the library was combined with denatured PhiX Control v3 DNA (Illumina UK, Cambridge, UK), this was to act as an internal control alongside the low diversity amplicon library. The final library was then loaded into the MiSeq cartridge (Illumina UK, Cambridge, UK) alongside the custom primers. A clean flow cell (Illumina UK, Cambridge, UK) is also required for the reaction, so this was added before the run was started. The MiSeq run was then initiated. The resulting FASTQ files were then taken forward to analysis.

2.10 Sequencing Data Processing using Mothur

Mothur v1.39.5 was used to process the 16S rRNA gene sequencing data. Mothur is an open-source, platform-independent and community supported software for describing and comparing microbial communities (Schloss et al. 2009). In order to process the raw data in a timely manner, the HIVE multicore processor computer (School of Bioscience, Cardiff University, UK). Due to the remote nature of this computer the relevant files needed to be uploaded using a File Transfer Program (FTP). Due to working on both a Macintosh and Windows system both Cyber Duck and WinSCP, respectively, were used to facilitate files transfer. Furthermore, the inbuilt Terminal app on the Macintosh and the PuTTy program was used on a Windows system. This software facilitated access to the remote HIVE computer in order to give commands during the data processing. A full copy of all the commands used in the Mothur package to process the raw sequencing data can be found in Appendix 3.

The aim of this protocol was to remove erroneous sequences and other unwanted data, this allowed the grouping of the remaining sequencing into Operational Taxonomic Units (OTUs) prior to data analysis. To begin with, the complementary forward and

reverse reads, another name for sequencing data, are combined to form contigs, or overlapping reads. Next, sequences that are greater than the 97.5th percentile and with more than 12 ambiguous bases are removed. All unique sequences are then combined to reduce time during further analysis. The resulting sequences are then aligned to the SILVA, version 132, reference database of 16S rRNA gene sequences (Quast *et al.* 2013). Once the sequences have been aligned to 16S rRNA gene, sequences that do not cover the target sequence, V3-V4, are removed along with sequences that contain a maximum homopolymer length of 6. The regions V3-V4 were chosen as they provide the best coverage for bacterial species. The sequences are then trimmed by removing the overhangs at each end of the target sequence. The unique sequences command was then re-run, as the alignment and overhang stages will have generated more. Following this, all sequences with up to 2 base pair differences were combined into the most abundant group. Here it is assumed that these differences are due to sequencing errors and not phylogenetic differences. All chimeric sequences were then removed. The entire dataset is then split into rare and abundant sequences, with rare meaning one copy of the sequence. One of the most crucial steps in assigning each sequence a taxonomic classification. For this command a group of files called Trainset 16 was downloaded from the Ribosomal Database Project (RDP), these are the files Mothur uses to determine the identity of a sequence (Cole *et al.* 2014). Once the origin of the sequences was known, all organisms of non-bacterial origin can be removed, as they are not relevant to this study. The identity of all the sequences are them grouped into OTUs. A file was then created detailed the number of times each OTU appeared in each sample, this is one of the files taken forward to data analysis. However, this is a large amount of data and it was favourable to create a smaller sub sample of the larger dataset to take forward for data analysis. Therefore, all samples containing less than 1000 reads were excluded from further analysis. Using this subsampled dataset, the uncorrected pairwise distances between aligned DNA sequences were calculated, quantifies the relatedness between sequences and therefore samples. In addition, the alpha diversity indices of the subsampled data were determined, this allows for analysis into the relatedness of organisms within each sample.

Outside of the Mothur software, further analysis was performed. Firstly, a phylogenetic tree using the FastTree software was drawn (Price *et al.* 2009). However, in Mothur the weighted UniFrac distances can be calculated using the phylogenetic tree drawn using FastTree. Secondly, using a file containing a representative sequence of each OTU the

species identity of each OTU can be determined using the USearch command and the 16S rRNA Gene Sequencing Training Set reference database from the RDP (Edgar 2010). A threshold of 97 % exact sequence match to the database was needed in order to assign a species identity to each OTU.

During later analysis it was found that some of the species identified could be wrong, in other words it did not fit with previously published data in the area. As a result, the accuracy of the species identified was checked by comparing the representative sequence for the OTU with sequences in the RDP for the identified genus. Only samples where the sequence showed greater than 97 % similarity to only 1 species within the RDP, were taken as correct and remained in the dataset. In cases where more than one species shared greater than 97% identity with the representative sequence, the sequencing was determined not be of an adequate depth to identify species.

2.11 Data Analysis

2.11.116S rRNA Gene Sequencing Data

The resulting output from the Mothur software was reformatted, in Microsoft Excel (2013), in order to be imported into R, a statistical computing and graphics software (Colin *et al.* 2017). The version of R used for all analysis was 3.4.1 (Single Candle). A core package used for this analysis was Phyloseq, allowing for the graphical representation of 16S rRNA gene sequenced microbiome data (McMurdie and Holmes 2013). This package was used in the R environment.

The Mothur output was reformatted into an OTU and a taxonomy table. The taxonomy table contained a list of all the OTUs identified by Mothur and the taxonomy for each OTU. The OTU table contained a list of all the OTUs and the number of times they appeared in each sample. In order to make the analysis more meaningful, in other words to examine the most predominant organisms, the OTUs containing less than 10 reads and that constituted less than 5% of the total sample, were removed. These resulting files were then imported into R and further analysis performed.

In the R environment, independent of the analysis to be done there are several key steps. Firstly, the relevant packages, such as phyloseq, and working directory, i.e. the location of the files to be imported into R, have to be set. Furthermore, another step that is repeated across analysis of sequencing data is to transform the sample counts. This step

converts the number of reads per sample into a percentage in order to display the results as relative abundance. Then, depending on the analysis to be done the steps are different.

2.11.1.1 Visualising 16S rRNA Gene Sequencing Data Using the R Software

The bar charts were produced using the following method. The packages needed for this analysis are phyloseq and ggplot2. Firstly, using the OTU and taxonomy files a 'phyloseq object' can be created, a matrix data frame containing the information from both sources. Then using this object, a bar chart is plotted to show the presence of each organism in each sample. Furthermore, the taxonomic levels of the bar chart can be changed from Phylum to Genus, as appropriate.

In order to determine the spread of data between groups, a combination of a boxplot and bee swarm can be used. The boxplot function is inbuilt to the R software, so an accessory package is not required, however for the bee swarm the 'beeswarm' package is needed. Firstly, the data is imported into the R environment. Then the boxplot is drawn first, followed by the bee swarm, which overlays onto the boxplot. This allows for the visualization of the spread of data alongside the mean and other statistics.

2.11.1.2 Diversity Analysis of 16S rRNA Gene Sequencing Data

Commonly the next step in microbiome data analysis is to determine the alpha diversity within samples. The data used during this analysis was generated in Section 2.10 using the Mothur software. In order to visualize the data, it was imported into Excel (Microsoft Office). Using the Excel software, histograms were drawn to show the change in alpha diversity over time.

Beta diversity analysis, using Principal Coordinates Analysis (PCoA), started with the initial steps listed above and used the R software. PCoA plots are used to determine relationships between samples in a dataset. Furthermore, meta data on the samples can be included, which allows for investigations into the causes of potential clustering. Firstly, the necessary packages within R were loaded, these were phyloseq, ggplot2 and vegan. For this analysis three data files were needed, the OTU, taxonomy and a meta data file containing information such as the gender etc. In the same process as before a phyloseq object is created from the three data files. Specifically, to the PCoA plots the data contained within the phyloseq object was ordinated using the Bray-Curtis

dissimilarity index (Beals 1984). The Bray-Curtis dissimilarity index is a statistical method to quantify the compositional dissimilarity between two different sites, or samples, which is based on the counts, number of reads of an OTU, at each site. Then using gglot2 it was possible to plot the ordination data in 2-dimensional space on a PCoA plot. The full script for this analysis can be found in Appendix 4.

2.11.2 General Data Analysis

2.11.2.1 Statistics.

For this project a p value of 0.05 was used. A Mann-Whitney U test, conducted in SPSS was used to determine significance for this thesis.

2.11.3 STAMP

STAMP as described by its creators is "a software package for analysing taxonomic or metabolic profiles" (Parks *et al.* 2014). Furthermore, the software allows you to choose the most appropriate statistical test for the data you have, this allows for more accurate reporting of results. The OTU and Meta data files are all that is needed for this software. For this project, this software was used to determine any significant effects of clinical data on microbiome profiles. The main test used to determine these differences was extended error bars.

2.12 Determination of Protease Activity

2.12.1 Preparation of Casein Fluorescein Isothiocyanate (FITC-Casein)

Manufacturer's instructions show that 5 mg/ml is an appropriate concentration for FITC-Casein for storage, with the following adjustments. 5 mg of FITC-Casein is dissolved in 1 ml of sterile deionised H_2O. This solution was then aliquoted into 20 µl volumes in sterile 1.5 ml tubes (Eppendorf UK, Stevenage, UK) and stored at -20 °C until further use. In addition, all FITC-Casein aliquots and working solutions were stored away from light. Upon use the 20 µl FITC-Casein would be diluted with 980 µl of sterile deionised H_2O, to a working concentration of 50 µg/ml.

2.12.2 Preparation of Standards

Trypsin was used as the standard for all protease activity assays during this project. Each set of standards were prepared for each assay and were never stored or frozen, this was to ensure 100% activity. The top standard was prepared at 1 mg/ml of Trypsin in 1 x PBS. The standards were then made using a serial dilution to 10 ng/ml. For each assay 100 µl of the standard was used.

2.12.3 FITC-Casein Assay for the Determination of Total Protease Activity

FITC-Casein aliquots and samples were defrosted on ice for at least 2 hr before the assay was set up. Whilst the samples were defrosting the Trypsin standards were prepared and stored on ice. All reagents were stored on ice to prevent breakdown of the products, or premature proteolytic reactions. Once the FITC-Casein was defrosted, 980 µl of sterile deionised H_2O was added and the mixture thoroughly vortexed. In a 96 well plate (Nunc, Thermo-Fisher, Loughborough, UK) 50 µl of the working stock of FITC was pipetted into all appropriate wells. Then 100 µl of the sample or standard to be tested was pipetted into the necessary wells. All samples and standards were measured in triplicate. The plate was then sealed and incubated at room temperature in a light free environment. Numerous samples were tested for protease activity using this method such as Faecal Total Extracellular Protein Extracts (FTEPEs), supernatants from bacterial cultures, Faecal Total Protein Extracts (FTPEs), for example. Upon completion of the 1 hour incubation the pate was then unsealed and placed into the Tecan plate reader (Tecan, Switzerland). The fluorescence was then measured using an excitation and emission filter of 485 nm/538 nm respectively. Furthermore, a blank in triplicate of the sample diluent was assayed alongside the samples and standards.

2.12.4 Analysis of Protease Activity Data

The inclusion of the standards not only provided a positive control to the experiments but were also used to provide a quantitative measurement of the relative protease activity of the unknown samples. The final fluorescence measurement, once the background fluorescence had been deducted (negative control), for the top Trypsin standard was taken as 100% protease activity. The majority of samples tested were normalised to a total protein concentration of 1 mg/ml. Therefore, if a sample showed

100% protease activity then the 1 mg/ml protein content was purely proteases. Once the unknown samples had the background reading removed, it was compared to the top Trypsin standard and a percentage of activity compared to 1 mg/ml of Trypsin was given.

2.13 ProteaseArrest Assay

The samples tested using this assay were the faecal total extracellular protein extracts, previously normalised to 1 mg/ml total protein. The same protocol, as per manufacturer's instructions, was followed for both the bacterial and human ProteaseArrest (G-Biosciences, USA). Furthermore, all reagents were kept on ice during this experiment and both the ProteaseArrest and samples were defrosted before use. In a 96 well plate (G-Biosciences, USA) add 100 µl of the sample in question to the appropriate wells. To each sample 1.5 µl of ProteaseArrest and 1.5 µl of 100 x EDTA solution (G-Biosciences, USA) was added to each sample. Finally, 50 µl of FITC-Casein was added to each well. The plate was then sealed and incubated for 1 hour at room temperature. The fluorescence was then measured on Tecan (Switzerland) to determine the inhibitor effect of the ProteaseArrest. These assays were used because there were no other suites of protease inhibitors available.

2.14 ProteSeeker Assay

During this assay all reagents were defrosted and kept on ice. In a 96 well plate (Nunc, Denmark) 100 µl of the sample to be test was added, this was followed by 1.5 µl of the 100 x protease inhibitor to the appropriate well. In total there was 12 protease supplied in the kit that were tested, separately, upon the same sample. Finally, 50 µl of FITC-Casein was added, the plate sealed and incubated for 1 hour at room temperature. The fluorescence was then measured on Tecan to determine the level of inhibition by the 12 inhibitors of the ProteoSeeker kit (G-Biosciences, USA).

2.15 Determination of the Dominant Protease Profile of Preterm Infant Faecal Samples.

2.15.1 Determination of Protease Origin using the G-Biosciences ProteaseArrest Kits

The samples tested using this assay were the faecal total extracellular protein extracts, previously normalised to 1 mg/ml total protein. The same protocol was followed for both the bacterial and human ProteaseArrest, as per manufacturer's instructions (G-Biosciences, USA). Furthermore, all reagents were kept on ice during this experiment and both the ProteaseArrest and samples were defrosted before use. In a 96 well plate (G-Biosciences, USA) add 100 µl of the sample in question to the appropriate wells. To each sample 1.5 µl of ProteaseArrest and 1.5 µl of 100 x EDTA solution (G-Biosciences, USA) was added to each sample. Finally, 50 µl of FITC-Casein was added to each well. The plate was then sealed and incubated for 1 hour at room temperature. The fluorescence was then measured on Tecan (Switzerland) to determine the inhibitor effect of the ProteaseArrest. Again, this was the only kit available on the market.

2.15.2 Determination of Protease Families using the G-Biosciences ProteSeeker Kit

Protocol was followed as per manufacturer's instructions. During this assay all reagents were defrosted and kept on ice. In a 96 well plate (Nunc, Denmark)) 100 µl of the sample to be test was added, this was followed by 1.5 µl of the 100 x protease inhibitor to the appropriate well. In total there was 12 proteases supplied in the kit that were tested, separately, upon the same sample. Finally, 50 µl of FITC-Casein was added, the plate sealed and incubated for 1 hour at room temperature. The fluorescence was then measured on Tecan (Switzerland) to determine the level of inhibition by the 12 inhibitors of the ProteoSeeker kit (G-Biosciences, USA). Again, this was the only kit available on the market.

2.15.3 Fluorescence Assay to Determine the Protease Inhibition of Preterm Faecal Samples.

The results of the ProteSeeker kit from Section 2.14, showed that only 2 of 12 inhibitors were significantly effective on a sample of the total faecal extracellular protein extracts.

Therefore, for the entire cohort it was decided to only test all the samples for the presence of these 2 proteases. This protocol was conducted as follows.

The protease inhibitors were made as a concentrated stock solution, aliquoted and stored at -20 ˚C, prior to experiments. Phenylmethylsulfonyl fluoride (PMSF) (Melford, Suffolk, UK) was made to a stock concentration of 100mM, 0.01742 g of PMSF was diluted in 1 ml of 98 % ethanol. The solution was then aliquoted into 20 µl and stored at -20 ˚C. The second protease inhibitor was ethylenediaminetetraacetic acid (EDTA). EDTA (Acros Organics, Fisher Scientific, Loughborough, UK) was made to a concentration of 10 mg/ml in sterile H_2O. The solution was stored at -20 ˚C in 100 µl aliquots.

All reagents for the experiments were defrosted on ice for at least 1 hour prior to setting up, and then kept on ice throughout. Firstly, the PMSF stock solution was diluted 1 in 10 to produce a working concentration of 10mM. Next 100 µl of each sample was added in triplicate to the appropriate wells. This was followed by 7.5 µl of PMSF and 3.75 µl of EDTA to each sample in the appropriate wells. Finally, 50 µl of FITC was added to each well, the pate sealed and incubated in the dark, at room temperature for 1 hour.

After the 1 hour incubation, the plate was read in the Tecan Infinite M200 Pro (Tecan, Switzerland) at an excitation and emission filter of 485 nm and 538 nm.

2.15.4 Analysis of the protease inhibition of preterm faecal samples.

Included in the experiment is each sample without inhibition, in triplicate, the fluorescence of this sample is taken as 100 % activity. The sample containing an inhibitor are then compared to the 100 % value and the percentage inhibition calculated. For example, if a sample gave 56 % activity in the presence of an inhibitor, this would give a 44 % inhibition. This data analysis was performed using Excel from Microsoft Office.

Further analysis from this allowed for the determination of the protease content of the sample in question. For example, using the previous example, if there was 44% inhibition by PMSF, but no inhibition by EDTA, then I would define this sample as having a varied protease profile. This is because the level inhibition by a serine protease does not constitute more than half of the inhibition of that sample, and there is no inhibition by EDTA. To give another example, if there was more than 50% inhibition by

both PMSF and EDTA, then this sample would be defined as having a serine and metalloprotease profile. All of the samples were evaluated in this way to determine the families of proteases present in each sample.

2.16 Enzyme-Linked Immunosorbent Assays (ELISAs) Protocol

The concentration of IL-6 and IL-8 in the total faecal extracellular protein extracts was tested using an enzyme-linked immunosorbent assay or ELISA. The principal of this assay is to quantify the amount of IL-6 or IL-8 in a sample by comparing it against the samples used. The amount of absorbance measured is proportional to the amount of the IL-6 or IL-8 in a sample. The ELISA utilises a capture antibody to the inflammatory molecule in question, this antibody is bound to the bottom of well in a 96 well plate (R&D Biosystems, Minnesota, USA). A biotinylated detection antibody is then incubated with capture antibody and molecule in question, it then binds to capture antibodies containing the molecule in question. All unbound detection antibody is washed away before the next step. When the sterptavidine HRP is added to the well it binds to the detection antibody, a blue colour develops when tetramethylbenzidine substrate solution is added to the well. This blue colour is proportional to the amount of inflammatory molecule bound to the capture antibody. Concentrated acid was added to stop the reaction and a yellow colour develops. The absorbance is then measure.

2.16.1 IL-6

The DuoSet IL-6 ELISA kit (R&D Systems, MN, USA) was used to measure the IL-6 concentration in total extracellular faecal protein extracts, as per the manufacturer's instructions and as detailed below. All antibodies and reagents were provided in the ELISA kit. This kit was chosen above others as it contains both IL-6 and IL-8.

Firstly, the 96 well plate (R&D Biosystems, MN, USA) was prepared by diluting the capture antibody to a concentration of 2 μg/ml in plate coating buffer (1 x PBS, supplied in kit). 100 μl of the capture antibody solution was added to each well, the plate sealed and incubated at room temperature, protected from light overnight.

The next day, the plate was washed with 200 μl of wash buffer (R&D Biosystems, MN, USA), in each well, this process was repeated 3 times. The plate was then incubated for 1 hour at room temperature with a 1% BSA in reagent diluent solution, 200 μl of this solution was added to each well. The aim of this step is to 'block' any areas of the plate

from being able to contain any IL-6 when the same is added, so only the capture antibody contains the IL-6, this gives a more accurate reading. Following the incubation, the plate is washed as per above. Following this, 100 µl of either standard, sample or blank (reagent diluent, protein storage buffer) was added to the appropriate wells. IL-6 standards were prepared as follows: a stock solution of recombinant human IL-6 (supplied in kit) was prepared into a 1:2 serial dilution form 600 pg/ml to 9.375 pg/ml. The standards were diluted in reagent diluent. The standards, samples and blanks were incubated for 2 hours at room temperature, protected from light. The samples were defrosted and kept on ice before addition to the plate. Once again, the incubation was followed by the wash step. Next, 100µl of the biotinylated detection antibody, made in reagent diluent to 50 ng/ml was added to the appropriate wells and incubated for 2 hours, in the same conditions. The wash step was repeated before the addition of the streptavidine-HRP. The Strepavidine-HRP was diluted 1 in 40 in reagent diluent, and 100 µl was added to each well and incubated for 20 minutes at room temperature, protected from light. Another wash step was performed. The substrate solution (supplied in kit) was prepared using equal volumes of reagent A and reagent B, 100 µl of this was then added to each well and incubated in the dark for 20 minutes. Wells containing IL-6 turned blue. After 20 minutes 50 µl of stop solution (2 N Sulphuric acid, supplied in kit), this turned the blue to a yellow colour. The absorbance of each well was then measured using a plate reader (Tecan, Switzerland) at 450 nm.

2.16.2 IL-8

A DuoSet IL-8 ELISA kit (R&D Systems, MN, USA) was used to measure the IL-8 concentration in total extracellular faecal protein extracts, as per the manufacturer's instructions. All antibodies and reagents were provided in the ELISA kit. This kit was chosen above others as it contains both IL-6 and IL-8.

Firstly, the 96 well plate (R&D Systems, MN, USA) was prepared by diluting the capture antibody to a concentration of 2 µg/ml in plate coating buffer (1 x PBS, supplied in kit). 100 µl of the capture antibody solution was added to each well, the plate sealed and incubated at room temperature, protected from light overnight.

The next day, the plate was washed with 200 µl of wash buffer (R&D Systems, MN, USA), in each well, this process was repeated 3 times. The plate was then incubated for 1 hour at room temperature with a 1% BSA in reagent diluent solution, 200 µl of this

solution was added to each well. The aim of this step is to ‘block’ any areas of the plate from being able to contain any IL-6 when the same is added, so only the capture antibody contains the IL-6, this gives a more accurate reading. Following the incubation the plate is washed as per above. After the wash step the standards, samples and blanks (reagent diluent and protein storage buffer) are added to the appropriate wells and incubated for 2 hours at room temperature, protected from light. The IL-8 standards were prepared were prepared using a stock of recombinant human IL-8 supplied with the kit, and reagent diluent prepared as follows: 0.1% BSA in Tris-buffered saline and 0.05 % Tween 20. A 2:1 serial dilution was prepared from 2000 pg/ml to 31.25 pg/ml in reagent diluent. The standards, samples and blanks were incubated for 2 hours at room temperature, protected from light. The samples were defrosted and kept on ice before addition to the plate. Once again, the incubation was followed by the wash step. Next, 100µl of the biotinylated detection antibody, made in reagent diluent to 50 ng/ml was added to the appropriate wells and incubated for 2 hours, in the same conditions. The wash step was repeated before the addition of the streptavidine-HRP. The Strepavidine-HRP was diluted 1 in 40 in reagent diluent, and 100 µl was added to each well and incubated for 20 minutes at room temperature, protected from light. Another wash step was performed. The substrate solution (supplied in kit) was prepared using equal volumes of reagent A and reagent B, 100 µl of this was then added to each well and incubated in the dark for 20 minutes. Wells containing IL-6 turned blue. After 20 minutes 50 µl of stop solution (2 N Sulphuric acid, supplied in kit), this turned the blue to a yellow colour. The absorbance of each well was then measured using a plate reader (Tecan, Switzerland) at 450 nm.

2.16.3 Analysis of ELISA Results

The value of the blanks was subtracted from the samples and standards, in order to begin analysis. A standard curve was drawn using the values obtained from the known standards, using a 4-parameter logistic curve fit. The concentration of the unknown samples was then interpolated from the standard curve. The software used to perform this analysis was GraphPad Prism. Shown below is an example of the standard curve generated from the analysis if an IL-6 ELISA.

2.17 qPCR for Total Bacterial Load

2.17.1 Preparation of Standards

Escherichia coli (*E. coli*) genomic DNA was obtained using the following method. A pure culture of *E. coli* was grown from a freezer stock provided by Prof Julian Marchesi. The freezer stock was prepared from tryptic soy broth (TSB) and 10 % DMSO. *E. coli* was grown on plates of nutrient agar overnight at 37 ˚C and 5 % CO_2. A single colony was then picked using a sterile loop in aseptic conditions and grown in TSB overnight with shaking at 37 ˚C and 5 % CO_2. The overnight culture was then centrifuged at 4000 rpm for 30 minutes at 4 ˚C. The supernatant was removed, and the cell pellet taken for DNA extraction using the method outlined in Section 2.8.1. Concentration of the DNA extract was determined using a Qubit fluorometer as detailed in Section 2.7. This was determined to be 24.5 μg/ml.

The *E. coli* genome contains 7 copies of the 16S rRNA gene (Klappenbach *et al.* 2001) and in total the genome comprises of 4.6×10^6 base pairs (Blattner *et al.* 1997). Furthermore, the average molar mass of a single base pair is 650 g/mol/bp. Below is the calculation to determine the number of copies of the 16S rRNA gene in the extracted DNA from a pure *E. coli* culture, where the DNA concentration was measured to 24.5 μg/ml.

1. Calculate the weight of 1 mole of the *E. coli* genome:

= Size of the *E. coli* genome x Molar mass per base pair

= 4.6×10^6 base pairs x 650 g/mol/bp

= 2.99×10^9 g/mol

2. Calculate the Molarity of the *E. coli* genome within the solution of extracted DNA:

= Concentration of DNA in solution / Molar mass of *E. coli* genome

= 24.5×10^{-3} g/L / 2.99×10^9 g/mol

= 8.194×10^{-12} M

3. Multiply by Avogadro's constant to calculate the number of copies of the *E.*

coli genome in solution:

$= 8.194 \times 10^{-12} \times 6.022\ 10^{23}$

$= 4.934 \times 10^{12}$ copies/L

$= 4.934 \times 10^{9}$ copies/ml

4. Multiply by 7 to find the number of copies of the 16S rRNA gene within the extracted DNA:

$= 4.934 \times 10^{9}$ copies/ml x 7

$= 3.454 \times 10^{10}$ copies of the 16S rRNA gene per ml

5. It was decided that seven standards would be used, ranging from 1×10^{7} to 1×10^{1} copies per ml. Therefore, the standards were prepared using the following calculations:

1ml (1000 µl) / 2.5 µl (volume used in the qPCR reaction)

= 400

6. Therefore, the first standards needs to be 400 times more concentrated in order for 2.5 µl to contain 1×10^{7} copies of the 16S rRNA gene:

$1 \times 10^{7} \times 400 = 4 \times 10^{9}$ copies per ml

7. Using the $C_1 \times V_1 = C_2 \times V_2$ calculation:

4×10^{9} copies/ml x 1 ml = 3.454×10^{10} copies/ml x V_2

$= 4 \times 10^{9} \times 1 / 3.454 \times 10^{10}$

= 0.1158 ml

= 115.8 µl of the extracted *E. coli* DNA

8. Therefore, a serial dilution was made using 115.8 µl and 884.2 µl nuclease free water (Severn Biotech).

2.17.2 qPCR Protocol for Quantification of Bacterial Load

The assay to quantify bacterial load in samples was adapted by Dr David Gallacher and performed by myself, from the BactQuant protocol published by Liu *et al* 2012 (Liu et al. 2012). All experiments were performed using strips of white Thermo-Tubes in white

with clear caps (Thermo Scientific, UK) on the Bio-Rad machine. Furthermore, the MxPro software package was used to collect and export the data. A master mix was prepared as per Table 11.

Table 11. qPCR for bacterial load reaction components and volumes. The total reaction volume used is 10 µl. All reagents were stored at -20 ˚C and kept on ice during any experiments.

Volume	Reagent
5 µl	Platinum qPCR Supermix – UDG – with ROX (Thermo Scientific, Leicestershire, UK)
2.12 µl	Nuclease Free H_2O (Severn Biotech Ltd, Worcestershire, UK)
0.18 µl	Forward Primer (100 µM) (Thermo Scientific, Leicestershire, UK)
0.18 µl	Reverse Primer (100 µM) (Thermo Scientific, Leicestershire, UK)
0.0225 µl	TaqMan Probe (Thermo Fisher, Leicestershire, UK)

The sequences of the forward and reverse primers used during this protocol are as follows (Liu *et al.* 2012):

1. Forward Primer: 5’ – CCTACGGGDGGCWGCA - 3′
2. Reverse Primer: 5′- GGACTACHVGGGTMTCTAATC -3′

The above primers were obtained in a lyophilised form (Thermo Scientific, Leicestershire, UK) and reconstituted to 100 µM, according to manufacturer’s instructions, using nuclease free H_2O.

A TaqMan Probe incorporating a 6-FAM (6-Carboxyfluorescein) reporter and a Molecular-Groove Binding Non-Fluorescence Quencher (MGBNFQ) was used for this assay (ThermoFisher, Leicestershire, UK). The sequence is detailed below:

TaqMan Probe: (6FAM) 5′-CAGCAGCCGCGGTA-3′ (MGBNFQ)

Prior to each reaction the pipettes used during this assay were irradiated using UV light for 5 minutes. Furthermore, set up of the reaction is done under a flow hood. This was done to reduce contamination. For each reaction 7.5 µl of the master mix, detailed in Table 11, was added to each tube. In addition, to the master mix 2.5 µl of standard DNA or unknown sample was added. This makes the total reaction volume to be 10 µl.

The thermal cycling conditions are listed below:

- 3 minutes at 50 ˚C for UNG treatment
- 10 minutes at 95 ˚C for Taq activation
- Followed by 40 cycles of:
    - 15 seconds at 95 ˚C for DNA denaturation
    - 1 minute at 60 ˚C for annealing and extension

A standard curve was included in every reaction, this was achieved by running the previously made standards, Section 2.17.1, in triplicate. In addition, all samples were run in triplicate for each experiment. Analysis of the data can be found in Section 2.17.3.

2.17.3 Analysis of qPCR Results

qPCR results were analysed using Excel. The standards allowed for the creation of a standard curve. This standard curve was then used to determine the total bacterial load in the unknown samples. The total bacterial load was then determined per mg of stool within that stool by dividing the total by the number of mg of stool used during sample processing.

Chapter 3

The Development of the Preterm Infant Gut Microbiome in the First Month of Life

Chapter 3. The Development of the Preterm Infant Gut Microbiome in the First Month of Life.

3.1 Introduction

The microbiome is a term used to describe the whole study environment, which is often a particular site on or within the human body. Of particular interest in the research community is the gut microbiome. Therefore, the gut microbiome encompasses all organisms contained within that environment, the surroundings, such as the gut lumen, and transient components, such as digested food, and finally all the genetic material contributed by each component of the system (Marchesi and Ravel 2015). This symbiotic relationship contributes numerous advantages to the host (Gill *et al.* 2006; Momose *et al.* 2008; Cantarel *et al.* 2012; Tremaroli and Bäckhed 2012).

Colonisation of the gut microbiome begins before birth and continues to develop throughout the first two years of life. From birth until 3 months of age, the infant microbiome is dominated by *Firmicutes*, after 3 months the community is dominated by *Actinobacteria*. This continues until 1 year of age when *Bacteroidetes* and *Firmicutes* are the main constituents (Koenig *et al.* 2011b; Azad *et al.* 2013). At two to three years of life the infant microbiome is beginning to resemble that of an adult. This is a result of the abundance of *Clostridia* becoming predominant alongside the *Bacteroidia* (Avershina *et al.* 2016). Finally, the healthy adult microbiome is dominated by the *Bacteroidetes* and *Firmicutes* phyla, and remains relatively stable until old age (Consortium *et al.* 2012).

The gut microbiome development of preterm infants deviates significantly from the route described above and is a result of numerous factors. Current research indicates the initial colonisation of the preterm infant gut begins with Gram-positive cocci, such as the *Bacilli* family. These initial colonizers are then overtaken by facultative anaerobes, within the *Gammaproteobacteria* class. This then leads to a final strictly anaerobic state (Jacquot *et al.* 2011; Normann *et al.* 2013; Torrazza *et al.* 2013; Sim *et al.* 2015; Zhou *et al.* 2015). Therefore, the development of the gut microbiome of preterm infant's progresses towards a Proteobacteria dominated state (Moles *et al.* 2013).

Furthermore, as a result of the reduced diversity and richness seen in preterms, greater than 90% of all organisms found in the gut of preterm infants predominate the community. Furthermore, the *Gammaproteobacteria* class are proportionally over-represented in preterm infants, often comprising greater than 50% relative abundance. This is in contrast with less than 20% seen in full term infants (La Rosa *et al.* 2014; Ward *et al.* 2016).

Furthermore, as a result of the cumulative exposure to antibiotics in the NICU, overall species and richness and diversity in preterm infants is significantly reduced. In addition, the majority of preterm infant's sampled carried plasmid encoded antibiotic resistance genes for more than six classes of antibiotic (Gibson *et al.* 2016; Ward *et al.* 2016). Not only do treatments received in the NICU contribute to colonization of the gut microbiome, the abnormal environment has been found to play a significant role.

As described previously in Section 1.3.1.2, the NICU is a breeding ground for bacteria, therefore it is not surprising for these environmental bacteria to be seen in the guts of hospitalized infants. One study found that dominant gut taxa, such as *Staphylococcus epidermidis*, *Klebsiella pneumoniae*, *Bacteroides fragilis*, and *Escherichia coli*, were also found in over half of the samples taken from the NICU environment (Brooks *et al.* 2014).

Anomalous to preterm infants is the high inter-individuality seen in this sub population. In contrast to full term infants and adults, the preterm gut microbiome shares no common features or characteristic between individuals (Magne *et al.* 2006). Currently this is no explanation for this phenomenon, but it can be hypothesised to be a result of the NICU environment.

Gestation is not the only factor during infancy to affect the gut microbiome developmental process. The abundance of *Bacteroides* are demonstrated to be reduced in infants delivered by caesarean section (CS). Furthermore, the caesarean delivered infants were shown to have a reduced gut microbiome diversity (Grönlund *et al.* 1999; Jakobsson *et al.* 2014). Furthermore, CS has been found to seed the neonatal microbiome with opportunistic pathogens including species from the *Haemophilus*, *Enterobacter*, *Veilonella* and *Staphylococcus* genera (Dominguez-Bello *et al.* 2010; Bäckhed *et al.* 2015). A recent systematic review provided further evidence to a reduced

diversity in CS delivered infants, and showed that the significant differences observed in the infants during the first 6 months of life had reduced after this time (Rutayisire *et al.* 2016). A vaginal delivery is currently accepted as the more beneficial mode of delivery, as demonstrated in several studies. Infants delivered vaginally are enriched with *Escherichia/Shigella* and *Bacteroides* compared with infants delivered by CS (Azad *et al.* 2013). Furthermore, there is a high level of *Lactobacilli* in vaginally delivered infants (Dominguez-Bello *et al.* 2010; Aagaard *et al.* 2012; Avershina *et al.* 2014). However, mode of delivery has been shown to have no impact on the gut microbiome community in preterm infants (Stewart *et al.* 2017). This is an interesting finding as the birthing procedure is the same for both full and preterm infants, therefore similar differences should be found in both groups of infants.

In addition, the feeding routine influences the type of bacteria acquired during gut microbiome development in both full and preterm infants (Cong *et al.* 2017; Timmerman *et al.* 2017). More specifically, preterm infants fed their mother's own milk had the highest abundance of *Clostridales*, *Lactobacillales* and *Bacillales*. Whereas infants fed primarily donor breast milk or formula had a higher abundance of *Enterobacteriales*. Furthermore, the gut microbiome diversity in infant's fed their mother's own milk was higher compared to the other feeding types (Cong *et al.* 2017).

Location and environment dramatically affect the infant gut microbiome. For instance, the preterm gut microbiome has been shown to reflect that of the NICU (Groer *et al.* 2014). It is believed this to be a result of the handling, feeding and treatment regimens performed in the NICU (Brooks *et al.* 2014). However, there is no evidence to show if this is consistent throughout different NICUs.

Finally, gender has also been shown to be an important driving factor in preterm infant gut microbiome development. For example, the abundance of *Enterobacteriales* is higher in male infants, and numbers of *Clostridales* are increased in females (Cong *et al.* 2016).

One of the primary outcomes of preterm birth is the increased incidence of disease, such as NEC, and results in high rates of neonatal mortality and morbidity (Kosloske 1994; Holman *et al.* 2006; Fitzgibbons *et al.* 2009; Horbar *et al.* 2012). A recent systematic review concluded that NEC is preceded by an increase in the relative abundance of

Proteobacteria and a decreased relative abundance of *Firmicutes* and *Bacteroidetes* (Pammi *et al.* 2017). This result could translate that the *Bacteroidetes* phylum confers a form of "protection" against disease such as NEC. One reason for this could be its ability to confer stability in the gut ecosystem, such that it can utilize host polysaccharides in the absence of dietary ones (Bäckhed *et al.* 2005). Even though there are organismal changes occurring as a result of NEC development, no change in bacterial load has been observed as a result of the disease (Abdulkadir *et al.* 2016b).

3.2 Aims

It can be concluded that the gut microbiome of preterm infant's progresses towards a *Proteobacteria* dominated state. However, it is highly individualistic, and no common characteristics have been discovered. The development of the preterm and full-term gut microbiome is affected by several factors such as, gender and feeding. However, in contrast to full term infants the gut microbiome of preterms remains unaffected by mode of delivery. Gestation has been shown to be a significant factor in the development of the gut microbiome, especially in relation to disease. It has been demonstrated that preterm infants who do not develop NEC have a greater abundance of *Bacteroidetes*, and therefore this phylum has been labelled as 'protective' against the disease. Taking previous research into consideration, the aims of this study were:

1. To determine if the cohort of preterm infants recruited to this study, showed similar results to those previously published. For example, I hypothesise that the gut microbiome of the preterm infants in this cohort will be affected by most clinical factors such as gender and feeding.
2. In addition, I wanted to investigate the occurrence of a *Proteobacteria* dominated state occurred over time and if there were substantial inter individual differences.
3. Finally, I aimed to determine if the reduction in the presence of the *Bacteroidetes* phylum in infants who develop NEC is significant. This will enable us to substantiate this finding, which may then encourage the research community to investigate the mechanisms behind this 'protective' effect.

3.3 Materials and Methods

3.3.1 Determination of the Preterm Gut Microbiome using Illumina MiSeq Sequencing Techniques on DNA Extracted from Stool.

In order to determine the gut microbiome of preterm infants, stool samples were collected as previously described in Chapter 2, Section 2.2.4. These samples were then processed and the DNA extracted according to the methods detailed in Chapter 2, Sections 2.6 and 2.8.2. Sequencing of the 16S rRNA gene by Illumina MiSeq technology was performed by Dr Dagmar Alber at the Child Health Institute, Great Ormond Street Hospital, University College London, as referenced in Chapter 2 Section 2.9. Processing of the raw sequencing data was conducted using the Mothur software package, version 1.39.5.

3.3.2 Analysis of the Preterm Gut Microbiome Data.

The raw sequencing reads were processed using the Mothur software, v1.39.5, as detailed in Section 2.10. As part of the data processing protocol, all samples containing less than 1000 reads were removed, as they were deemed to not accurately reflect the microbial community present in that sample. This was determined by using the coverage value, calculated during the Mothur pipeline, Section 2.10. A coverage value ranges from 0 to 1, with a score of 1 indicating that the microbial community is 100% accurately reflected at that sub sampling threshold. As a result, a value ≥ 0.99 was used as the cut off during this study. Using this value, a sample needed to contain ≥ 1000 reads in order to accurately reflect the microbial community. Therefore, the dataset was subsampled at 1000 reads to create the Microbiome Cohort. The resulting dataset, the Microbiome Cohort, was visualised and analysed as detailed in Section 2.11.

3.4 Results

3.4.1 Determination of the Optimum Starting Material and DNA Extraction Methods for Bacterial DNA from Preterm Infant Stool

The aim of this investigation was to determine the best DNA extraction method and weight of stool to use for this project. There were two methods available at the start of this project, the Qiagen QIAamp DNA Stool Mini Kit and the Promega Maxwell 16 instrument for automated DNA extraction. Established protocols within the lab used a bead beating step prior to DNA isolation with said kits, however stool was tested with and without bead beating during this investigation. As defined in the Qiagen protocol, 200 mg of faecal material was used with the Maxwell protocol.

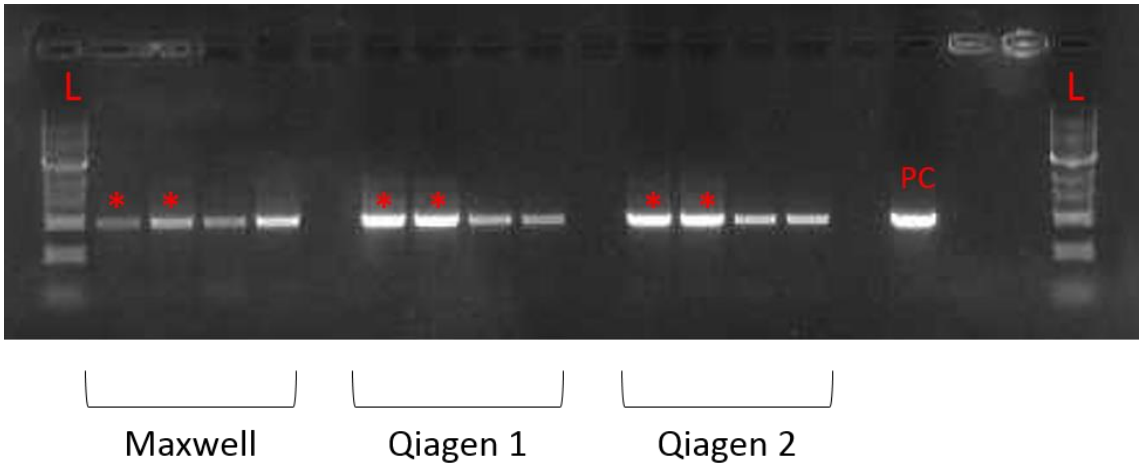

Figure 15. A Comparison of Different DNA Extraction Methods using 16S rRNA Gene PCR. Notations are as follows: L = Ladder (1kb), PC = Positive Control, * = Samples were bead beaten prior to DNA extraction. The samples extracted with the Maxwell instrument are within the bracket labelled "Maxwell", each sample was done in duplicate. The samples extracted by a colleague (Ms. Giulia Masetti) with the Qiagen kit are within the bracket labelled "Qiagen 1". The samples extracted by me, using the Qiagen kit, are within the bracket labelled "Qiagen 2".

From this experiment Figure 15, shows that the Qiagen DNA extraction yielded more bacterial 16S rRNA than the Maxwell protocol. Furthermore, it also showed that the bead beating step improved the total bacterial 16S rRNA yield. As a result, all DNA extractions used a bead beating step and the Qiagen QIAamp DNA Stool Mini Kit.

Next the minimum weight of stool required to extract the highest yield of 16S rRNA, was determined. This was important as many of the samples were less than 200 mg. Figure 16, clearly shows that the weight of stool used during DNA extraction did not

affect the yield of 16S rRNA, thus showing that the starting weight of the stool will not affect the DNA yield. Even a stool weight of 10 mg yielded sufficient DNA. It was decided that 200 mg of stool would be used as standard.

Figure 16. The Effect of Faecal Weight on the Yield of 16S rRNA gene DNA. Notations are as follows: L = Ladder (1 kb), PC = Positive Control. The numbers 1 to 8 represent a series of DNA extractions on stool ranging from 200 mg to 10 mg in weight. Note these samples were collected from a full-term infant. The numbers 9 to 16 represent a series of DNA extractions on stool ranging in weight from 200 mg to 10 mg. Note these samples were collected from a preterm infant. The weights of stool used during the extractions was as follows: 1 and 9 = 200 mg, 2 and 10 = 150 mg, 3 and 11 = 100 mg, 4 and 12 = 80 mg, 5 and 13 = 60 mg, 6 and 14 = 40 mg, 7 and 16 = 20 mg, 8 and 16 = 10 mg.

3.4.2 Illumina MiSeq 16S rRNA Gene Sequencing to Map the Gut Microbiome of Preterm Infants

A total of 210 DNA extractions, extracted by me, were sent to our collaborators, Dr Dagmar Alber and colleagues. As detailed in Section 2.9, after PCR amplification on samples with a DNA concentration greater than 0.5 ng/μl were taken forward to sequencing. This resulted in 194 (92%) of samples being returned with sequence data to be processed in the Mothur pipeline and other downstream packages, as detailed in Section 2.11.1. During the Mothur pipeline a number of samples were removed from the dataset, as they contained less than 1000 reads. A threshold of 1000 reads was used as this was the lowest number possible to achieve a coverage level of ≥ 0.99. The aim of this study was to determine as much detail on the colonisation of the preterm gut microbiome, therefore a high coverage was necessary to ascertain the most detailed results possible. Thus, 151 (72%) of samples were included in the analysis of the preterm gut microbiome, the Microbiome Cohort.

Table 12. Number of Samples Collected by Day of Life. Data included in this are the number of samples that were sent for sequencing (Sent); the number of samples that were returned with sequencing data (Sequenced); the percentage of the samples sequenced and included in the Microbiome Cohort.

Samples	Day of Life						Total
	1 to 5	**6 to 10**	**11 to 15**	**16 to 20**	**21 to 25**	**> 25**	
Sent	31	49	37	23	35	35	**210**
Sequenced	23	45	35	22	35	34	**194**
Percentage	74%	92%	95%	96%	100%	97%	**92%**
Analysed	10	35	28	19	30	29	**151**
Percentage	32%	71%	76%	83%	86%	83%	**72%**

The quality of the sequencing data produced can be a very informative result, indicating the overall quality of the techniques and methods used. This information can be found in Table 13. Observing the controls first, it can be seen that the numbers are vastly lower than that for the samples, the only exception being the minimum values. Overall, this shows that the DNA extraction method was successful and minimal amounts of contamination was introduced during this process. The organisms present in these samples was assessed and determined to not affect the results and were removed from the Microbiome Cohort. The number of reads generated in the samples vastly outnumber those seen in the controls, further validating the results observed to be biologically relevant. In addition, the high mean and median show that overall the sequencing was successful and able to produce a high number of reads. Therefore, this should be accurate regarding the bacterial content of the preterm infant gut microbiome. However, due to the low biomass seen in samples from this section of the population, it is not unexpected to find a minimum number of reads to be one. Actually, this is more likely due to reaction failure rather than low biomass or DNA content, as these samples will have already been removed prior to sequencing, under the threshold mentioned in Section 2.9. The high standard deviation score seen is a result of the low read numbers, in contrast, it could be due indication of the variety in the microbiome data seen in Section 3.4.6. As will be discussed in Section 3.6.2, the variation in read numbers per sample made it difficult to choose an accurate number of reads in which to subsample the data.

Table 13. Simple Statistics for the Number of Reads Generated by Illumina MiSeq 16S rRNA Gene Sequencing. This table provides a comparison of simple statistics between the samples and controls.

Statistic	Samples (n=194)	Controls (n=5)
Minimum	1	5
Maximum	481757	206
Median	45314	64
Mean	59329	89
Standard Deviation	65829	80

The simple statistics shown in Table 13 are relevant but are limited in presenting the spread of data in the number of reads generated per sample. Therefore, a box and bee swarm plot can be seen in Figure 17, and clearly shows the dispersal around the mean. Without visualising this data, it would be impossible to see the high number of samples clustered towards the bottom of the graph. This shows that a high number of samples did not sequence very successfully, and indicates as to why the mean is low, as these low numbers are skewing the mean. Importantly, it shows that the controls, coloured in green, cluster at the bottom of the graph, confirming they are low in number and why they were not included in downstream analysis. Ignoring the lower values, there is a good spread of samples, further implying a diversity in microbiome communities between individuals and samples.

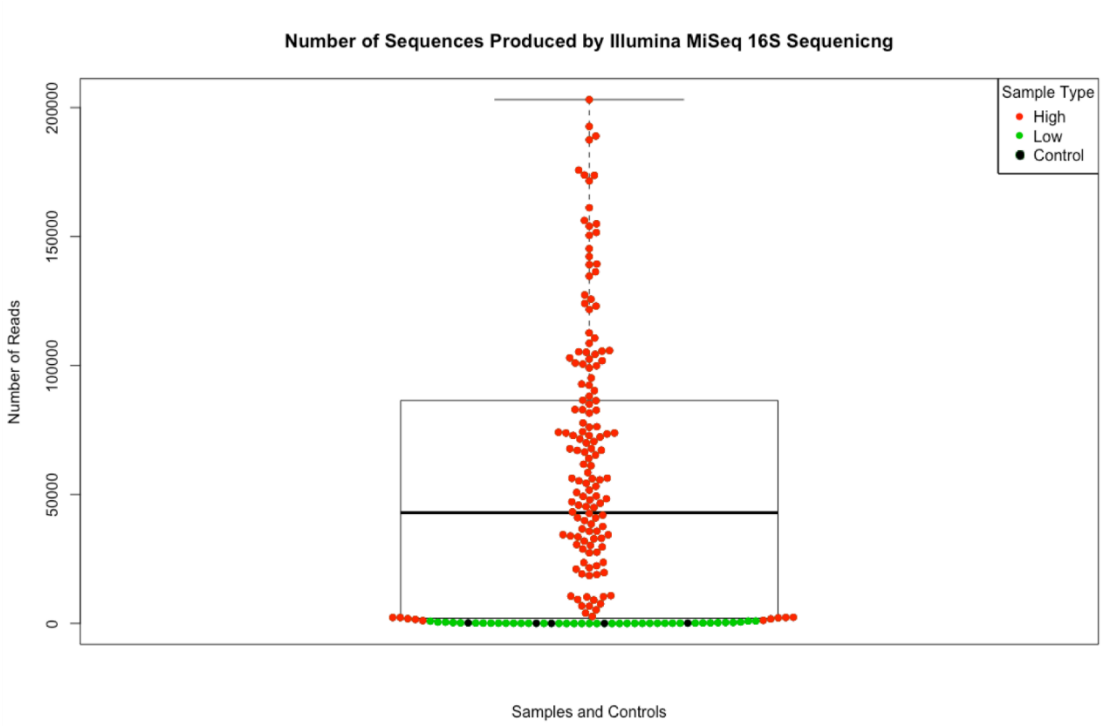

Figure 17. The Number of Reads Generated by the Controls, Study and Microbiome Cohorts. This box and bee swarm plot includes all positive samples that were included in the Microbiome Cohort (Red); all samples that were removed due to containing less than 1000 reads (Green); and the DNA extraction controls (Black). The box plot underneath the swarm points shows the mean, upper and lower quartiles, and the minimum and maximum values.

To summarize, the number of reads generated by the sequencing was high and will provide an accurate representation of the bacterial content of the preterm infant gut microbiome. Furthermore, the Illumina MiSeq process was successful and only produced minor errors.

3.4.3 Comments on the DNA Extraction Controls and Sub-Sampling

DNA extraction controls were used to ensure the process had been carried out in a manner that introduced little or no bacterial contamination. The organisms present in Figure 18 were very similar to those present in Microbiome Cohort, for example *Klebsiella*, *Escherichia*, *Staphylococcus*, and *Enterococcus*, were present in all samples and controls. However, the number of reads produced during sequencing were substantially lower than those included in the Microbiome Cohort. Taking into account the similarities in community structure and the dramatic reduction in numbers, it was decided that these sample would be removed from the Microbiome cohort. Due to the presence of similar organisms present in both the extraction controls and samples, it was not possible to remove the organisms found in the controls from the sample

communities. The predominance of the organisms, seen in the controls, within the sample communities demonstrated that these were integral components of the microbial community, rather than artefacts of the DNA extraction methods. Therefore, the decision to acknowledge and report, but removed was chosen.

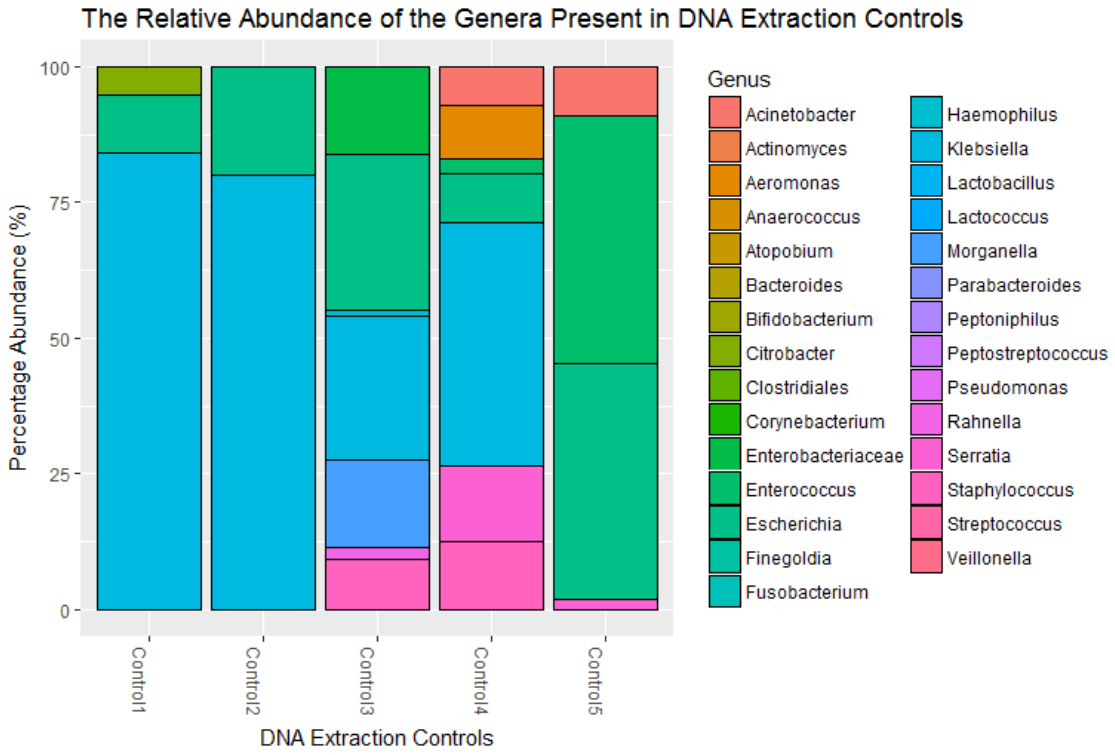

Figure 18. The Relative Abundance of the Genera Present in the DNA Extraction Controls. The relative abundance of each organism within a sample is shown by the y axis with each colour denoting a different organism. The number of reads in each control is as follows: Control 1 = 38, Control 2 = 5, Control 3 = 87, Control 4 = 184, and Control 5 = 55.

3.4.4 Patient Recruitment and Sample Demographics

The aim of this analysis was to determine if there were any simple differences between the infants who provided samples that sequenced successfully and those who did not. Furthermore, it provided details on the infants recruited described as a cohort. This allowed analysis to be conducted to determine if there was any bias during the recruitment process, such as more males than females recruited to the study.

Table 14 shows the demographics for the cohort before and after sequencing. The complete cohort is referred to as the Study Cohort and the successfully sequenced cohort are referred to as the Microbiome Cohort. As mentioned in Section 3.3.2, as part of the 16S rRNA gene sequencing data processing, a number of samples were removed due to containing less than 1000 reads. This was a result of the coverage level, as any samples with a 1000 reads or more gave a coverage of more than 0.99. As a result, the only samples included in the Microbiome Cohort were those containing ≥ 1000 reads. As demonstrated in Table 2 all samples collected form infants residing on the post-natal ward (PNW) did not generate a coverage value ≥ 0.99 and were therefore removed from the Microbiome Cohort.

Overall, from Table 14, it can be seen that not all of the samples used during this study were sequenced adequately in order to be included in the Microbiome Cohort, there are numerous explanations for this detailed in Section 7.1 (Chapter 7). In total there were nine infants without sequencing data. However, there was only one infant for which all of the sample provided, a total of 4, which were not included, this was infant 19.

Not included in the table is data referring to the average gestational age between males and females. When calculated there is very little difference. The average gestational age for males is 27 weeks compared to 25 weeks for female preterm infants.

Table 14. Table Containing the Patient and Sample Demographics. Abbreviations are as follows: Post Natal Ward (PNW), University Hospital Wales (UHW), North Bristol Trust (NBT), Caesarean Section (CS), and Necrotising Enterocolitis (NEC). The Study Cohort includes all infants recruited to the study and their associated demographics. However, the Microbiome Cohort includes only the samples that produced more than 1000 reads during sequencing and were taken forward to analysis.

	Study Cohort				Microbiome Cohort		
	Preterm	**Term**	**PNW**	**Total**	**Preterm**	**Term**	**Total**
N							
Infants	52 (90%)	3 (5%)	3 (5%)	58 (100%)	49 (96%)	2 (4%)	51 (100%)
Samples	199 (95%)	8 (4%)	3 (1%)	210 (100%)	140 (96%)	6 (4%)	146 (100%)
Sex							
Male	33 (64%)	2 (67%)	2 (67%)	37 (64%)	29 (62%)	1 (50%)	30 (61%)
Female	19 (37%)	1 (33%)	1 (33%)	21 (36%)	18 (38%)	1 (50%)	19 (39%)
Sample Site							
UHW	18 (35%)	3 (100%)	3 (100%)	24 (41%)	15 (32%)	2 (100%)	17 (35%)
NBT	34 (65%)	0 (%)	0 (0%)	34 (59%)	32 (68%)	0 (0%)	32 (65%)
Gestation (Complete Weeks)							
Mean ± Standard Deviation	26 ± 2.5	38 ± 2.3	40 ± 1	NA	26 ± 2.5	37 ± 0	NA
Birthweight (grams)							
Mean ± Standard Deviation	869 ± 309.8	2783 ± 752.2	3640 ± 471.6	NA	873 ± 324	2350 ± 71	NA
Length of Hospital Stay (Days)							
Mean ± Standard Deviation	102 ± 45.4	23 ± 11.8	2 ± 0	NA	102 ± 40	28 ± 11	NA
Survival							
Yes	46 (89%)	3 (100%)	3 (100%)	52 (90%)	44 (94%)	2 (100%)	46 (94%)
No	6 (11%)	0 (0%)	0 (0%)	6 (10%)	3 (6%)	0 (0%)	3 (6%)
Delivery							
Vaginal	29 (56%)	3 (100%)	2 (67%)	24 (50%)	27 (57%)	2 (100%)	29 (59%)
CS	23 (44%)	0 (0%)	1 (33%)	24 (50%)	20 (43%)	0 (0%)	20 (41%)
Developed NEC							
Yes	13 (25%)	0 (0%)	0 (0%)	13 (22%)	11 (23%)	0 (0%)	11 (22%)
No	39 (75%)	3 (100%)	3 (100%)	45 (78%)	36 (77%)	2 (100%)	38 (78%)
Maternal Age							
Mean ± Standard Deviation	29 ± 5.7	22 ± 5.0	37 ± 3.1	NA	29 ± 6	20 ± 4	NA

There was no change in average gestation and maternal age between the Study and Microbiome Cohorts. The birthweight of the preterm infants increases slightly in the Microbiome Cohort. The length of hospital remained static and there was a decrease in infants from the survival and NEC groups in the Microbiome Cohort.

As a result of the reduction in samples from term infants in the Microbiome Cohort, it was decided that it would be difficult to make meaningful comparisons between the microbiome of term and preterm infants. Therefore, the term samples were removed from the final data set and the focus of this thesis became to report to the development of the preterm gut microbiome. To confirm the number of samples included in later analyses using 16S rRNA gene sequencing data was 140, unless otherwise stated.

3.4.5 qPCR for Total Bacterial Load in the Stool of Preterm Infants

In contrast, to the later analysis of the bacterial community in the microbiome, total bacterial load was determined for all samples collected, the Study Cohort. Unless otherwise stated the results shown in this section included all 210 samples that were collected. The aim of this experiment was to determine the total bacterial load of preterm infant stool and compare this to the bacterial community present in the stool and clinical factors such as mode of delivery and feeds.

Table 15. Simple Statistics for the Total Bacterial Load of Preterm Infant Stool. Total bacterial load was calculated and given as copies of the 16S rRNA gene per mg of stool. Copies of the 16S rRNA gene were determined using a standard curve produced during each reaction of a known copy number of *E. coli* genomic DNA. This number was then divided by the mg of stool used during the DNA extraction protocol. 199 samples were included.

Statistic	Copies of 16S Gene/mg of Stool
Minimum	0
Maximum	368,800,000
Mean	6,861,649
Standard Deviation	28,689,414
Median	420,000
25^{th} Quartile	12,773
75^{th} Quartile	5,631,100

From Table 15, there is a large variation in the total bacterial load from preterm infant stool. However, on average there is a substantial bacterial load contained within the stool of preterm infants. As a result, it was determined that 16S rRNA gene Illumina MiSeq sequencing was a viable option for determination of the bacterial community

within these samples. Moreover, associations between the total bacterial load and clinical factors could be investigated.

Firstly, to reflect on the total bacterial load and sequencing results, it was found that the mean bacterial load significantly, $p < 0.01$, affected the positivity of sequencing results. This can be seen in Figure 19. A major reason for this difference is detailed in a later chapter, Chapter 5. The main result, in terms of future preterm microbiome research, is that samples with a total bacterial load of less than 2,000,000 copies per mg of stool, should not be considered for high through put sequencing as they are more than likely to give a negative result. Therefore, this would save time and money in future research.

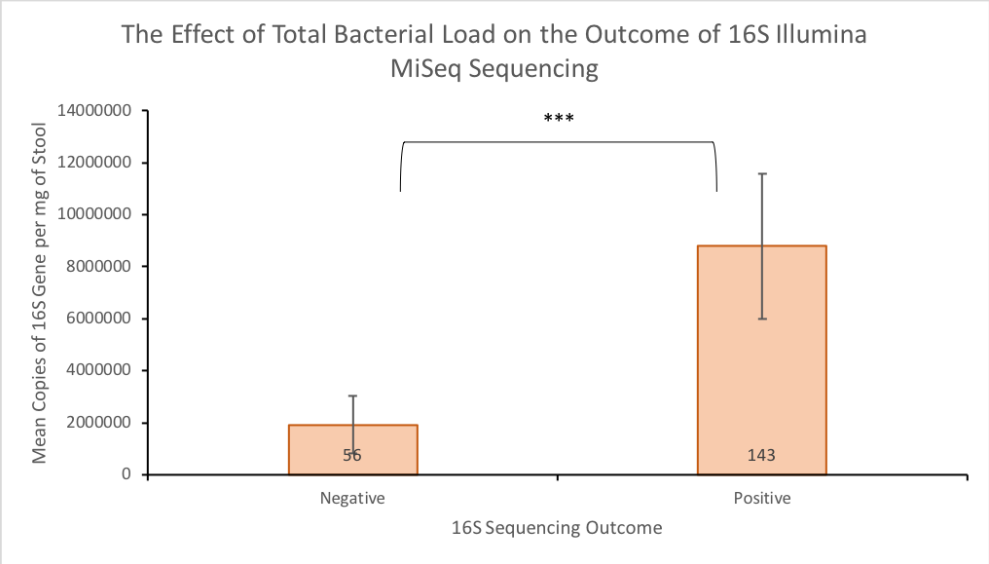

Figure 19. The Effect of Total Bacterial Load on the Outcome of 16S rRNA gene Illumina MiSeq Sequencing. Significance was tested using an independent samples Mann-Whitney U test in SPSS, and significance was shown by the *. Number of samples included in each group is given by the number at the base of each bar. The error bars are the standard error of the mean (SEM).

The design of this study allowed for investigations into the development of the preterm gut microbiome over the first 30 days of life. As a result, the total bacterial load over this time was determined, Figure 20. The lowest total bacterial load was present at the earliest days of life, with the highest followed very soon after at days 6 to 10 life. The total bacterial load steadily decreased until 20 days of life, where it peaked again until a minor decrease after 25 days of life. Therefore, the total bacterial load during the first 30 days of life is not stable and peaks at 6 to 10 days of life.

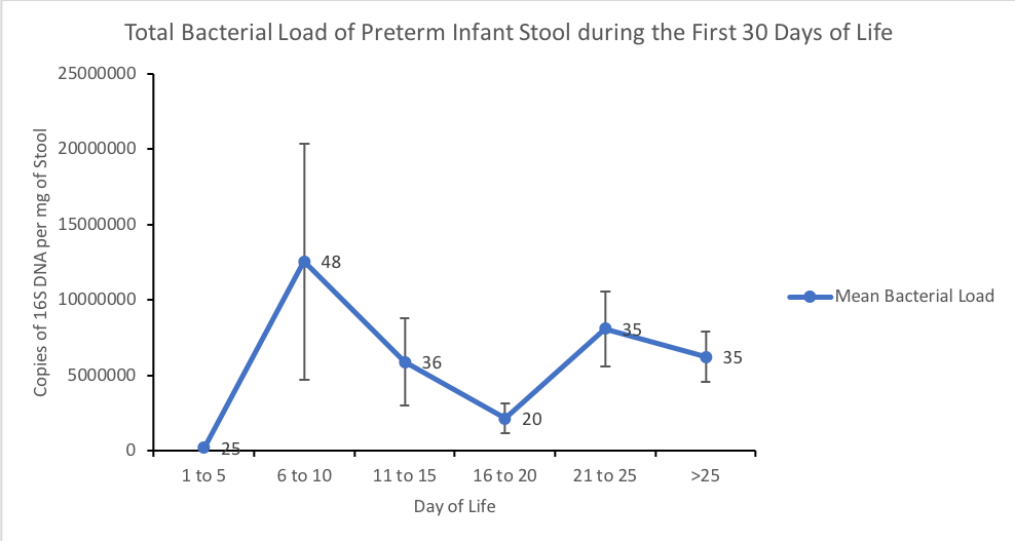

Figure 20. Total Bacterial Load of Preterm Infant Stool during the First 30 Days of Life. The error bars represent the standard error of the mean. The number callouts are the number of samples used to calculate the mean for that day of life grouping.

One of the aims of this research was to determine if there was a difference in gut microbiome development in infants from separate NICUs. There was no significant difference between samples collected from the two NICUs. However, Figure 21 does shows that the mean bacterial load for samples taken at UHW is higher than that of NBT. Furthermore, the range of bacterial loads from UHW varied greater than those taken from NBT. Therefore, sampling site does not significantly affect bacterial load.

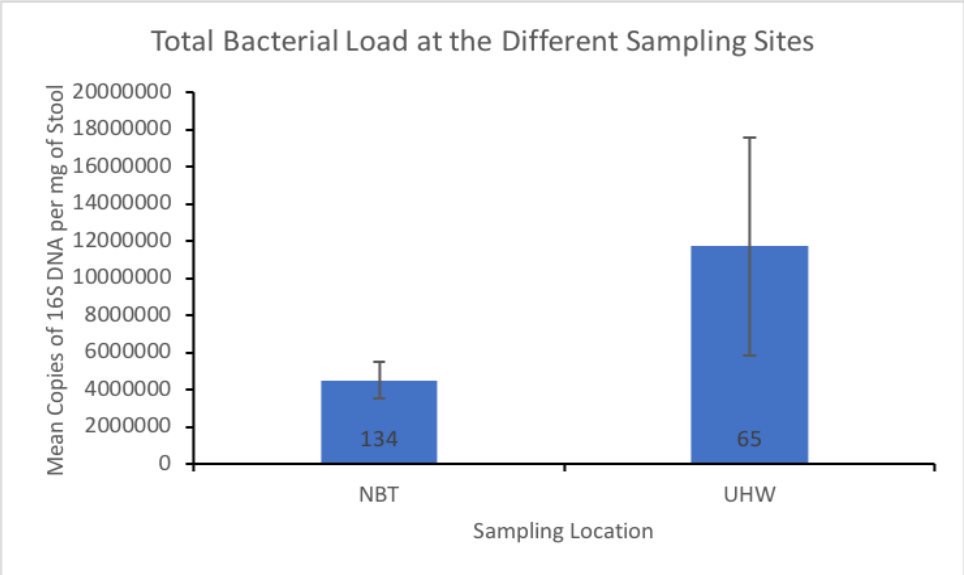

Figure 21. A Comparison of Total Bacterial Load from Samples Collected at Different NICUs. The error bars are the standard error of the mean. The number at the bottom of the bars are the number of samples collected from each site. Abbreviations are as follows: NBT (North Bristol Trust) and UHW (University Hospital Wales).

The sampling sites differed by location but also by the administration of probiotics. During the sampling period only UHW was administering probiotics to the recruitment population, and then only to infants who met the minimum feeding criteria. As a result, probiotics and their effect of total bacterial load was investigated. There was no significant differences in mean total bacterial load as a result of gestational age, a Mann-Whitney U test was performed, as seen in Figure 22. As seen in Figure 22, there is a large variation in total bacterial load in the samples taken in the presence of probiotics. Therefore, probiotics does not significantly affect total bacterial load of preterm infant stool.

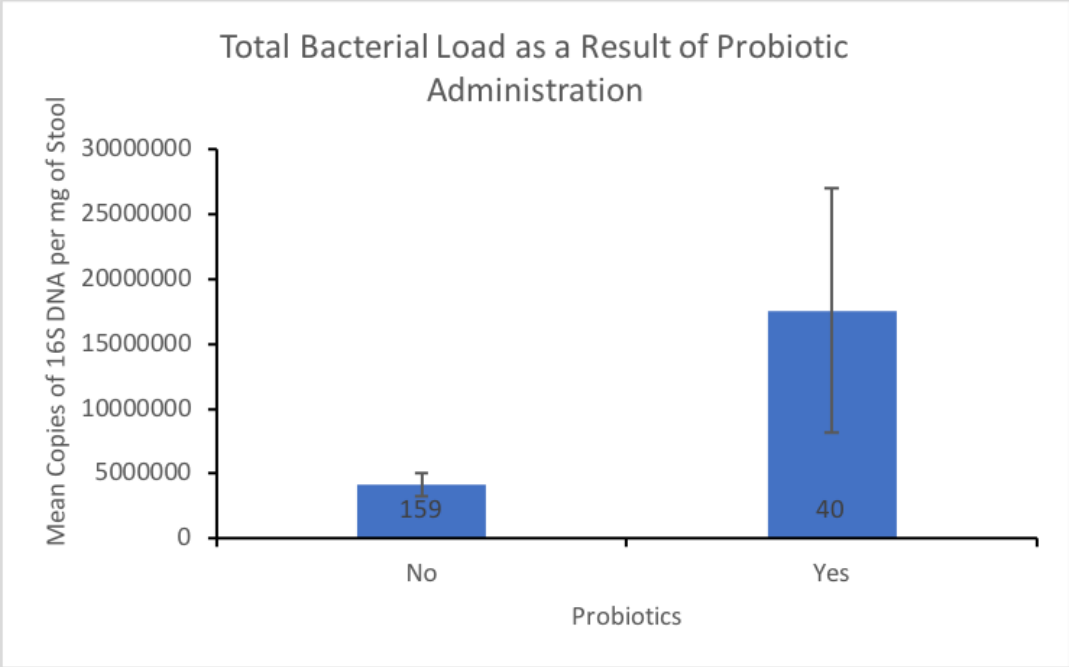

Figure 22. The Effect of Probiotics on Total Bacterial Load of Preterm Infant Stool. For this graph the error bars are given as the standard error of the mean and the number of samples included in each group are given at the bottom of each bar.

It was shown that the trend in total bacterial load was increasing over time, therefore the effect of gestation on bacterial load was also investigated. Figure 23, showed that there were no significant differences in mean total bacterial load as a result of gestational age, a Mann-Whitney U test was performed. In contrast, there was a decrease in mean total bacterial load in the samples taken from full term infants. Once again, there is a large range in total bacterial load taken from the extremely preterm group. This indicates that full term infants harbour a lower total bacterial load compared to full term infants, but due to the limited number of full-term infants recruited to this study, this remains inconclusive.

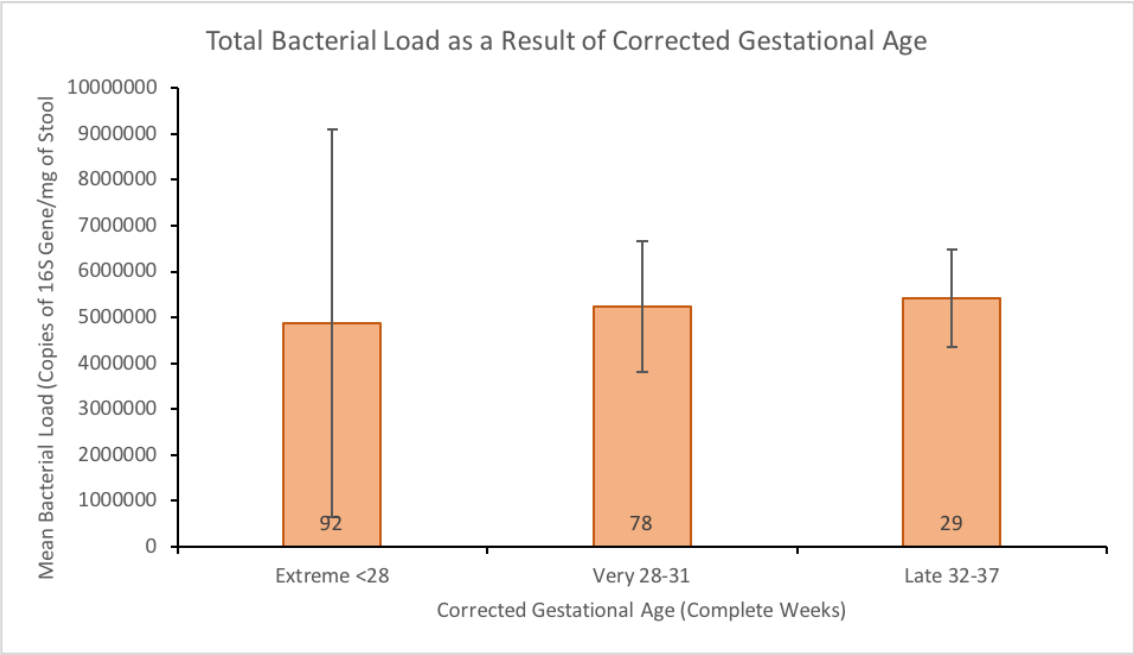

Figure 23. A Comparison between Corrected Gestational Age and Total Bacterial Load. Again, the error bars are included to show the standard error of the mean and the numbers at the bottom of each bar are the number of samples included in each group.

During sample collection if the attending clinician had any concerns for sepsis in the infant, this was noted as part of the clinical data. It was found that there was no significant difference in the total bacterial load of preterm infant stool taken from infants with and without clinical sepsis concerns. However, as shown in Figure 24, there is a decrease in bacterial load from the infants with no reported sepsis concerns, and a large variation in total bacterial load from the infants who were noted as possibly septic. Therefore, this indicates at a possible role for specific bacteria in the pathology of sepsis rather than the overall bacterial load.

As detailed previously feeding type has a dramatic effect on the bacterial community within both full and preterm infant stool. As a result, this variable was analysed in this cohort to determine if these findings were repeatable. As shown in Figure 25, there was a substantial increase in bacterial load in samples taken from infants who were exclusively fed a preterm formula diet. The bacterial load in preterm infant stool taken from infants fed either mother's own milk or a donor breast milk was highly comparable. Finally, to be expected, the lowest total bacterial load was seen in infants receiving no enteral feeds. These results suggest that the administration of a preterm infant formula as the primary diet component increases faecal bacterial load in preterm infants.

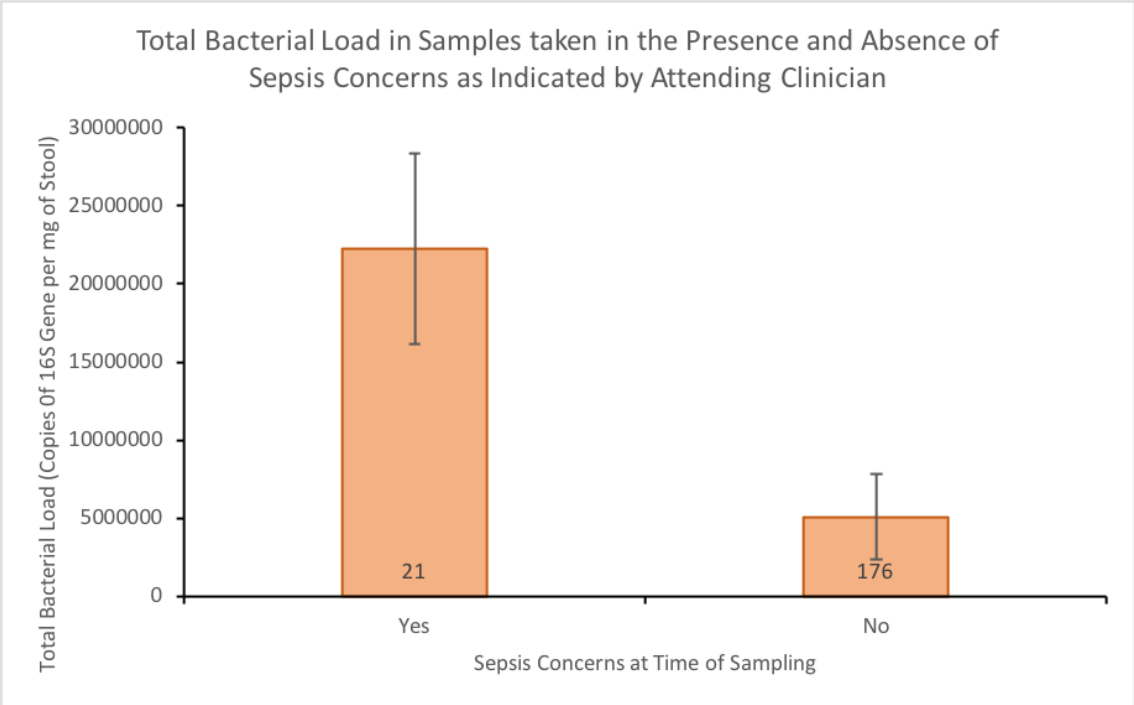

Figure 24. A Comparison in Bacterial Load taken from Infants who did and did not Present Sepsis Concerns at the Time of Sampling. Again, the error bars are included to show the standard error of the mean and the numbers at the bottom of each bar are the number of samples included in each group.

The relationship between gender and bacterial load was also investigated as part of this study. There was a large increase in total bacterial load from male preterm infant stool compared to females, Figure 26. However, this was found to not be significant. Yet the difference in total bacterial load between the genders is an interesting finding, that is yet to be explained.

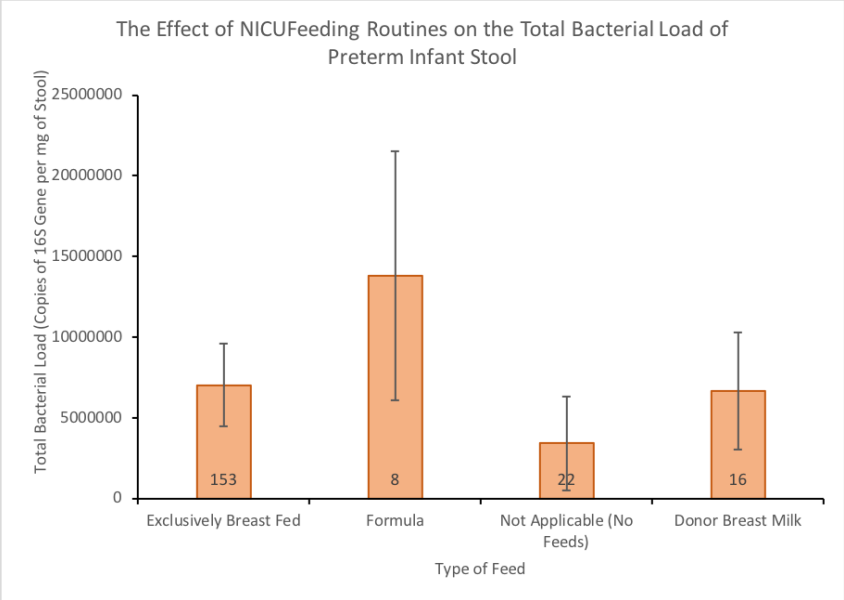

Figure 25. A Comparison between the Total Bacterial Loads in Preterm Infant Stool as a Result of Diet. The error bars are included to show the standard error of the mean and the numbers at the bottom of each bar are the number of samples included in each group. When tested it was not significant.

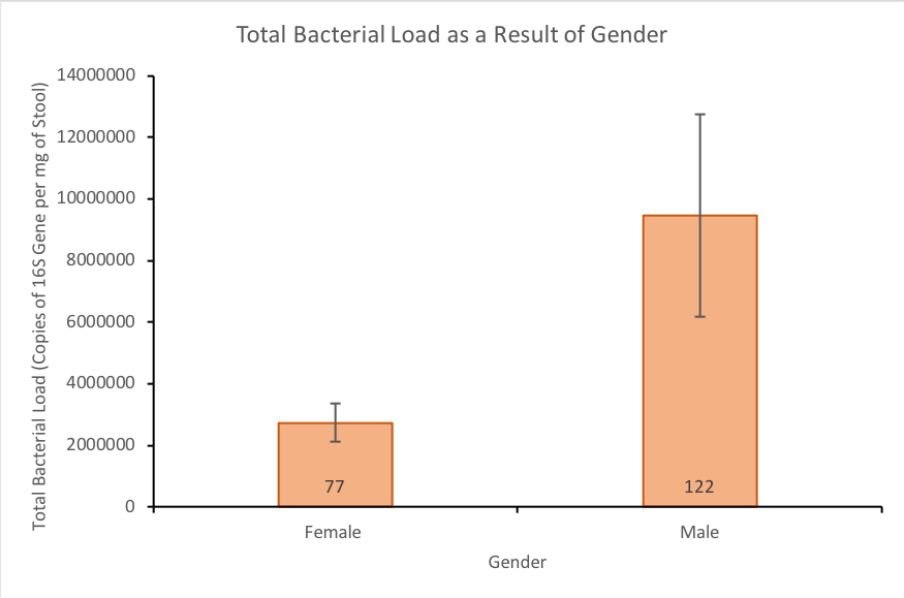

Figure 26. A Comparison between Total Bacterial Load in Female and Male Infants. The error bars included to show the standard error of the mean and the numbers at the bottom of each bar are the number of samples included in each group. When tested it was not significant.

As detailed in Section 1.2.4.3.1, the development of the gut microbiome is influenced by the mode of delivery. This difference was conformed in this study. There is a significant increase, $p = 0.045$, in total bacterial load in the stool of infants who are delivered by caesarean section (CS), as shown in Figure 27. Further analysis is needed to determine if this is the result of the difference in bacterial content encountered via the different mode of delivery.

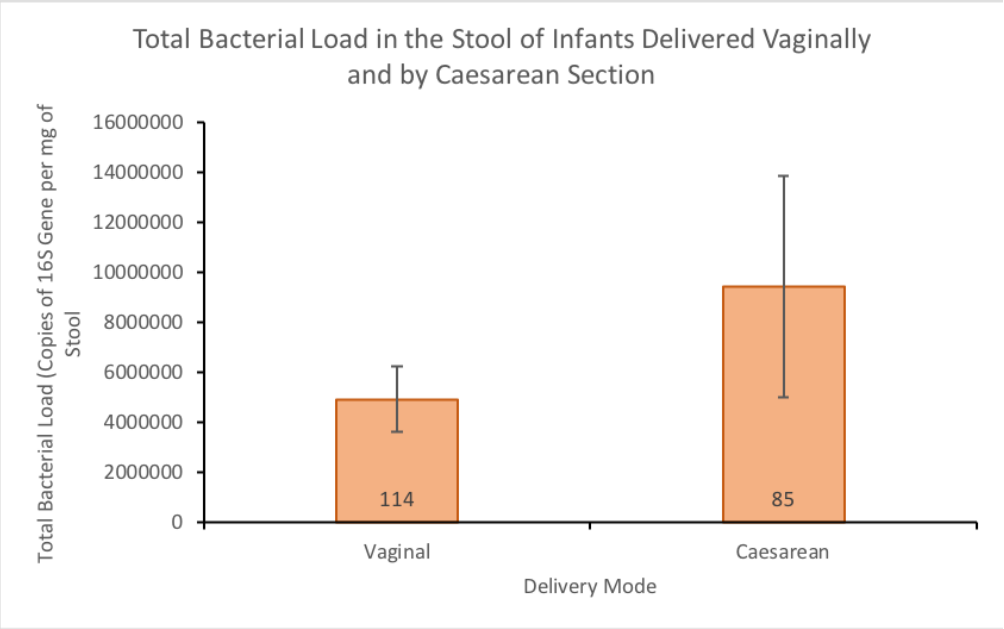

Figure 27. A Comparison of Total Bacterial Load in the Stool of Infants Delivered Vaginally and by Caesarean Section. The error bars are included to chow the standard error of the mean and the numbers at the bottom of each bar are the number of samples included in each group.

NEC is one of the primary causes of neonatal mortality and morbidity and has been linked to changes in the gut microbiome. Therefore, it was important to determine if there was a change in bacterial load in infant stool prior to the development of NEC. There was a substantial increase in total bacterial load in samples taken from infants who developed NEC, but this was not determined to be significant, as shown in Figure 28. Furthermore, there was a greater variety in total bacterial load in stool from infants who did develop NEC compared to those who did. This indicates that unlike bacterial community structure in the microbiome the total bacterial load is not linked to the development of NEC.

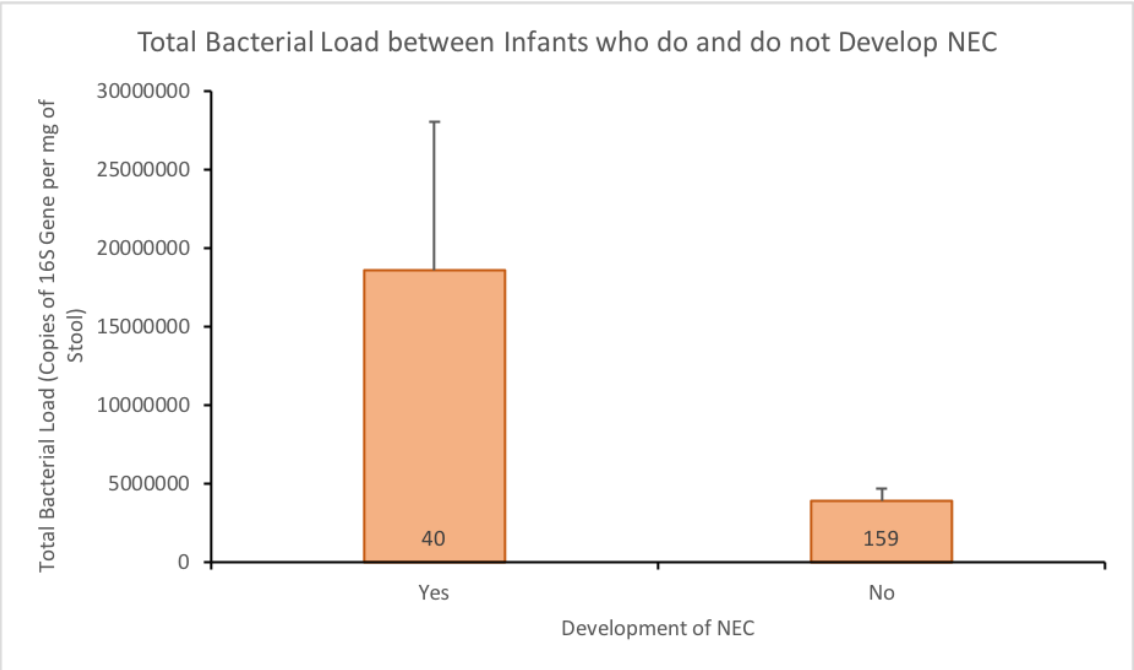

Figure 28. A Comparison between the Total Bacterial Load in Infants who did and did not Develop NEC. The error bars are included to show the standard error of the mean and the numbers at the bottom of each group.

The length of an infant's stay in hospital in determined by their health, and so length of hospital stay has the potential to act as a proxy for health. Therefore, the effect of the duration of hospital stay on the total bacterial load was determined. The bacterial load was higher in infants whose hospital stay was greater than 3 months, as shown in Figure 29. However, this was not significant. This indicates that an increase in faecal bacterial load during the first month of life may affect the length of hospital stay, and as a result the possibility of reduced health.

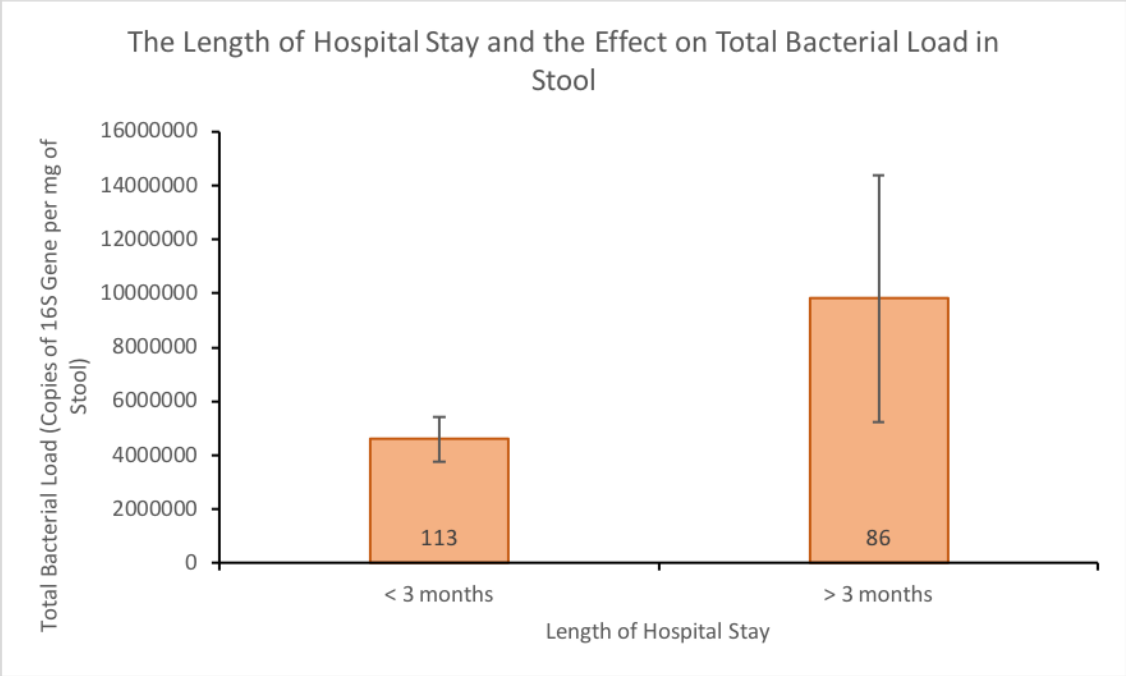

Figure 29. A Comparison between Length of Hospital Stay and Total Faecal Bacterial Load. The error bars are included to show the standard error of the mean and the numbers at the bottom of each bar are the number of samples included in each group.

Directly related to an infant's hospital stay is there survival to hospital discharge, as shown in Figure 30, there is a clear link between bacterial load and survival. A significant increase, $p = 0.028$ in faecal bacterial load is associated with survival to hospital to discharge. This indicates that bacterial colonisation of the gut is essential for survival in this population.

Antenatal steroids are often given to women threatened by preterm delivery, as a result a large portion of the infants recruited to this study were prenatally exposed to these drugs. Therefore, it was interesting to determine if antenatal steroids had an effect on the infant microbiome. As shown in Figure 30, antenatal steroids had no significant effect upon the total faecal bacterial load.

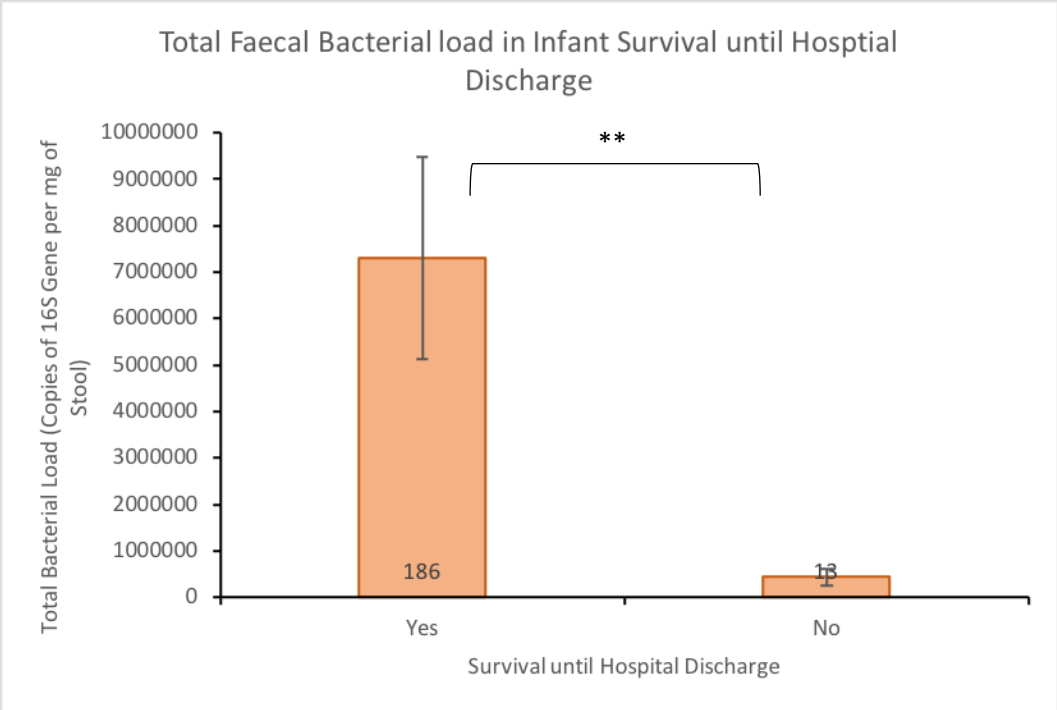

Figure 30. A Comparison of Total Faecal Bacterial Load in Infants who did and did not survive until Hospital Discharge. The error bars are included to show the standard error of the mean and the numbers at the bottom of each bar are the number of samples included in each group.

In order to determine if the dominance of the microbiome by a certain organism was linked to total bacterial load, Figure 15 was constructed. As can be determined from Figure 31, there was no association between a high bacterial load and the dominance of the gut microbiome by one organism. This indicates that total bacterial load is not affected or influenced by the dominance of one particular organism or set or organisms. In other words, the gut microbial community changes in response to the increased growth of an organism. Furthermore, this suggests that there have been decreases in other members of the community in order to accommodate this dominance by one specific organism or set of organisms. Otherwise, there would be an increase in bacterial load as a result of the dominance of the microbial community by a particular organism.

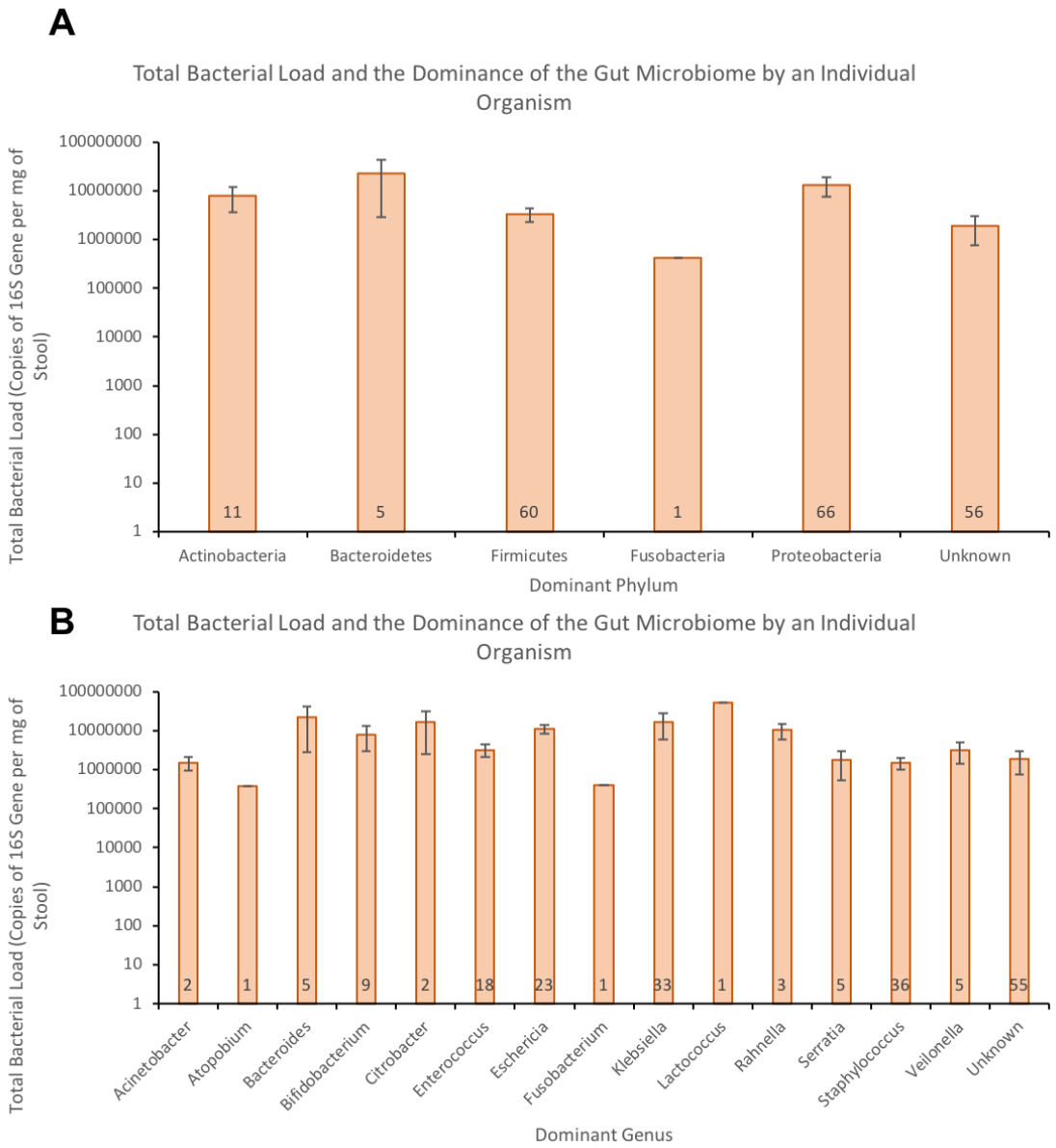

Figure 31. Total Bacterial Load and the Dominance of the Gut Microbiome by an Individual Organism. The error bars represent the standard error of the mean for the mean bacterial load represented by each bar. The numbers at the bottom of each bar are the number of samples dominated by the particular organism. Panel A shows the mean bacterial load for each group dominated by a certain genus. Panel B shows the mean bacterial load for each group of samples dominated by a certain genus.

3.4.6 The Development of the Preterm Gut Microbiome during the First Month of Life.

The Microbiome Cohort of infants showed an overall trend towards a *Proteobacteria* dominated community during the sampling period, Figure 32. From birth until approximately 1 week of age, the majority of infants are shown to have a gut microbiome community dominated by organisms from the *Firmicutes* phylum. After this milestone the microbial community in the majority of infants shifts towards being dominated by organisms from the *Proteobacteria* phylum. Also, of note from Figure 16, is that the *Actinobacteria* phylum appears at varying levels of dominance during the first 30 days of life in preterm infants. Furthermore, the *Fusobacterium* phylum appears in only 2 samples, in complete dominance in one, during the very first days of life, 1 to 5. Finally, the presence of the *Bacteroidetes* phylum decreases in presence and dominance during the first 30 days of life in the preterm infant.

These results translate into the genus taxonomic level. The *Firmicutes* dominating the preterm infant microbiome at the beginning of life are *Staphylococcus*. The *Proteobacteria* dominating the preterm gut microbiome at the end of the first month of life are *Escherichia* and *Klebsiella*. Therefore, the bacterial community of the preterm infant gut transforms from a *Staphylococcus* dominated environment to an *Escherichia* and *Klebsiella* dominated community. The varying appearance of the *Actinobacteria* phylum translates into the *Bifidobacterium* genus and the decreasing presence of the *Bacteroidetes* phylum is a gradual decrease in the presence of *Bacteroides* during the first month of life.

In contrast to the overall progression to a *Proteobacteria* dominated microbiome, infants can demonstrate individual patterns very different to that of the cohort in general. This is demonstrated in Figure 33.

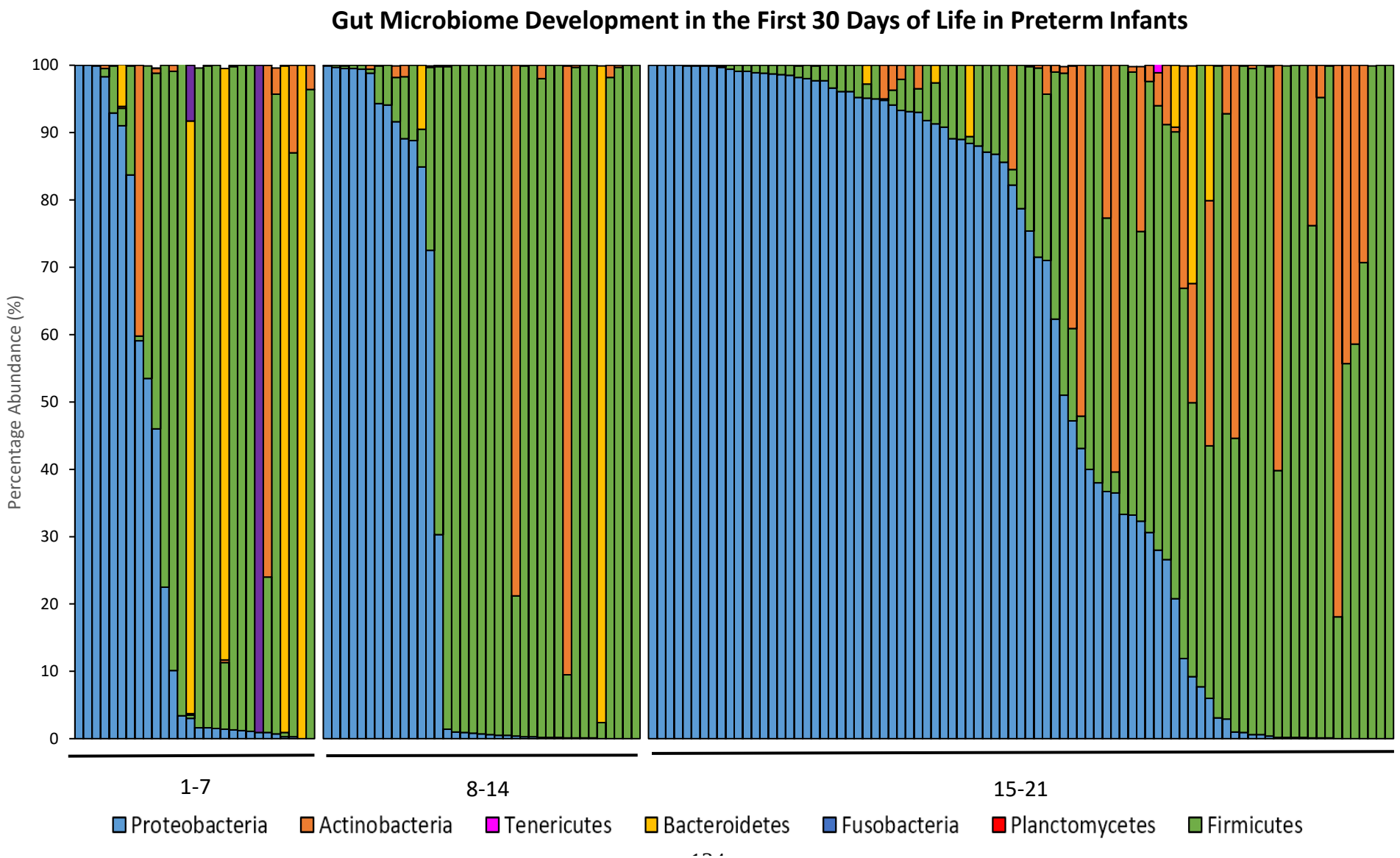

Figure 32. The Development of the Gut Microbiome Community in the First 30 Days of Life in Preterm Infants. The y axis shows the percentage abundance of each organism from the respected phylum in that sample, where each sample is represented by a bar on the chart. The x axis has been split to show the samples that are included in each day of life category. The samples have been ordered by both day of life and the *Proteobacteria* phylum. This enabled for the visualisation of the trend towards a *Proteobacteria* dominance over time.

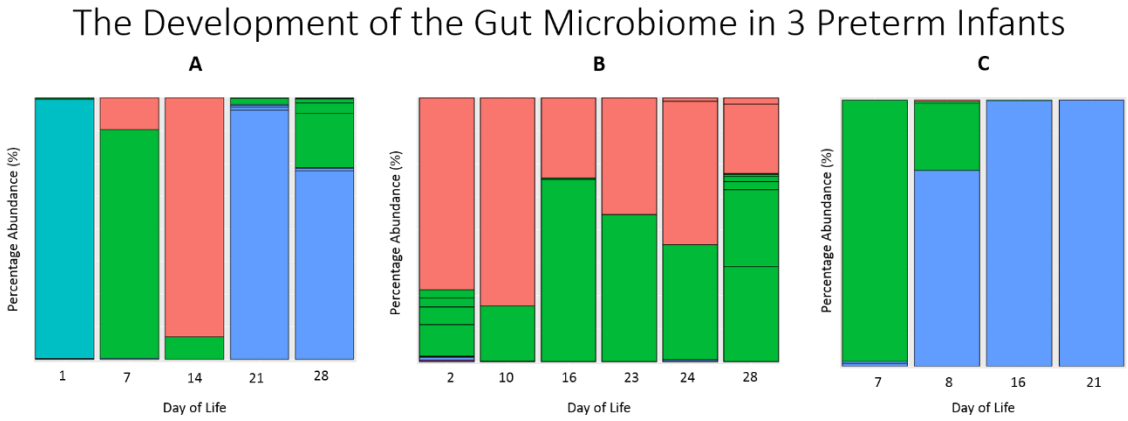

Figure 33. The Individualistic Nature of Gut Microbiome Development in Preterm Infants. Graphs A, B and C represent an individual infant and each bar a sample taken at the noted day of life on the x axis. The y axis represents the percentage abundance, to a total of 100%. The colours represent the following organisms: Cyan = *Fusobacteria*, Green = *Firmicutes*, Red = *Actinobacteria*, Blue = *Proteobacteria*.

Firstly, infant A of Figure 33, clearly shows the trend towards *Proteobacteria* dominance over time, however the dominance of *Firmicutes* at the beginning of life does not occur until day 7 of life. Secondly, infant B of Figure 33, never develops a *Proteobacteria* dominance at the end of 30 days of life. There is an exchange of dominance between the *Firmicutes* and *Actinobacteria* phyla. Lastly, infant C of Figure 33, clearly shows the majority pattern of a *Proteobacteria* dominated phylum towards the end of the first month of life. Moreover, infant C also demonstrates the highly changeable nature of the preterm infant gut microbiome. For instance, on day 7 of life the gut microbiome is dominated by organisms from the *Firmicutes* phylum. However, 24 hours later this has shifted to a *Proteobacteria* dominance.

Figure 34 shows a progression from *Staphylococcus* dominated to a more *Escherichia* and *Klebsiella* dominated one. Moreover, this figure further demonstrates the dominant nature of the preterm gut microbiome.

In summary, the gut microbiome of preterm infants is individualistic and subject to dramatic change within a short amount of time. However, when observed as a group the development of the gut microbiome community begins with a dominance of *Firmicutes*, this then progresses towards a *Proteobacteria* dominated state during the first 30 days of life. Other key contributing organisms during development are *Actinobacteria* and *Bacteroidetes*.

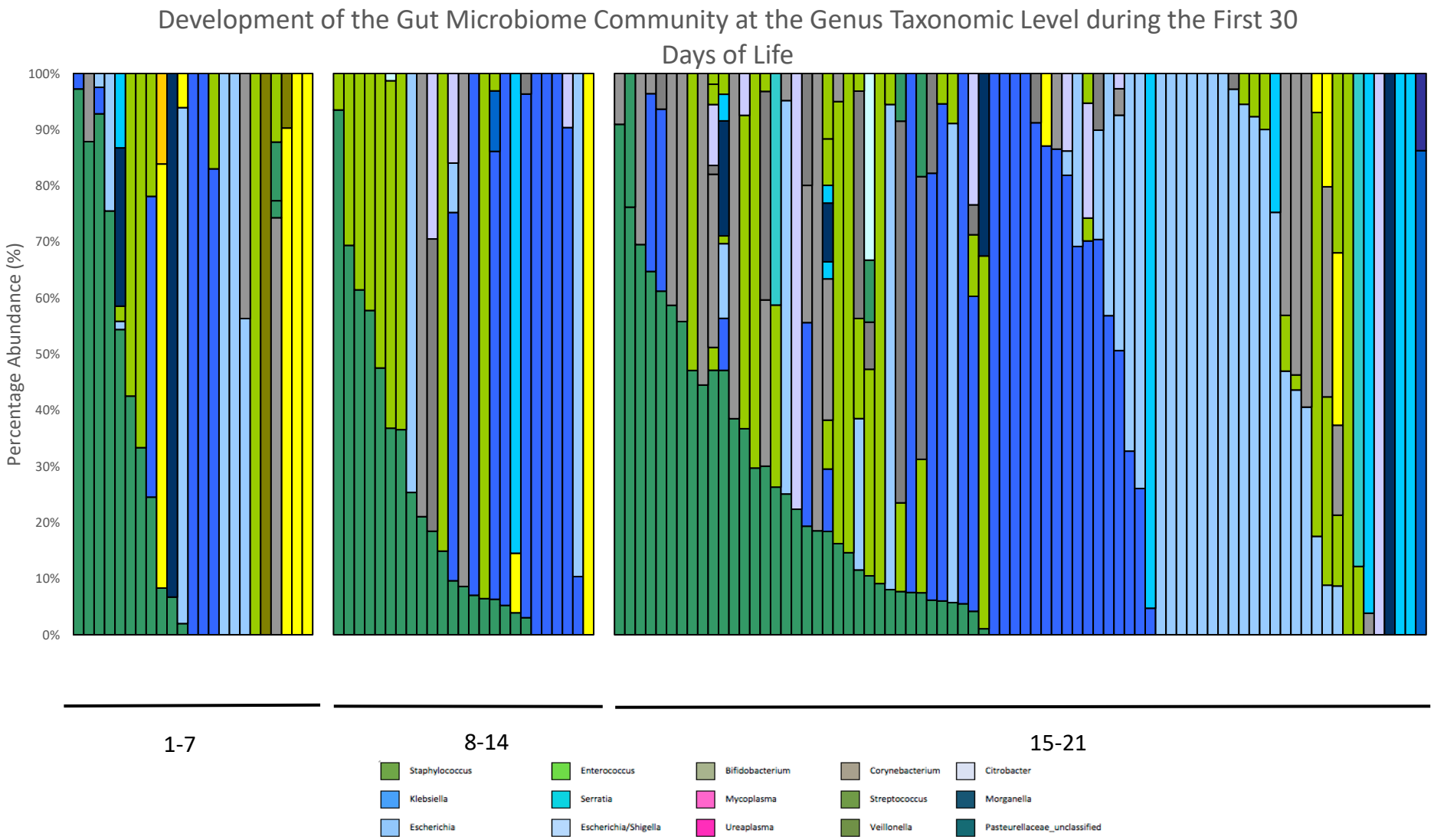

Figure 34. The Development of the Gut Microbiome Community at the Genus Taxonomic Level during the First 30 Days of Life. The y axis shows the percentage abundance of each organism from the respected phylum in that sample, where each sample is represented by a bar on the chart. The x axis has been split to show the samples that are included in each day of life category. The samples have been ordered by both day of life and the *Klebsiella* genus. This enabled for the visualisation of the trend towards a *Klebsiella* and *Escherichia* dominance over time.

3.4.7 The Effect of Development on the Diversity of the Preterm Infant Gut Microbiome.

As a result of the observations in the previous Section 3.4.6, it was hypothesised that the overall diversity of the preterm infant gut microbiome would be low. This is a result of the dominance of the microbial community by an individual organism. Taking these findings into account, the results of several alpha diversity measures can be found in Figure 35. To note alpha diversity measures the diversity of organisms within samples.

Firstly, charts A, B and C of Figure 35 shows different measures of richness within samples. Measures of richness indicate how many different organisms are observed within a sample. The three different indices used during this analysis was SOBS (Species Observed), chart A, Chao1 index, chart B, and the Ace index, chart C. All three measures of richness show that the number of organisms observed is the highest during the first week of life. It then gradually decreases until 3 weeks of life where it increases once more. Therefore, very soon after birth the gut microbiome of preterm infants contains the highest number individual organisms. Over the next 3 weeks certain organisms begin to outcompete the others, thereby reducing the number of species present. After 3 weeks of life different organisms begin to thrive, as a result increasing the richness of the microbial community in the gut.

In comparison, charts D and E of Figure 35, show the diversity of samples during the first month of life. More specifically, chart D shows the Simpson diversity index and chart E the Shannon diversity index. Measures of diversity take into account both number of species and the number of organisms within a species that are present. Therefore, both D and E of Figure 35, show that the diversity of samples remains consistent until 3 weeks of age. At 3 weeks of age it increases slightly and then falls after 25 days of life. Overall, these results show that during the first month of life there are no significant changes in the diversity of the preterm gut microbial community.

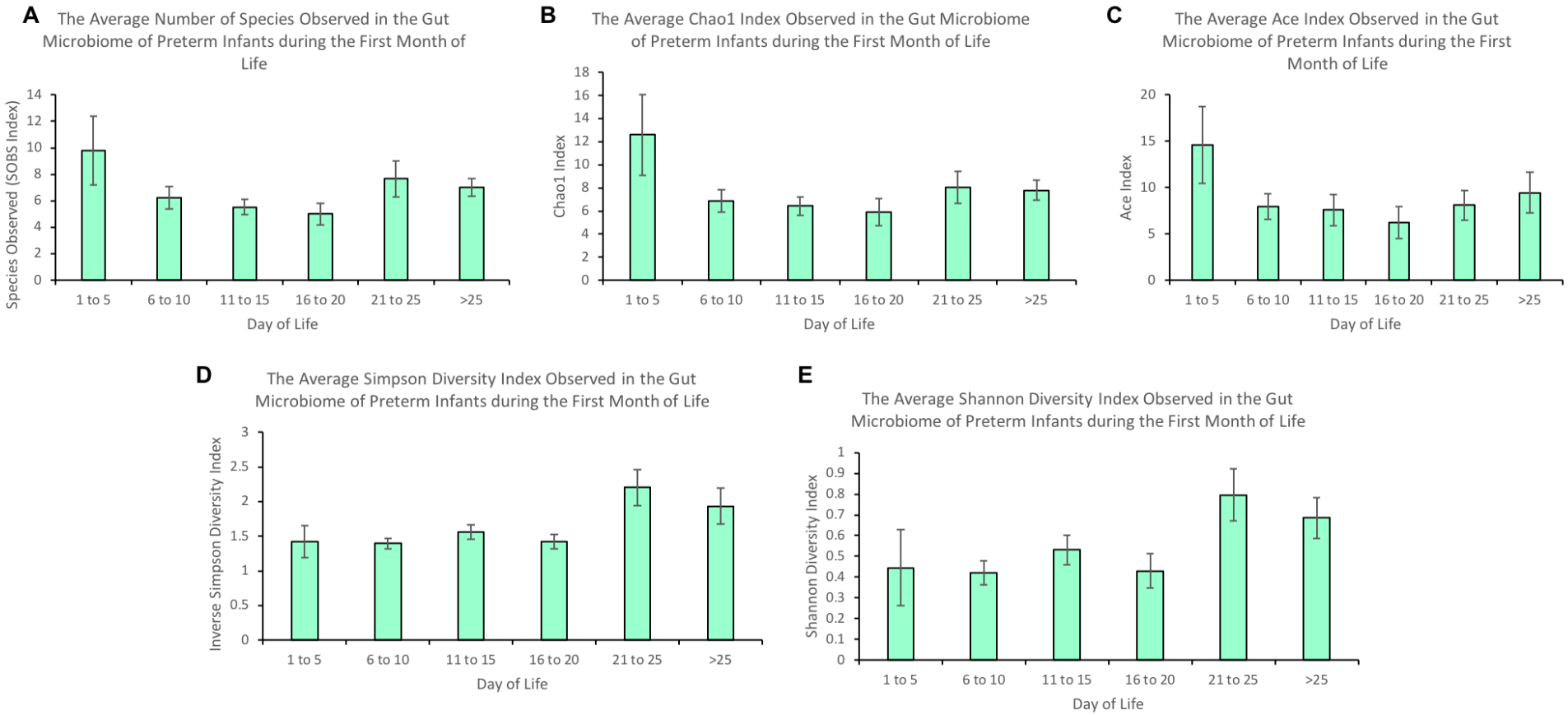

Figure 35. The Development of the Preterm Gut Microbiome Community in Terms of Alpha Diversity. For all charts in the figure, the day of life group is given on the x axis and the measure of alpha diversity on the y axis. In total 142 samples were used during this analysis, if more than one sample within the same day of life grouping from the same infant was present, the average of these samples was taken. All error bars in the figure represent the standard error of the mean.

In summary, the alpha diversity of the gut microbiome of preterm infants remains relatively stable throughout the first 30 days of life. There are no dramatic shifts as a result of time. The effects of clinical factors such as antibiotics, feeding etc. will be discussed in later section in this chapter, Section 3.4.8.

The second measure of diversity for analysis into the microbial community of the gut microbiome is beta diversity. Beta diversity measures the diversity between samples. The aim of this analysis was to determine if development had an impact upon the diversity between samples from early in life to those taken during the fourth week of life. Upon testing the effect of development and numerous other factors, such as gender or feeding, the beta diversity remained unaffected by any of these factors. Instead it was discovered that samples clustered only according to the organism that was dominant in that sample, see Figure 36. From this analysis it was determined that the samples clustered into 5 distinct groups. The *Escherichia* and *Klebsiella* groups, ellipses A and B, cluster separately and away from the main group of samples. However, even though there is a clear progression towards a *Proteobacteria* dominance over time, these results were not reflected in the beta diversity analysis. In the main group of samples, there are 3 clusters, *Staphylococcus*, *Enterococcus* and a mixed group. The organisms of *Firmicutes* origin cluster away from the *Proteobacteria* is clear, and provides some explanation as to the pattern observed, however a clear explanation remains to be determined.

In summary, when these results are compared to those in previous sections, it clearly demonstrates that the dominance of certain organisms within the preterm gut are unaffected by confounding factors such as age. Therefore, the dominance of organisms within a sample affect the community structure and subsequent beta diversity analysis, to such an extent that clinical factors do not.

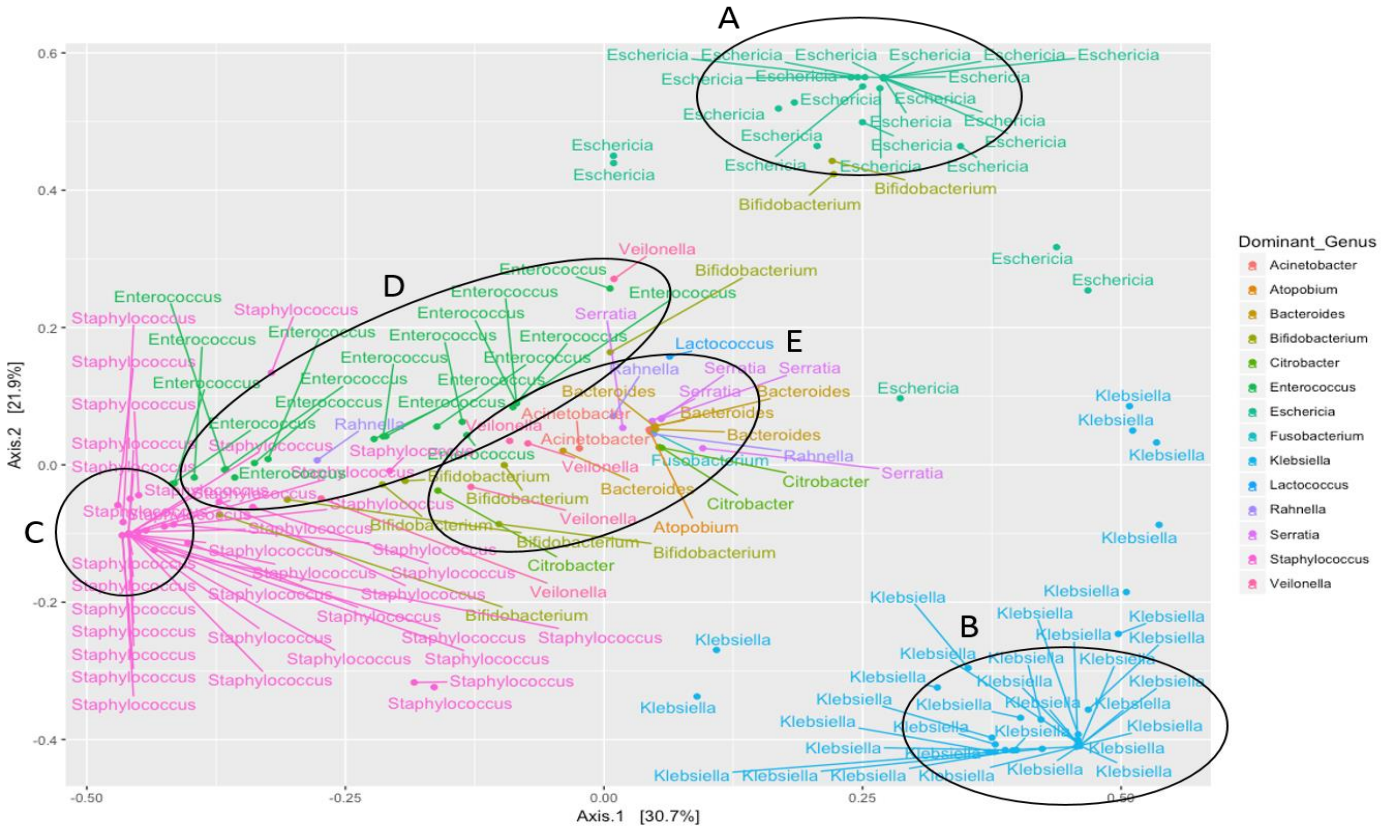

Figure 36. The Effect of Dominant Organisms on the Clustering of Samples using the Bray-Curtis Dissimilarity Index. Each sample is represented on the graph by a coloured dot. The colour and text refer to the dominant genera within that sample. The ellipses are drawn to clarify the individual clusters. This is a Principal Co-Ordinate Analysis (PCoA) using the Bray-Curtis dissimilarity index. This method calculates the similarity between samples and then organises them in a 3D space. The plot then allows for this matrix to be seen in 2 dimensions; this is why there is a lot of overlap between samples. Cluster A is *Escherichia* dominated, cluster B is *Klebsiella* dominated, cluster C is *Staphylococcus* dominated, cluster D is *Enterococcus* dominated and cluster E contains a mix of dominant organisms. On each axis there is a percentage, 30.7% and 21.9%, this refers to the amount of variability between the samples that is explained by this plot.

3.4.8 The Effect of Clinical Factors on the Development of the Preterm Gut Microbiome.

Previously it has been determined that the dominance of certain organisms remains unaffected by clinical factors when analysed using beta diversity techniques. However, when looking at the effects of clinical factors on the whole cohort, significant differences were determined. One of the most influential factors was the administration of antibiotics. As a result, the effect of antibiotics on the developing gut microbiome is detailed in Chapter 5.

Gender is determined at the moment of conception and the results of this study show that it can have a dramatic impact on the development of the preterm gut microbiome. The results in Figure 37, illustrate the effect of gender of the gut microbial community. Firstly, panel A of Figure 37, shows that the abundance of *Proteobacteria* is increased in male infants, whereas the number of *Firmicutes* is increased in female infants. At the genus level, panel B of Figure 37, the abundance of *Proteobacteria* translates into increases of *Klebsiella* and *Escherichia* in the male infants. On the other hand, the increase in *Firmicutes* in the female infants translates into an increase in *Staphylococcus* at the genus level, panel B of Figure 37. Finally, when analysed using the software package STAMP the increase in *Proteobacteria* observed in male infants was found to be significant, $p = 0.029$, panel C of Figure 37. Therefore, confirming the effect of gender on the development of the preterm gut microbiome.

The mode of delivery can sometimes be a choice of preference by the mother, however in preterm delivery it is often a decision made by the attending clinicians in order to give mother and baby the best outcome. Due to the differences in the skin and vaginal microbiome, mode of delivery can have a dramatic impact on the developing gut microbiome. The same was seen in this study, as shown in Figure 38.

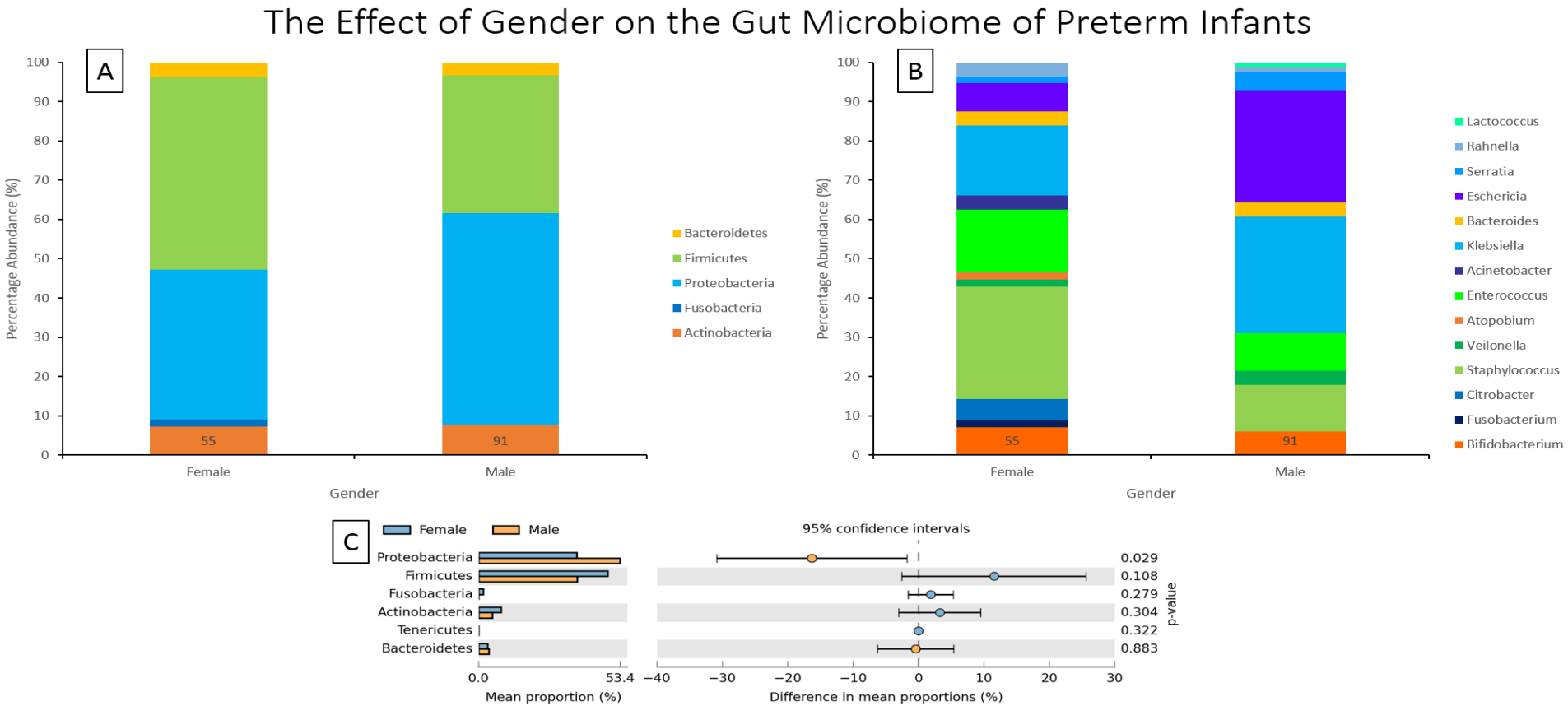

Figure 37. The Effect of Gender on the Gut Microbiome Community of Preterm Infants. Panel A shows the effect of gender on the gut microbiome development at the phylum taxonomic level. Panel B shows the effect of gender on the gut microbiome at the genus taxonomic level. The y axis in panels A and B shows the percentage abundance of each organism, and the x axis shows the gender. Furthermore, the numbers at the base of each bar on panels A and B, indicate the number of samples included within that group. Panel C is the output of an extended error bar using Welch's t test, a significance threshold of 0.05 was used during this analysis.

At the phylum taxonomic level, as shown in panel A of Figure 38. The *Bacteroidetes* phylum is only present in samples taken from infants who were delivered vaginally. Furthermore, the numbers of *Proteobacteria* are increased in samples taken from infants who were delivered by caesarean section. These differences are also demonstrated at the genus level in panel B of Figure 38. There is a greater abundance of *Klebsiella* in the samples taken from infants delivered by caesarean section. The *Bacteroides* genus is only present in samples taken from vaginally delivered infants. Moreover, these differences were shown to be statistically significant, as shown in panels C and D of Figure 38. In summary, mode of delivery has a significant effect on the development of the gut microbiome in preterm infants.

Often unreported is the effect of sampling site if multiple centres are used during a study of this kind. Therefore, the aim of this analysis was to determine if samples taken from infants residing on geographically different NICUs were significantly different. Ultimately, this analysis was a necessary prerequisite in order to determine if these infants could be treated as a collective group.

The effect of sampling site on the preterm gut microbiome can be seen in Figure 39. Firstly, as shown in panel A of Figure 39, there is a substantial increase in the abundance of *Actinobacteria* in the samples taken from infants residing on the NICU at UHW. In contrast, the infants at NBT have a higher level of both *Firmicutes* and *Proteobacteria*. Secondly, in panel B of Figure 39, the increase of *Actinobacteria* in the UHW infants translates into an increase of *Bifidobacterium*. On the other hand, the infants from NBT have an increase in *Escherichia* and *Enterococcus*. Upon statistical analysis, it was found that the increase in *Actinobacteria* in the gut microbiome of preterm infants from UHW was statistically significant, $p < 0.01$. Moreover, the abundance of *Bifidobacterium* in these infants was also discovered to be statistically significant, $p < 0.01$. Whereas at NBT, the samples taken from infants residing on the NICU at NBT were found to harbour statistically significant levels of *Escherichia* and *Serratia*, $p < 0.01$ and $p = 0.02$ respectively.

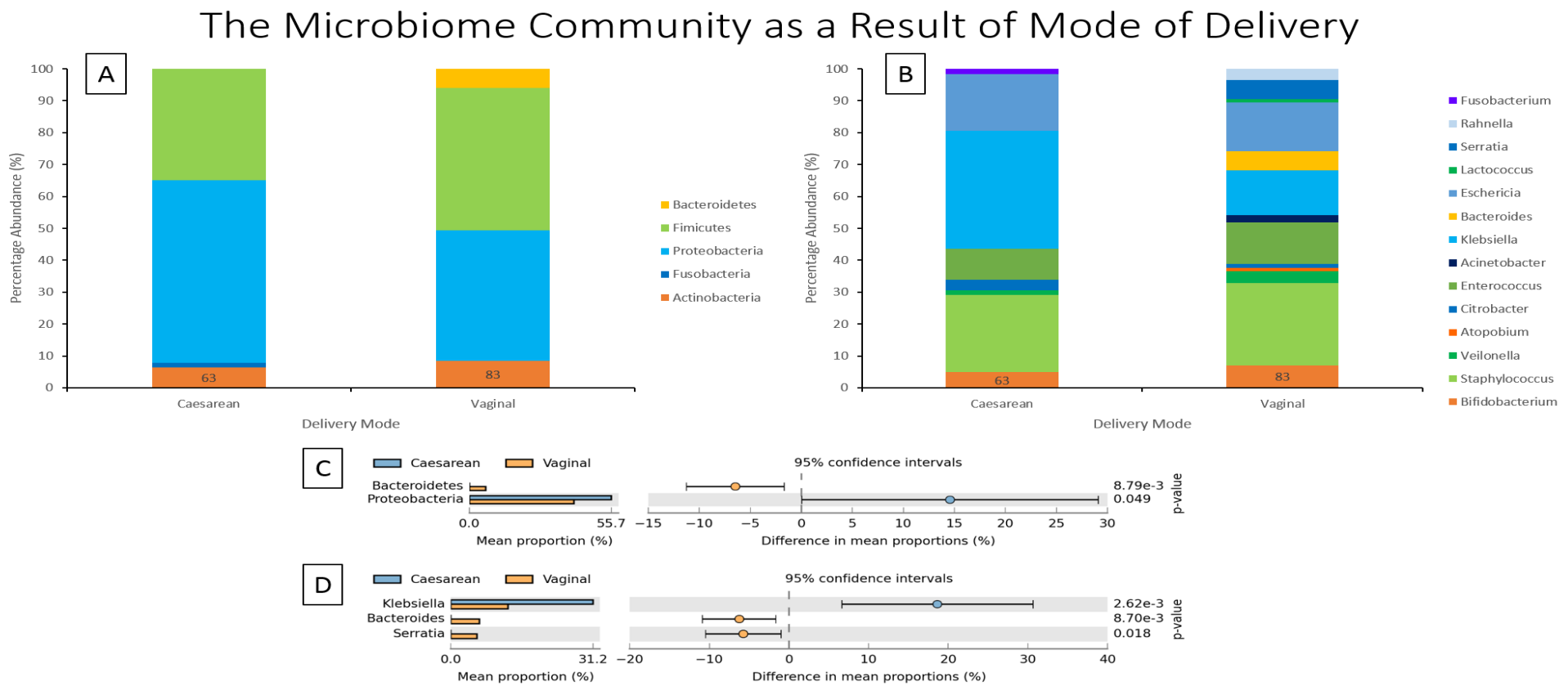

Figure 38. The Effect of Mode of Delivery on the Gut Microbiome of Preterm Infants. Panel A shows the effect of delivery mode on the gut microbiome community at the phylum taxonomic level. Panel B shows the effect of mode of delivery of the gut microbiome community at the genus taxonomic level. The y axis in panels A and B shows the percentage abundance of each organism, and x axis the delivery mode. Furthermore, the numbers at the base of each bar on panels A and B, shows the number of samples within that group. Panel C is the output of an extended error bar using Welch's t test, a significance threshold of 0.05 was used during this analysis. This was done at the phylum taxonomic level. Panel D is the output of an extended error bar using Welch's t test, a significance threshold of 0.05 was used during this analysis. This was done at the genus taxonomic level.

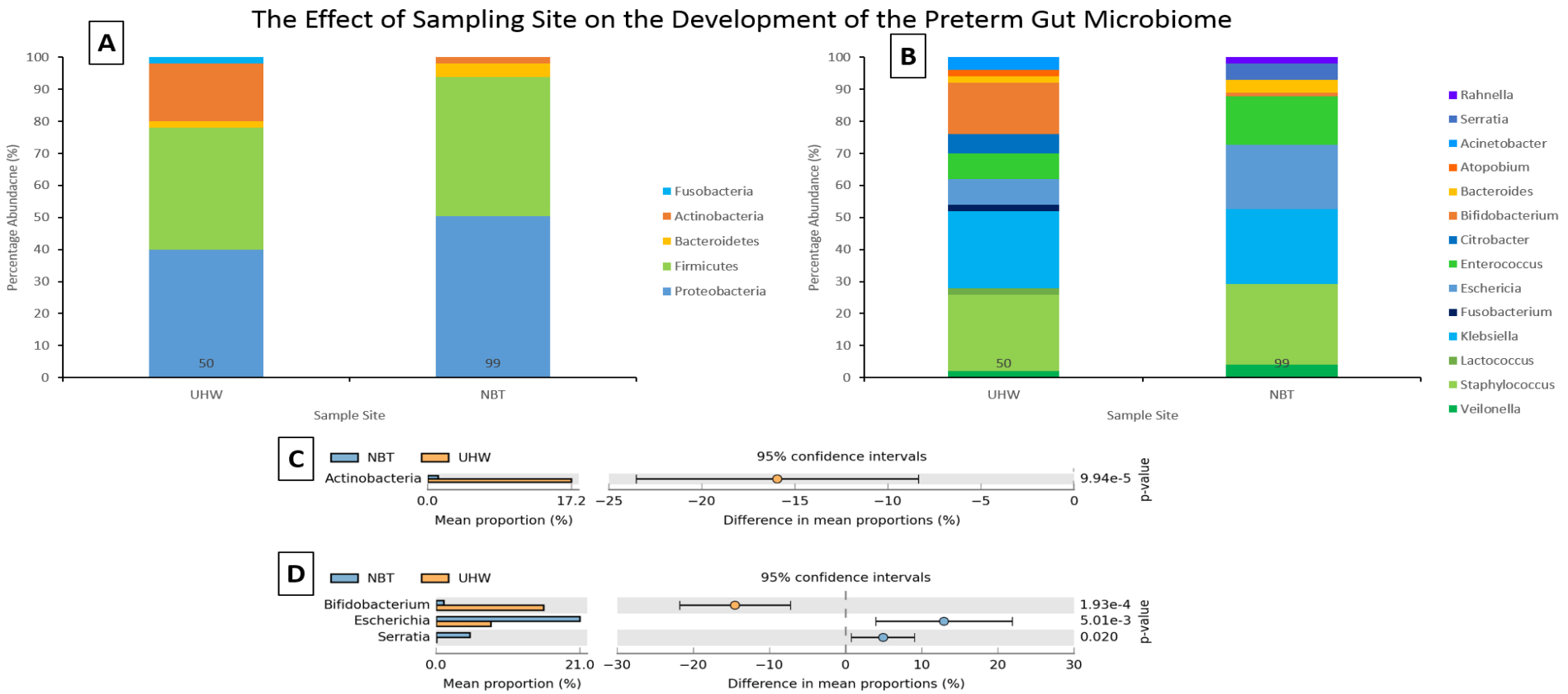

Figure 39. The Effect of Sampling Site on the Development of the Preterm Gut Microbiome. Panel A shows the effect of sampling site on the gut microbiome community at the phylum taxonomic level. Panel B shows the effect of sampling site on the gut microbiome community at the genus taxonomic level. The y axis in panels A and B shows the percentage abundance of each organism, and x axis the sampling site. Furthermore, the numbers at the base of each bar on panels A and B, shows the number of samples within that group. Panel C is the output of an extended error bar using Welch's t test, a significance threshold of 0.05 was used during this analysis. This was done at the phylum taxonomic level. Panel D is the output of an extended error bar using Welch's t test, a significance threshold of 0.05 was used during this analysis. This was done at the genus taxonomic level.

In order to determine if the two sample sites could be combined into a single dataset, the origin of the significantly increased levels of *Bifidobacterium* in the samples taken from UHW infants, were investigated. The result of this investigation can be found in Figure 39. Firstly, in panel A of Figure 39, it can be seen that the use of probiotics results in a significant increase in the abundance of *Actinobacteria*. Secondly, in panel B of Figure 39, there is a significant increase in the number of *Bifidobacterium* as a result of probiotic administration. Furthermore, there is a significant increase in the abundance of *Escherichia* and *Serratia* in the samples taken from infants who did not receive probiotics. To clarify, the probiotic, Infloran, given to infants contained only 2 organisms a *Bifidobacterium* and a *Lactobacillus* species. Moreover, this probiotic was only administered to infants at UHW during the sampling period.

When comparing the results of Figure 38 and Figure 39, they are strikingly similar, especially in regards to the results of the STAMP software. Therefore, the differences observed in samples taken from infants at NBT and UHW is a result of probiotic administration, and not as a result of any extraneous environmental variable. As a result, it was possible for the UHW and NBT datasets to be combined during further analysis, to determine the development of the preterm gut microbiome during the first month of life.

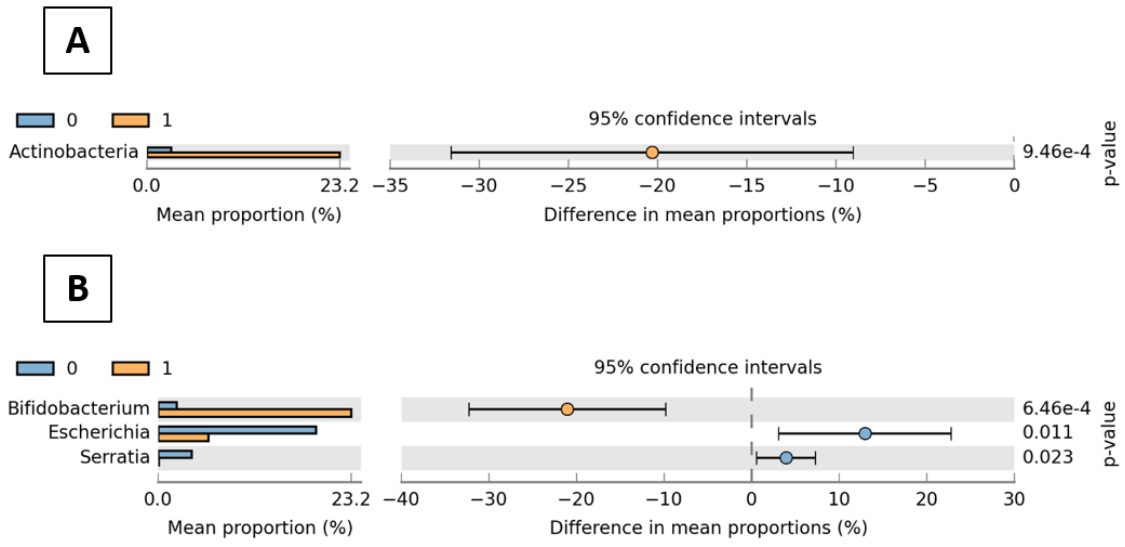

Figure 40. The Effect of Probiotics on the Microbial Content of the Preterm Infant Gut Microbiome. Panel A of this figure is an extended error bar plot using Welch's t test for significance, p = 0.05 threshold. This was the data analysed at the phylum taxonomic level. Panel B of this figure is another extended error bar plot using Welch's t test for significance, p = 0.05. This data was analysed at the genus taxonomic level. The 0 or blue colour are the samples taken in the absence of probiotics, and the 1 or orange colour are the samples taken in the presence of probiotics. There were 124 samples taken in the absence of probiotics and 27 in the presence of probiotics, therefore a total of 151 samples were used during this analysis.

A key factor affecting the gut microbiome is mode of feeding. As part of this study, the effect of feeding on the development of the preterm gut microbiome was investigated. The results of this investigation can be found in Figure 40. Firstly, panel A of Figure 40, shows that the *Bacteroidetes* phylum is only present in infants who are exclusively breast fed. The greatest abundance of *Actinobacteria* can be found in the infants who received exclusively formula feeds. Finally, the infants who received exclusively donor breast milk had the greatest abundance of *Proteobacteria*. At the genus taxonomic level, panel B of Figure 40, there are no *Klebsiella* present in the infants receiving no enteral feeds at the time of sampling. The *Enterococcus* genus is only present in the infants who received exclusively mother's own breast milk or no enteral feeds. Furthermore, *Lactococcus* organisms are only found in the infants who received exclusively formula feeds during the sampling period. The greatest abundance of *Staphylococcus* was found in samples taken from infants receiving donor breast milk. Finally, levels of *Escherichia* remain relatively consistent across all feeding types.

In summary, the gut microbiome of preterm infants is significantly affected by several factors before, during and after birth. Therefore, this indicates that the development of the preterm gut microbiome to be a multifactorial process, with numerous of influencing variables.

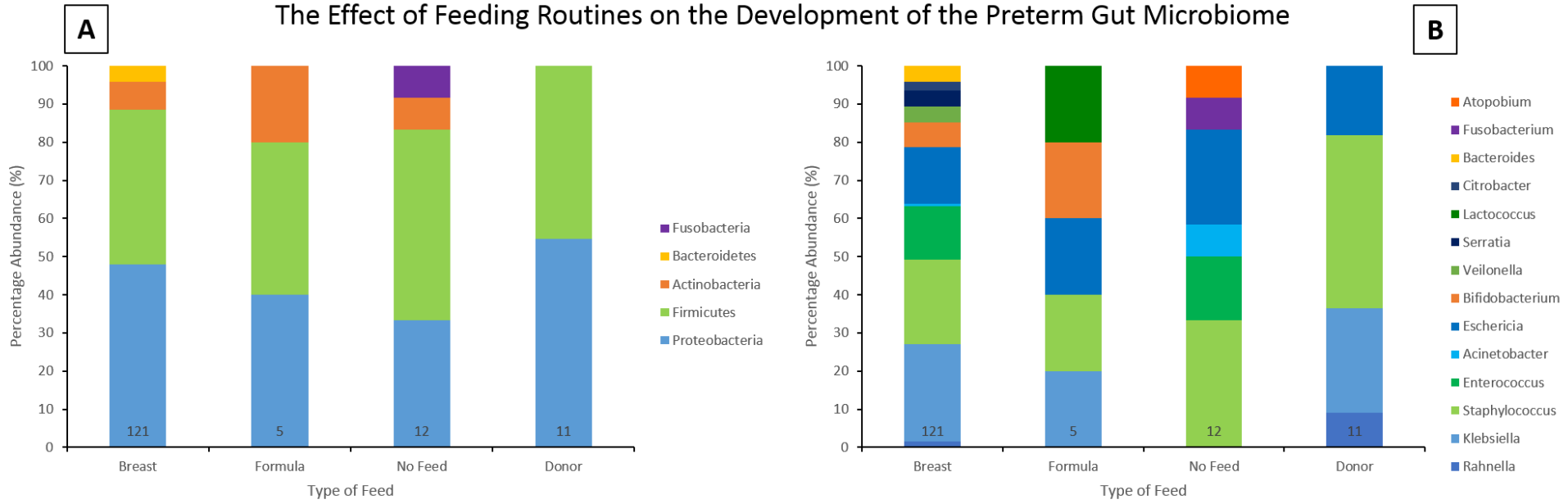

Figure 41. The Effect of Feeding Routines on the Development of the Preterm Gut Microbiome. Panel A shows the percentage abundance, y axis, of organisms at the phylum taxonomic level as a result of different feeding routines received on the NICU, x axis. Panel B shows the percentage abundance, y axis, of organisms at the genus taxonomic level as a result of different feeding routines received on the NICU, x axis. The numbers at the bottom of all the bars are the number of samples included in each group. The total number of samples used during this analysis was 149.

3.4.9 . The Gut Microbiome of Preterm Infants who Develop NEC Significantly Differs from those Who do not.

NEC is a neonatal disease with high incidences of mortality and morbidity. Currently the cause remains unknown, therefore it is of great interest to the scientific community to determine the risk factors for the development and progression of this disease. The gut microbiome has been implicated in the pathology of this disease, but a specific organism has yet to be identified. Therefore, the aim of this investigation was to further investigate the potential links between the gut microbiome and the development of NEC.

Table 16 summarizes the clinical data of the groups of infants who did and did not develop NEC. In total there were 11 infants who provided 29 samples recruited to this study. The average birthweight and gestation of the infants who developed NEC was lower than those who did not. Interestingly, more male infants developed NEC, however there were more male infants recruited to the study as a whole. Furthermore, all infants who received donor breast milk did not develop NEC. Finally, the majority of infants who developed NEC were delivered vaginally, however this could also be a spurious result of there being more vaginally delivered infants in the study as a whole.

Table 16. A Demographics Table Focusing on the Development of NEC. The percentages in brackets are comparisons within groups. The NAs are non-applicable to that grouping. For an infant to be grouped into the 'yes' column they developed Grade 1 to 3 NEC, according Bell's criteria.

Development of NEC	Yes	No	Total
N			
Number of Infants	11	36	47
Number of Samples	29	117	146
Birthweight (mean ± standard deviation) (grams)	759 ± 154	989 ± 457	NA
Gestation (mean ± standard deviation) (complete weeks)	25 ± 1	27 ± 3	NA
Gender			
Male	19 (66%)	73 (62%)	92
Female	10 (34%)	44 (38%)	54
Feeding Routine			
Mother's Own Milk	22 (76%)	97 (83%)	119
Formula	2 (7%)	3 (3%)	5
Parenteral	5 (17%)	7 (6%)	12
Donor Breast Milk	0 (0%)	10 (9%)	10
Delivery Mode			
Caesarean	12 (41%)	49 (42%)	61
Vaginal	17 (59%)	68 (58%)	85

In order to determine the influence of the gut microbiome on the development of NEC, the infant and the samples they gave were grouped according to the development of NEC status. All infants who developed NEC of Grade 1 or higher, according to Bell's criteria, were classed as having developed NEC and were included in the 'yes' category. The results of this analysis can be found in Figure 42. Firstly, panel A of Figure 42, shows that the *Bacteroidetes* and *Actinobacteria* phyla were only present in samples taken from infants who did not develop NEC. However, these differences were found to be not significant. Secondly, panel B of Figure 42, shows that the *Enterococcus* genus is enriched in infants who do develop NEC. This is an interesting result as the *Enterococcus* genus is known to produce proteases and has been linked to sepsis. The abundance of *Klebsiella*, *Escherichia*, and *Staphylococcus* are increased in infants who do not develop NEC. Finally, panel C of Figure 42, shows the significant differences between the 2 groups of infants, with *Citrobacter* and *Veilonella* significantly increased in infants who do not develop NEC. Furthermore, the increase in *Enterococcus* in infants who do develop NEC is significant, $p = 0.021$.

3.5 Summary

The results of this Chapter have shown that the preterm gut microbiome progresses from a *Firmicute* dominated environment to a *Proteobacteria* one, during the first 30 days of life. Differences as a result of several clinical factors have been found. Moreover, the total bacterial load of infants as a result of different conditions have also been shown. However, the diversity of the preterm gut microbiome did not change over time or as a result of clinical factors. Finally, significant differences between the gut microbial communities in infants who do and do not develop NEC were found. These results will be discussed further in the following section.

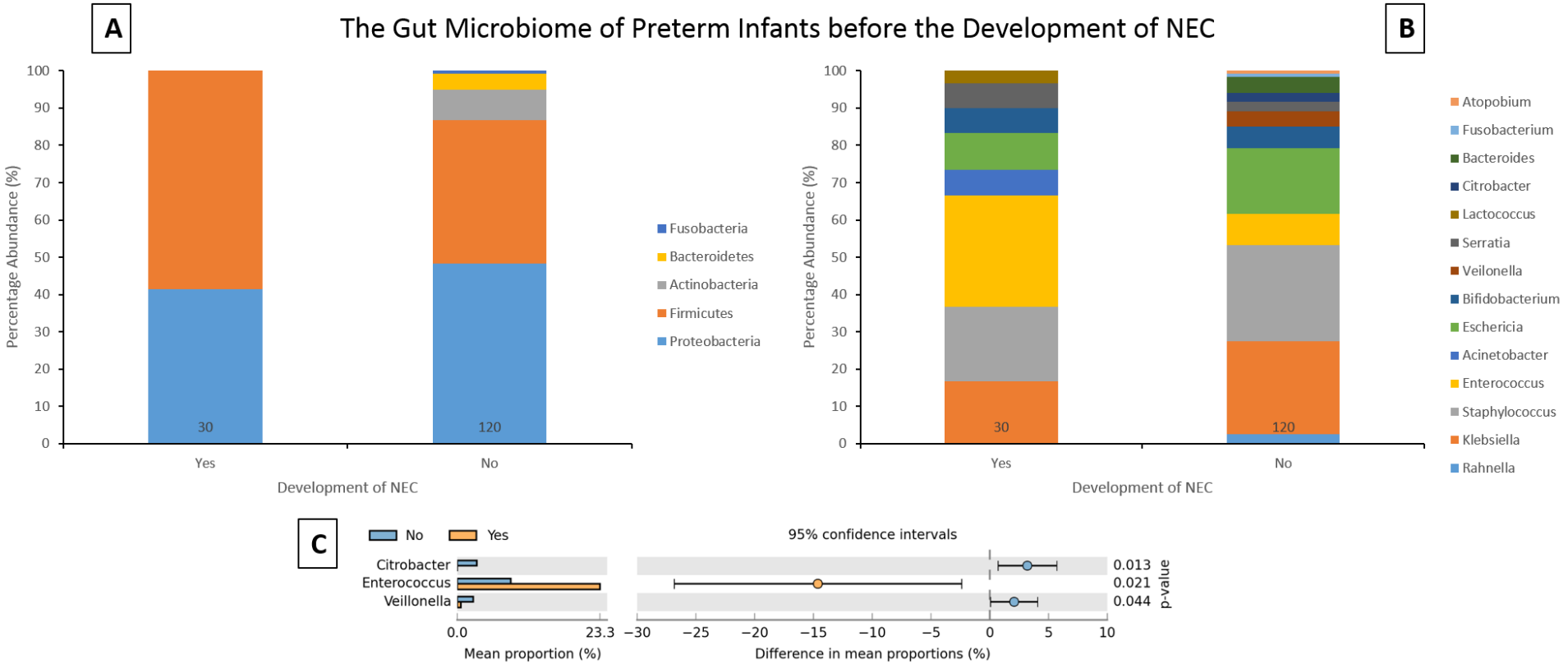

Figure 42. The Gut Microbiome of Preterm Infants before the Development of NEC. Panel A shows the percentage abundance, y axis, of organisms at the phylum taxonomic level. The samples for this analysis were grouped according to the development of NEC in the infant the samples were given by, x axis. Panel B shows the percentage abundance, y axis, of organisms at the genus taxonomic level. The samples for this analysis were grouped according to the development of NEC in the infant the samples were given by, x axis. The numbers at the bottom of each bar in panels A and B, are the number of samples included in the group. These sample number apply to the groupings in panel C. In total 150 samples were used during this analysis.

3.6 Discussion

3.6.1 The Optimum Stool Weight and DNA Extraction Method for Studies into the Preterm Gut Microbiome.

The aim of this investigation was to confirm the optimum weight of stool necessary to produce adequate 16S rRNA gene. In addition, multiple DNA extraction kits with established protocols were being used by colleagues. There have been several studies investigating the best kits and methods to extract 16S rRNA gene from faecal samples. However only one study has focused specifically on preterm stool samples. Therefore, it was necessary to determine which one would be optimal during this study.

Research into the efficacy of different DNA extraction methods has shown that kits such as the MoBio PowerSoil, Qiagen QiAmp, and MPBio are highly successful in yielding 16S rRNA gene DNA (Holland *et al.* 2000; McOrist *et al.* 2002; Li *et al.* 2003; Nechvatal *et al.* 2008; Mirsepasi *et al.* 2014; Wesolowska-Andersen *et al.* 2014). Furthermore, these kits are often less time consuming than manual protocols such as phenol-chloroform extraction. Further research has investigated the role of mechanical extraction methods such as bead beating. These studies have shown that in order to gain an accurate reflection of a microbiome community, bead beating is a necessary step (de Boer *et al.* 2010; Salonen *et al.* 2010; Yuan *et al.* 2012; Wesolowska-Andersen *et al.* 2014). This is to ensure complete breakdown of the thick peptidoglycan cell wall in gram positive organisms. Therefore, the two methods tested during this study tested including a bead beating step, before the sample was taken forward to the kit protocol proper.

A key issue of research using preterm infants as the recruiting population, is the limited sample size that can be collected. Therefore, it was important for this study to ensure that any samples given below the manufacturer recommended volume would be viable. As shown by the data in Section 3.4.1., the volume of starting material did not impact the yield of 16S rRNA gene. As a result, all samples were extracted regardless of weight, however samples more than 200 mg were weighed and only 200 mg used.

These results were compared to previous research into the optimisation of DNA extraction from preterm stool samples by Alcon-Giner *et al* in 2017. Their research

concluded that the DNA extraction kit QIAmp Stool Mini by Qiagen yielded satisfactory 16S rRNA gene for sequencing. Secondly, they found that a prolonged bead beating step was necessary for optimal bacterial DNA extraction. Furthermore, they commented that this was especially important for infants receiving probiotics, of which a proportion of the study cohort did (Alcon-Giner *et al.* 2017).

In summary, this investigation and previously published research strongly established the DNA extraction method chosen for use during this study. Furthermore, all the necessary steps and experiments were conducted in order to confidently choose the optimal methods. As a result, the data discussed in this thesis accurately reflects the bacterial community of the preterm gut microbiome.

3.6.2 Illumina MiSeq 16S rRNA Gene Sequencing for Mapping the Preterm Gut Microbiome

Often overlooked but extremely informative is the number of DNA samples sent for sequencing, the number of samples returned, and the number of reads they produced during sequencing. In total, 151 (72%) of samples were in included in the Microbiome Cohort, this was just below three quarters and provided a clear and detailed analysis of the colonisation of the preterm infant gut colonisation. Furthermore, the higher percentage demonstrates the efficacy of the DNA extraction, sequencing and analysis techniques used.

Lest not forget, there were a number of samples that were not included in the Microbiome Cohort, this was a result of both low reads numbers but also a number of samples did not amplify sufficiently during Section 2.9. There are numerous possible explanations for this. For instance, infant 19, was administered antibiotics throughout the entire sampling period, thereby reducing bacterial abundance. To confirm this, the total bacterial load of each sample was determined, Section 2.17. Infant 19 had a mean bacterial load of 3342 copies/mg this was substantially lower than the mean for the Study Cohort. Moreover, this infant was exposed to three different antibiotics at each time of sampling. As research has shown (Dethlefsen and Relman 2011), this could cause a reduction in bacterial richness, therefore increasing the likelihood of a negative sequencing result. Furthermore, meropenem was used throughout the sampling period in this infant. Meropenem is a broad spectrum β-lactam antibiotic used to treat serious

bacterial infections (Kollef 2008). Therefore, the use of the broad-spectrum antibiotic, which is not often used in this cohort, may explain the limited bacterial numbers. As discussed in Chapter 5, these results indicated at a tangible link between antibiotics, bacterial load and sequencing results. Furthermore, infant 19 provided 2 stool samples of only 10 mg, therefore the lack of starting material could be another possible factor in the low bacterial DNA and subsequent sequencing results.

To summarize, there is no clear explanation for the lack of conclusive results for this infant, there are two possible explanations for this. Firstly, a failure of the Illumina MiSeq process may be the reason, however for it to affect all sample from only one infant is highly unlikely. Therefore, this indicates a contribution from each of the reasons discussed. However, this was not the only individual or sample to produce unusable results and upon further examination of the 59 samples removed, there were several reasons to explain these results. The following cause have been mentioned but are expanded and added to here.

Firstly, the Illumina MiSeq reaction may not have executed correctly and led to these samples being removed, because there was bacterial DNA present, albeit in low numbers. The reaction disruption could be due to a number of reasons for example, a lack of correct primer adhesion would lead to little or no amplification. Secondly, the mean weight of the stool collected for the missing samples was 628 mg, compared to 640 mg for the Microbiome Cohort. This shows that the weight of the stool is a contributing factor but not a definitive cause. Thirdly, as mentioned previously the effect of antibiotics on the positivity of sequencing results can be seen in the entire group of samples that were removed. A chi-squared test between antibiotic use and sequencing results gave a p value < 0.01. This clearly shows a significant effect of antibiotics on the production of positive sequencing results. Another contributing factor could be early day of life, as shown in Table 1. These samples are the most likely to contain the lowest levels of bacteria and it reflected in the 32% sequencing results.

There are some factors worth mentioning that do not affect the likelihood of a positive sequencing result. For instance, both the Study and Microbiome Cohorts have the same mean gestational age, therefore the degree of bacterial colonisation is not dependent upon gestation. Also, the rates of feeds given to these two groups is not different, demonstrating that if an enteral feed is given, this makes no impact on the positivity of a

sequencing result. The effect of feeds and the microbiome will be explored in later sections of this thesis.

The final reasons for decreasing positivity are as follows. As previously mentioned, the number of individuals not included in the Microbiome Cohort were infants that only produced a single sample for this study. This would be an indication that the infant is struggling to pass stool and who may be clinically unwell. As a result, the samples they were able to give were low in bacterial richness as a result of illness, or the illness caused a decrease in bacterial load. The cause and effect cannot be established. Lastly, results not shown, there is a slight decrease in the number of samples taken during the administration of probiotics. This could lead to a reduction in the organisms present in the gut of the infants who were not included in the Microbiome Cohort.

In summary, a substantial number of the samples taken during this study were of sufficient quality in order to be used for analysis to determine the microbiome development in preterm infants. This is a positive result as often these samples are difficult to obtain, process and analyse due to the prematurity of these infants. These individuals have high and changeable care needs, making them a difficult section of the population to study. This is further demonstrated by the numerous factors contributing to the exclusion of so samples from the Microbiome Cohort.

3.6.3 DNA Extraction Controls and Sub Sampling

Negative controls are a necessary part of every experiment in order to control for any extraneous variables. The aim of the controls included in this investigation were to detect any possible contaminants in the DNA extraction kits used, and contaminants included as a result of extraction methodology. A control was included for each DNA extraction kit used, totalling 5. Recent research has shown that kit contamination is ubiquitous among the most widely kits available (Salter *et al.* 2014). Therefore, it was not an unexpected result for the controls included this study to produce a positive sequencing result.

The results detailed in Section 3.4.3 shows that the organisms detected were common to the organisms of interest within the samples. There are recommended methodologies used during microbiome analysis to control for the organisms detected in negative

controls (Salter *et al.* 2014). For instance, removing the organisms detected in the controls from the samples of interest. However, due to the concordance of organisms detected in the controls and samples of interest this was not appropriate. Therefore, the number of reads produced by the control samples was analysed. It was determined the average number of reads produced by these samples was significantly lower than that produced by samples of interest. As a result, it was concluded that even though the same organisms were detected, they were present in sufficiently low numbers to be negligible. Ultimately, the controls were considered negative. Therefore, the results were considered an accurate reflection of the preterm gut microbiome composition.

Another key step in producing a final dataset during microbiome analysis is the decision to subsample. The process of subsampling takes a random number reads per sample to produce a smaller, but accurate, representation of the organisms in the larger dataset. The aim of subsampling is to produce a dataset that is representative of the original dataset, which is much smaller in order to allow for faster analysis in downstream analysis. The accurateness of this subsample is given by the coverage score, produced during data processing using the Mothur software. For this study, coverage scores of 0.99 or greater were taken as the threshold. A score of 0.99 or greater shows that a minimum of 99% of the organisms detected in the original sample are represented in the subsampled dataset. Even though this method provides accurate results it can be considered inappropriate for highly diverse samples as it can simplify the array of organisms detected, possibly resulting in the loss of relevant data (McMurdie and Holmes 2014). However, due to the minimal diversity and predominance of organisms seen in samples from this recruitment cohort, as a whole and in this study, subsampling during this study was deemed appropriate, and performed to a depth of 1000 reads.

In summary, the use of subsampling as part of the methodology employed during this study was appropriate and would be considered favourable for future research in this area. The only caveat is that the recruitment group would remain the preterm and full-term infant population in order to reduce data negligence, as the gut microbiome becomes more densely populated and diverse with age. Furthermore, DNA extraction controls are a necessity in microbiome studies with potentially low biomass samples, such as the preterm gut microbiome.

3.6.4 Patient Recruitment and Sample Demographics

One of the aims in collating demographic data for a study is to be able to identify any bias that may have occurred during the recruitment. From the data shown in Table 2, the average gestation at birth of the preterm infants was 26 weeks. Therefore, all the participants of this study were classified as extremely preterm. This was a result of the recruitment criteria for the study, such that infants were only recruited if given ventilation during the first 24 hours of life. It is known that respiratory distress syndrome (RDS) is directly related to prematurity (Usher *et al.* 1971). Therefore, if the criteria recruitment necessitated the need to ventilation 24 hours after birth, due to RDS, then this would explain the extremely low average gestational age in this cohort. However, the aim of the study was to determine the development of both the lung and gut microbiome in ventilated preterm infants. As a result, it was expected that the average gestation would be very low, as ventilation was a key variable during this study. Therefore, the study by design was biased towards extremely preterm infants.

3.6.5 qPCR Results for Total Bacterial Load in the Stool of Preterm Infants.

The aim of this experiment was to determine the total bacterial load of preterm infant stool. These results were then compared to community structure and clinical factors to provide a more in-depth analysis of the preterm infant gut microbiome. Total bacterial load in preterm infant stool has been assessed previously and found to increase with gestational age (Korpela *et al.* 2018). However, links with clinical factors such as feeding and disease remain ambiguous (Abdulkadir *et al.* 2016b).

An interesting finding of this experiment was the possible use of a copy number threshold for the inclusion of a sample in high through put sequencing. This has the potential to reduce time and monetary costs associated with future microbiome studies. The results of this experiment suggest that samples containing less than two million copies of the 16S rRNA gene should not be considered for high throughput sequencing studies. However, with the economic cost of high throughput sequencing reducing, this may not be a worthwhile prerequisite (van Dijk *et al.* 2014). Taking this into account,

for studies with limited financial resources or time, this threshold could be useful tool in future research.

A significant finding of this investigation was that bacterial load in preterm infant stool peaks at 6 to 10 days of life. This change in bacterial load as a result of day of life has not been reported before, therefore it is currently impossible to determine if this is a normal occurrence in preterm infants. Moreover, it makes the cause of this decrease difficult to determine, however possible suggestions can be made. The most likely explanation is that upon admission to the neonatal ward antibiotics are administered as part of routine care. Subsequently, if the infant is doing well and there are no clinical indications of infection, such as high CRP or white blood cell count, the antibiotics are stopped. Primarily this occurs at approx. one week of life. Therefore, the increase in bacterial load detected at days 6 to 10 of life in preterm infant stool could be a result of antibiotic withdrawal, resulting in a bacterial bloom. Secondly, at days 6 to 10 of life the infant will more than likely be in a stable condition upon the NICU and clinically considered ready for enteral feeds. As a result, the influx of macro molecules into the intestinal tract provides a substantial resource for the residing bacteria. Upon the addition of excessive resources bacteria will multiply rapidly in order to utilize the resource, therefore resulting in a spike of bacterial load in the stool. Lastly, the length of time the infant has been residing on the NICU could explain this sudden rise in total stool bacterial load. However, this is the least likely explanation as there is a decrease at two weeks of life, which would not be expected if stool bacterial load was a result of the environmental bacterial load. As will be demonstrated in Chapter 4, antibiotics are the most likely cause for these changes in bacterial load.

In contrast, to the peak at 6 to 10 days of life, the total stool bacterial load decreases until the third week of life. Again, the most likely explanation for this is that after the withdrawal of antibiotics at days 6 to 10 of life, the infant is vulnerable to infection. Therefore, upon clinical signs of infection the attending clinician prescribes further antibiotics, thereby reducing bacterial load. Furthermore, if an infection is acquired it is most likely nosocomial and will be more resistant to antibiotic treatment, resulting in a, longer course of broad spectrum antibiotics causing the continual decline, seen in 3.4.5, until 3 weeks of age. Furthermore, as a result of infection the infant may become too unwell to receive enteral feeds, thereby further confounding the total stool bacterial load

decrease. In summary, there is a distinct change in total stool bacterial load in preterm infants, of which antibiotic use is the most likely explanation.

As a result of the recruitment process, infants were included in the study from two different sites, UHW and NBT. Analysis into the total bacterial load of preterm stool found a non-significant difference between the two sites, but a difference nonetheless, Figure 21. These results were repeated when comparing the bacterial load between samples taken from infants who received probiotics and those that did not, Figure 22. As indicated in Section 1.2.4.3.2, the definitive difference between the two sampling sites was the administration of probiotics. Therefore, providing a reasonable explanation for the difference in stool bacterial between infants from the different NICUs.

As discovered in a previous study the total bacterial load of preterm infant stool increases as a result of gestational age (Korpela *et al.* 2018). These results were somewhat repeated in this study with the marginal increase on bacterial load as a result of corrected gestational age, Figure 6. In addition, this study showed that full-term infants had a reduced total bacterial load in their stool. A notable finding from this study is the massive variation in total stool bacterial load in extremely preterm infants. This result has not previously been published, thereby making explanations difficult to confirm. A possible explanation is the known link between intrauterine infections and the incidence of preterm birth. This also leads into one of the major limitations with this study, the lack of maternal information. Therefore, it cannot be confirmed if the high levels of bacterial load seen in the extremely preterm infants is a result of an intrauterine infection, such as chorioamnionitis. Secondly, as mentioned previously, antibiotics are administered to infants upon admission to the NICU. The majority of infants requiring the specialist care of the NICU are those born extremely premature. Therefore, it is plausible that the low levels of stool bacterial load detected in the stool of extremely premature infants is a result of antibiotic exposure. Thirdly, the infants with the lowest corrected gestational age are the youngest, by day of life, in the cohort. Previous research has shown total bacterial load to increase by day of life. As a result the extremely preterm group are the most likely to have the lowest total bacterial load in their stools (Sharma *et al.* 2012).

A potential biomarker for detecting neonatal sepsis is the total bacterial in stool. Data shown in Figure 25, indicate that the total bacterial load in stool given by infants who

displayed clinical sign of sepsis on the day of sampling, is on average greater than those who did not. However, this difference was not significant due to the extremely high levels of variation seen in the group's samples taken from infants with clinical symptoms of sepsis. On the other hand, it indicates at a potential biomarker for the detection of sepsis. DNA extraction and qPCR for total bacterial load can be performed in 24 hours, much faster than culture-based methods, ultimately leading to a faster diagnosis and treatment. Further research is needed to determine if this could be a potential biomarker.

As mentioned previously bacterial load is likely to increase as a result of feeding, and this can be seen in Figure 8, as infants receiving exclusively parenteral nutrition have the lowest average stool total bacterial load. Interestingly, formula fed infants have the highest average stool bacterial load, shown in previous studies (O'Sullivan *et al.* 2015). The result of this increase in bacterial load has yet to be confirmed but could be a result of several factors. The most likely explanation is that the formula feeds are made in an industrial setting and then prepared by hospital staff. This process is known to introduce bacterial contamination (Anvarian *et al.* 2016). Therefore, it is this production process that is the most likely culprit for the additional bacterial load seen in the samples taken from exclusively formula fed infants. Formula feeding has been linked to a higher incidence of neonatal disease such as NEC (McGuire and Anthony 2003). However, a greater bacterial load in the stool of infants who develop NEC has not been found (Abdulkadir *et al.* 2016b). Therefore, it is difficult to associate this increase in bacterial load with, as a result of formula feeding, adverse outcomes such as NEC.

Previous research has shown gender to have a significant effect on the microbial community of the gut microbiome in both adults and preterm infants (Cong *et al.* 2016; Haro *et al.* 2016). As yet there has been no confirmation on the effect of gender on the total stool bacterial load. Therefore, the results present in Figure 26, demonstrate the dramatic effect gender has on bacterial load, albeit not significant. This study has shown that the average stool bacterial load of preterm infants is higher in males. Explanations for this gender are limited and highly speculative. The most apparent differences between the genders are hormones and genetics, and these differences have been used to explain microbiome differences (Gomez *et al.* 2015). However, due to the infancy of preterm infants, these factors are unlikely to be the main causative factor. Other

research into explanations in the gut microbiome differences between the sexes were carried out exclusively in animal models, thereby providing possible explanations but by no means revealing definitive mechanisms. For instance, studies into rodent models found that males and females responded differently to dietary components and specific microorganisms, both probiotic and pathogenic. Furthermore, the inflammatory responses, such as cytokine levels, significantly differed between the two sexes (Karunasena *et al.* 2014; Shastri *et al.* 2015). Therefore, sex differences are a key part of gut microbiome development.

Several studies have reported on the gut microbiome community differences as a result of mode of delivery (Grönlund *et al.* 1999; Dominguez-Bello *et al.* 2010; Azad *et al.* 2013). However, current research suggests that mode of delivery does not significantly impact the gut microbiome of preterm infants (Stewart *et al.* 2017). As a result, the data presented from this study provides new evidence towards a significant effect of mode of delivery on the development of the preterm gut microbiome. This study demonstrates a significant increase in total bacterial load from infants delivered by caesarean section. The reason for this difference remains to be determined, but further investigations into the bacterial community and clinical factors provides some explanation.

Investigations into the relationship between total stool bacterial load and NEC, were conducted as part of this study. It was determined that the bacterial load in stool taken from infants who developed NEC was not significantly different from this who did not. This is a key finding as it shows that community structure and the species of organisms present in the gut are the risk factors or biomarkers rather than the total bacterial load present in the gut. More simply, it is the organisms present rather than the amount of these organisms contributing to the development and progression of NEC. Therefore, more research needs to be conducted to conclusively determine if an individual, *Klebsiella*, or the combination of several organisms are a key factor in the development of NEC.

The length of stay in hospital is directly related to health, therefore length of hospital stay can act as a proxy for health when analysing clinical data. This study found an increase in bacterial load in the stool of infants who remained in hospital for more than 3 months. This indicates that a higher bacterial load during the first month of hospital stay may lead to a greater amount of time spent in hospital, therefore suggesting worse

outcomes for preterm infants. However, these results could be biased. For instance, the longer an infant remains on the NICU they will be accumulating gut microbiome inhabitants in the same way they would outside of hospital. Thereby, resulting in an increase in bacterial load. However, all the samples analysed were taken during the first month of life. As a result, this study shows that an increased bacterial load at less than 1 month of age, increases the likelihood of a hospital stay greater than 3 months. These results were not significant, so further research with a larger cohort is needed to determine if bacterial load could act as an indicator of hospital stay and therefore health.

In contrast to the results of the previous paragraph, where increased bacterial load indicates worse outcomes, this study found that an increase in bacterial load significantly affected survival to hospital discharge. Moreover, an increased bacterial load resulted in survival to hospital discharge. This and the previous findings contradict one another, which casts doubt upon the validity of the results. However, there are possible explanations as to why the bacterial load was significantly decreased in infants who did not survive. Firstly, as mentioned previously, the total bacterial of preterm infant stool increases with day of life. In addition, who are more seriously ill are removed from enteral feeds. Therefore, the samples included from the infants who did not survive may have been from early days of life. As once they became ill and were removed from enteral feeds, they would be less likely to pass stool and therefore provide further samples for the study. Unfortunately, this is not the case as the average day of life for the samples given by infants who did not survive to hospital discharge was 18 compared to 15, from infants who did survive. Secondly, the significant decrease in bacterial load, as mentioned previously, is likely to be a result of antibiotic use. As infants who did survive are more likely to have been the most seriously ill and will have therefore received a greater amount of antibiotics. This is true as 65% of samples taken from infants who did not survive were taken in the presence of antibiotics, whereas only 45% of samples taken from infants who survived to hospital discharge were taken in the presence of antibiotics. Overall, due to the conflict of results the total stool bacterial load is not a reliable variable for measuring the effect of the preterm gut microbiome on survival and length of hospital stay.

To further investigate the possible link between the preterm gut bacterial community and total stool bacterial load, the total bacterial load in samples dominated by particular

organism was analysed. The results of Figure 31, show that the bacterial load in samples dominated by different organisms, did vary but not significantly. The samples dominated by the *Bacteroidetes* phylum have the highest average bacterial load and the sample dominated by *Fusobacterium* contains the lowest bacterial load. However, overall there is not a lot of difference, therefore the predominance of a certain phylum is not linked to an overgrowth of bacteria. The same results can be seen at the genus level. This suggests the gut preterm gut microbial community changes in response to the increased growth of a specific organism. Moreover, it suggests that there have been decreases in other members of the community in order to accommodate this dominance by one specific organism or set of organisms. Otherwise, there would be an increase in bacterial load as a result of the dominance of the microbial community by a particular organism.

3.6.6 The Development of the Preterm Gut Microbiome during the First Month of Life.

The aim of this investigation was to confirm the results of previous findings and confirm the validity of the study by comparing the results with that of previously published research. Not only, does it confirm the results but also showed that the methodology employed during this study for sample collection, DNA extraction, and sequencing were correct and provide sufficiently accurate data to validate the study.

In line with numerous previously published research, the gut microbiome of preterm infants' transitions from a *Firmicutes* dominated environment to a *Proteobacteria* dominated community over the first 30 days of life. With the addition of the findings from this study, it can with high degree of confidence, that this is the normal development of the preterm gut microbiome during the first 30 days life. Furthermore, the infants were exposed to numerous to numerous differences in care, such as antibiotic and probiotic exposure, but as a cohort still produced this trend. Therefore, making this a firm outcome.

A further validation of previous and current findings is the individualistic nature of preterm infant gut microbiome development. This is more than likely a result of the tailored clinical care in the NICU to each infant's current and future needs. Moreover, a unique feature of the plasticity seen in the preterm infant gut microbiome, is that it has

the potential to be easily manipulated. Therefore, the potential of future therapies to change the preterm gut microbiome are an exciting area of research, because they have the potential to dramatically impact the development of the preterm gut microbiome. Ultimately, reducing the incidence of neonatal disease such as NEC.

Included as part of this investigation was the development of the preterm gut microbiome, ordered by day of life and predominant genus, Figure 34. This analysis was useful to determine the transition from a *Staphylococcus* dominated microbiome to either a *Klebsiella* or *Escherichia* dominated gut microbiome, occurred during the first 30 days of life. However, due to the number of genera included it makes any further deduction hard to interpret. Therefore, analysis into the change of individual genera over time, in future, should be conducted on an individual basis and displayed as such. This will improve presentation and interpretation of future data.

In summary, the results of previous and the study presented in this thesis have been further validated by the data presented in this thesis. As a result, the gut microbiome of preterm infants can be confidently stated as a transition from a *Firmicutes* to a *Proteobacteria* dominated state. Even though at an individual level the development of the microbial community varies significantly. This variability will improve efficacy of future therapies.

3.6.7 The Effect of Development on the Diversity of the Preterm Infant Gut.

Previous research has shown that gut microbiome diversity increases with age (Yatsunenko *et al.* 2012). Therefore, the aim of this investigation was to determine if this process also occurred in preterm infants, as the NICU provides limited exposure to potential gut colonizers. To the best of my knowledge there has been no previous reports on the change in gut microbiome diversity over time in preterm infants. The results of this study show that there is no significant increase in alpha diversity during the first 30 days of life in preterm infants. In more detail, the diversity only increases marginally until 3 weeks of life, where it falls drops. In contrast, measures of richness were discovered to peak at the beginning of life and steadily decrease till 3 weeks of. These results were the same across all three measures of evenness used during this study. There are numerous possible causes for these abnormal results. Firstly, during the

third trimester of pregnancy the foetus begins swallowing the amniotic fluid, thereby seeding the gut with organisms present in the amniotic fluid. As a result of premature birth this process does not get to occur, and the infant is delivered with a reduced number of bacteria present in the gut. However, this does not explain the high levels of richness seen at the beginning of life in Figure 35. As mentioned previously, this cannot be confirmed, but is likely a result of intrauterine infection, thereby increasing the presence of pathogenic bacteria. Secondly, and the most likely cause, is the near sterile environment the infants are exposed to in the NICU. If there is a low diversity of organisms present in the surrounding environment, then the gut is likely to be low in diversity also. Moreover, due to their extended periods of stay in the NICU the gut microbiome diversity is unlikely to increase significantly as a standard level of cleanliness is always achieved on the unit. Furthermore, the combination of a reduced environmental contribution, exposure to antibiotics, and a potential lack of enteral feeding all contributes to the low levels of diversity seen in preterm infant samples. In addition, the lack of a significant change during the first 30 days of life. For future studies into preterm gut microbiome diversity, all of the measures of richness and diversity seen in Figure 35, do not need to be used, they were used here in order to demonstrate the lack of significant changes. However, from this study the Simpson's diversity index and the SOBs give the most accurate information.

A key part of diversity analysis are the measures of beta diversity, which compare the differences in diversity between samples, often these samples are taken from different groups. The aim of this analysis was to determine if development has an impact upon the diversity between samples from early in life to those taken during the fourth week of life. This analysis showed that the dominance of organisms within samples remained unaffected by day of life and other clinical factors that were tested. This was an interesting finding as similar results have not been published, which also made explanations for this difficult to determine. One of the main explanations is entwined with the results itself. The fact that age, nor any other clinical factor, affected the predominance of certain organisms within the preterm infant gut, shows that there is no one specific driving force behind the development of the preterm gut microbiome community. Rather it is a multifactorial process, resulting in, sometimes significant, but subtle changes in the bacterial community, rather than individual drastic changes. Ultimately, this demonstrates that the key driving force behind the pattern of dominance

seen in the preterm gut microbiome is a result of prematurity. As this is the only common denominator between all the infants included in this analysis.

In summary, the results of this investigation are highly significant for this area of research. The knowledge that dominance in the preterm gut microbiome is a result of prematurity, as opposed to any clinical factor, allows for more focused research into the nuances of preterm gut microbiome development. Furthermore, it provides a new avenue of research to determine what factor of prematurity allows for the dominance of the preterm gut microbiome by organisms such as *Staphylococcus*, *Escherichia* and *Klebsiella*.

3.6.8 The Effect of Clinical Factors on the Development of the Preterm Gut Microbiome.

The aim of this analysis was to determine if the dominance of organisms within the gut microbiome of preterm infants was the result of an individual clinical treatment. The results of this investigation had the potential to influence future clinical treatment if it was found to significantly affect the health of the infant. However, the results of this study showed that no individual factor could account for the differences observed in the dominance of organisms within the gut of preterm infants. Therefore, the development of the gut microbiome of preterm infants is affected by multiple factors.

Even though an individual factor could not be used to explain the differences in dominant organisms in the gut of preterm infants, significant differences in microbial community can be explained by several factors. The first of which is gender. It was demonstrated that the abundance of *Proteobacteria* was significantly increased in male infants. Microbiome differences because of gender have previously been reported. More specifically, a 2016 study found that males had a greater abundance of *Proteobacteria* compared to females and the overall microbial community composition was different between the two genders (Cong *et al.* 2016). The results of this study are very similar to the findings of this project and in combination form the basis of our current knowledge on the impact of gender on the preterm infant microbial community. In addition, gender differences have been found to continue into adulthood (Haro *et al.* 2016). Various studies have speculated upon the origin of microbial community differences as a result of gender. Firstly, it has been suggested that hormone-immune-microbe interactions and

genetic traits are a likely cause (Gomez *et al.* 2015). Specifically, in adults' autoimmune disorders are more prevalent in females, when the gut microbiome is a contributing factor. Furthermore, sex specific hormones may shape the gut microbiome, as shown in dietary and autoimmune disease studies (Yurkovetskiy *et al.* 2013; Christine *et al.* 2015). However, this does not explain the differences seen in the pre-pubescent cohort of preterm infants, therefore there must be other mechanisms to explain gender differences beyond hormones. For example, studies in mice have found that males and females respond differently to diet, pathologic and probiotic organisms (Karunasena *et al.* 2014; Shastri *et al.* 2015). Moreover, the relative abundance of microbial species in the female GI tract were different compared to those in males (Karunasena *et al.* 2014). Furthermore, cytokine production was significantly different between the two genders in the colon, cecum and liver under basal conditions. This was repeated during experimental conditions. Therefore, demonstrating that gender is a key factor in the response to dietary and microbial changes (Karunasena *et al.* 2014; Shastri *et al.* 2015). These differences were found to extend to structural differences in the colon tissue of males and females, such as males had an increase of short chain fatty acids (butyrate and acetate) and females contained a greater concentration of o-phosphocholine or histidine (Karunasena *et al.* 2014).

In summary, sex specific metabolic and immune activities have been found to influence the gut microbial communities in adults and animal models. However, this research does not provide an explanation for the differences observed so early in life. Therefore, further research is needed to determine how immune functions and metabolic environment at the beginning of life affect the differences in microbial communities between the sexes.

Another key influencer on early life microbial community is mode of delivery. There has been a high volume of research into this area that has shown significant differences (Grönlund *et al.* 1999; Dominguez-Bello *et al.* 2010; Aagaard *et al.* 2012; Azad *et al.* 2013). However, mode of delivery has previously been found to have different impacts on the gut microbial community of preterm infants (Stewart *et al.* 2017). This is an interesting finding as the birthing procedure is the same for both full and preterm infants, therefore similar differences should be found in both groups of infants. In line with the research on full-term infants there were significant differences observed in the

microbial communities of male and female infants. The *Bacteroidetes* phylum was only found in infants delivered vaginally and infants delivered by CS had a higher abundance of *Proteobacteria*. Therefore, the results of this study contrast with those previously published in preterm infants. As a result, further investigations are needed to determine if mode of a delivery is a significant influence on the gut microbiome development in preterm infants.

Analysis into the effect of sample site on the development of the preterm gut microbiome provided an interesting result. It was determined that the difference in microbial communities between the sampling sites was a direct result of the administration of probiotics. The findings of these results are twofold. Firstly, that the administration of probiotics to preterm infants results in a significant increase in the abundance of the probiotic bacteria, *Bifidobacterium*. This bacterium was one of the two species included in the probiotic administered in the NICU, during the time of sampling. Therefore, probiotic bacteria can become constituent members of the microbial community of the preterm infant gut. Current research into the efficacy of probiotics is very mixed, with not one study able to conclusively state the effectiveness, or lack thereof, of probiotics (Al-Hosni *et al.* 2012; Janvier *et al.* 2014; Aceti *et al.* 2015; Abdulkadir *et al.* 2016a). As a result, a large scale, longitudinal study is required to conclusively determine the efficacy of probiotics in preterm infants.

In addition to probiotics, preterm infants are exposed to numerous feeding types dependant on several factors. Currently, mother's own breast milk is the accepted "gold standard" for both full and preterm infant feeding (Castanys-Muñoz *et al.* 2016). However, if mother's own milk is not available, formula and donor breast milk become the next options. These feeding routines have been shown to influence the type of bacteria acquired during gut microbiome development in both full and preterm infants (Cong *et al.* 2017; Timmerman *et al.* 2017). Furthermore, the gut microbiome diversity in infant's fed their mother's own milk was higher compared to the other feeding types (Cong *et al.* 2017). The results of this project show that diet has a dramatic impact the microbial community of preterm infants. For example, mother's own milk increased the presence of *Bacteroidetes,* and formula feeds increased the abundance of *Actinobacteria*. The difference in microbial community as a result of feeding routine is a result of the availability of different metabolites. For instance, the *Bacteroidetes*

phylum is known to metabolise fibre into short chain fatty acids more readily than other species (Aurélien *et al.* 2014). Therefore, dietary changes have the potential to manipulate the gut microbiome of preterm infants.

3.6.9 The Gut Microbiome of Preterm Infants who Develop NEC Significantly Differs from those who do not.

The aim of this analysis was to further investigate the potential links between the gut microbiome and development of NEC. A recent systematic review summarized the current evidence between the gut microbiome and the development of NEC. They concluded that the *Bacteroidetes* phylum is "protective" against the development of NEC i.e. the abundance of *Bacteroidetes* was increased in infants who did not develop the disease (Pammi *et al.* 2017). The results of this project confirm the conclusions of the Pammi *et al* study, with the *Bacteroidetes* phylum only present in samples taken from infants who did not develop NEC. Therefore, a "protective" effect of *Bacteroidetes* has been demonstrated, however, the mechanism of this protective effect has yet to be investigated.

An interesting finding of this research was that the *Bacteroides* genus was increased in infants who were breastfed but were decreased in infants who developed NEC. Therefore, adding further evidence suggesting that breastfeeding is protective against NEC.

Furthermore, similarly to previously published research, the infants in this study who developed NEC had a decreased gestation and birthweight. These are well established risk factors for the development of NEC (Lu *et al.* 2017; Duci *et al.* 2018).

In contrast to previous studies, an increase of *Klebsiella* or *Clostridium* was not associated with the development of NEC (de la Cochetière *et al.* 2004; Sim *et al.* 2015). Therefore, the cohort of infants recruited during this study was unique, as the *Enterococcus* genus was significantly associated with the development of NEC. Previous research has shown that *Enterococcus* species can produce proteases in the gut environment (Steck *et al.* 2013). Moreover, increased protease activity has been linked to the progression of the adult gastrointestinal disease, such as IBD (Carroll and Maharshak 2013). This process may be occurring in the gut of preterm infants during

the development of NEC. As a result, the increase in protease activity could be causing a breakdown of the tight junctions and necrosis of the gut epithelium. However, further research is needed to confirm the protease activity of preterm infant stool and the possible mechanism of action.

3.7 Conclusion

The development of the preterm gut microbiome is a highly individualistic and fluctuating process. However, as shown in this study and previous research, there is an overall progression towards a *Proteobacteria* dominated microbial community during the first 30 days of life. There are numerous factors affecting gut microbiome development in these infants such as gender and mode of delivery, but the effects are seen on a subtler scale as the dominance of organisms remains unaffected by these factors. In conclusion, this study provides further evidence towards the establishment of the gut microbiome community in the preterm infant and how exposure to early life factors affects the developmental process.

Chapter 4

The Protease and Inflammatory Content of Preterm Infant Stool.

Chapter 4. The Protease and Inflammatory Content of Preterm Infant Stool.

4.1 Introduction

Research into proteolytic enzymes began over 80 years ago as a result of increasing necessity to impede their action during the isolation and study of protein chemistry (Hans 1999). Since then, proteases have been shown to be essential regulatory components of numerous physiological functions such as mitochondrial maintenance, migration of neural crest cells during development, remodelling of skeletal muscles, blood coagulation and many more (Chakraborti and Dhalla 2017). Developments, such as those previously described, have led to the identification of proteases as potential targets for therapeutics.

Proteases are themselves polypeptides that contain a catalytic site, used for the degradation of proteins. The active site of each protease is specialised to target certain peptide bonds within the target protein. It is this specificity and highly regulated control mechanisms that make proteases highly effective components of cellular biology (Neurath 1989). The differences in mode of action has allowed proteases to be subdivided into several families as follows: threonine, aspartate, serine, cysteine and metalloproteases (Puente *et al.* 2005). As the name suggests it is specific amino acids that are the targets for proteases. The only exception are metalloproteases, who target metal ions on the inactive peptide.

More specifically, proteases are often part of complex pathways that involve the activation of several inactive precursor peptides, or zymogens. The active protein is released upon cleavage of the target peptide bonds, this cleavage can also result in structural changes in order to render the target active (Polgár 1989). Furthermore, the cleavage of a peptide bond by a protease can release an inhibitory peptide thereby activating the target peptide (Guasch *et al.* 1992). A typical example of a protease cascade is during nutrient digestion, such as pancreatic trypsinogen.

In summary, proteases are ubiquitous across the human body and critical for homeostasis. More specifically, the highest concentration of protease can be found in the gastrointestinal (GI) tract, this is not surprising as proteases are essential for the

breakdown of peptide bonds during food digestion (Antalis *et al.* 2007). Therefore, research into the content and function of proteases in the GI system is extensive and has provided interesting results.

4.1.1 The Role of Proteases in the Gastrointestinal System.

Proteases are present along the entire GI system, for example they are luminal, circulating, secreted, intracellular, intramembrane and pericellular. Once again, the plethora of locations for these molecules demonstrates their numerous functions in the human body. For example, pericellular proteolysis occurs as a result of the continual exposure of both apical and basolateral surface of intestinal epithelial cells (IECs) to circulating or secreted proteases (Medina and Radomski 2006). Furthermore, IECs themselves secrete proteases that remodel the extracellular matrix (ECM) (Medina and Radomski 2006).

The prevalence of proteases throughout the GI system indicates at the array of functions these molecules have in said system. For example, proteases are key components of signalling, with substrates such as growth factors and interleukins (Gschwind *et al.* 2001; Ohtsu *et al.* 2006). Moreover, a number of GI process are controlled by PARs, who are activated by proteases, some of these processes include the function of smooth muscle in the colon and small intestine (Vergnolle 2005). A further role for secreted proteases is in the maintenance of the intestinal mucosal barrier (Cenac *et al.* 2002). Furthermore, the epithelial layer underneath the mucosal barrier is also maintained by proteases. Matriptase increases the production of claudin-2 a tight junction protein (Buzza *et al.* 2010).

In order to maintain homeostasis, the activity of proteases must be tightly controlled. This can occur by the activation of zymogens, compartmentalization of active molecules, and the suppression or termination of activity by inhibitors (Antalis and Lawrence 2004; Jacob *et al.* 2005). In parallel to the gut microbial community, health is dictated by a balance of the types and numbers of proteases. Gastrointestinal disease has been linked to a disruption of these vital regulatory mechanisms (Bustos *et al.* 1998; Dunlop *et al.* 2006; Cenac *et al.* 2007; Róka *et al.* 2007b; Gecse *et al.* 2008; Shulman *et al.* 2008).

Combining knowledge of the microbial content of the gut microbial content and the significant role proteases have in the gastrointestinal system, research has shown bacteria also contribute the total proteolytic activity of the intestines. For example, serine, cysteine and matrix metalloproteases of bacterial origin have been found in the gut (Macfarlane *et al.* 1988; Gibson *et al.* 1989; Róka *et al.* 2007b). Akin to host proteases, protease of bacterial origin exists in many forms. For example, they can be excreted by the bacterium, remain attached to the cell surface or embedded in the membrane. Furthermore, bacterial proteases are components of numerous biological processes (Laskowska *et al.* 1996; Gottesman *et al.* 1997).

Specific to this thesis is the role of proteases in the gut of preterm infants. The primary source of proteases for preterm infants is breast milk. Several proteases have been found to be present in human breast milk, such as trypsin. Moreover, these proteases have been implicated in several physiological processes. However, little is known about the effect of several factors such as mode of delivery, may play in the presence and types of proteases present in the guts of preterm infants.

In summary, proteases in the GI tract are necessary to maintain homeostasis and there are complex mechanisms in place to regulate their activity. Furthermore, the dysregulation of these molecules has been found to contribute to the pathology of disease. In addition, like the microbiome, there are bacterial constituents to the total proteolytic activity of the gut. However, this thesis focuses on the preterm infant population.

4.1.2 The Role of IL-6 and IL-8 in the Gastrointestinal System of Preterm Infants

IL-6 is expressed by enterocytes in response to infection under the control of NF-κβ (Shimizu *et al.* 1990; Hunter *et al.* 2008b). The levels of IL-6 have been found to be elevated in the plasma of infants with NEC, furthermore, in concordance with severity of disease (Harris *et al.* 1994; Morecroft *et al.* 1994; Duffy *et al.* 1997). This indicates at a systemic response rather than a localised one, inflammatory levels would need to be found in the stool to indicate at a localised response. In a study of 62 new born infants with suspected sepsis or NEC, IL-6 levels were five to ten fold higher in infants with bacterial sepsis plus NEC at the onset of disease than in infants with bacterial sepsis

alone (Harris *et al.* 1994). This is not unsurprising as IL-6 is one of the products released a result of PAR activation. Furthermore, in a study of 60 preterm infants there was a trend to higher levels of IL-6 with a greater degree of NEC (Duffy *et al.* 1997).

IL-8 is generally regarded as a proinflammatory chemokine, and potent chemoattractant, predominantly produced by macrophages and endothelial cells. Importantly, exposure to amniotic fluid containing IL-8 has been indicated to be important for promoting intestinal health (Maheshwari *et al.* 2002). Not only are proteases present in breast milk, IL-8 is also present in significant concentrations. Furthermore, when human foetal and adult intestinal cells are treated with rhIL-8 *in vitro*, it stimulates cell migration, proliferation and differentiation (Maheshwari *et al.* 2002). IL-8 is also a potent chemoattractant for neutrophils and an angiogenic factor. As with several other proinflammatory cytokines, elevated IL-8 levels have been associated with human NEC and with an animal model of intestinal ischemia-reperfusion (Edelson *et al.* 1999; Chen *et al.* 2002; Benkoe *et al.* 2012). Furthermore, high IL-8 levels may correlate with human NEC severity (Edelson *et al.* 1999). One recent study reports that IL-8 appears to be a promising biomarker for the extent of intestinal necrosis (Benkoe *et al.* 2012). Cellular maturity may affect the response to bacterial challenge. Moreover, when compared to mature enterocytes, immature foetal intestinal cells have been shown to produce more IL-8 in response to LPS and flagellin (Nanthakumar *et al.* 2000). These developmental differences may predispose the premature intestine to inflammation.

4.2 Aims and Objectives

There is a significant lack of data on the role of proteases in the development of the preterm infant gut microbiome. Therefore, the aim of this study was to test the total faecal protease of preterm infant stool, in order to determine the significance of the role they play. Secondly, in order to potentially discuss roles for these proteases the types of proteases should be tested. Thirdly, proteases are integral members of the inflammatory response therefore it would be beneficial to investigate links between protease activity and inflammatory response of preterm infant stool. This will enable associations between protease activity and inflammation to be ascertained, in order to identify links with disease. Finally, all the previous data has the potential to be compared to the microbial content of the preterm infant gut. As a result, links between the microbial community and potential disease pathology could be made.

The above aims will be achieved by the following objectives:-

1. Determine the total faecal protease content of the preterm infant stool.
2. Use a suite of protease inhibitors to investigate the protease content of preterm infant stool.
3. Implement an ELISA assay to ascertain the levels of IL-6 and IL-8 in the stool of preterm infants.
4. Analyse the above findings, to compare the protease activity and inflammatory response of preterm infant stool. In addition, add data of the bacterial community to determine links between the microbial community, protease activity and inflammatory content.

4.3 Materials and Methods

All of the following assays were conducted on the 10% faecal slurry produced during Section 2.6.1. To determine total protease activity, assays were performed as detailed in Section 2.12.3. The results of these assays were then analysed according to Section 2.12.4, in order to determine the percentage of protease activity. One of the main aims of this study was to investigate the origin of the proteases present in the stool of preterm infants. Therefore, the following experiments were conducted. Firstly, assays using the ProteaseArrest kit from G-Biosciences were performed according to Section 2.15.1. When these experiments proved unsuccessful, it was decided that experiments into the families of protease present would yield better results, and these were conducted in Section 2.15.2 and 2.15.3. The results of these experiments were analysed according to Section 2.15.4, and showed the dominant proteases present in each sample.

In order to investigate the inflammatory response of the preterm infant gut ELISA assays were performed to determine the levels of IL-6 and IL-8 as detailed in Section 2.16. Analysis of these results was conducted as briefed in Section 2.16.3.

4.4 Results

4.4.1 Total Protease Activity of Preterm Infant Stool.

Firstly, it was important to determine if there was a change in total protease activity of preterm infant stool over time. Figure 43 below shows the average protease activity for the first 30 days of life. From this data it can be derived that there are no significant changes in total protease activity over time. However, when compared to the changes in bacterial load over the same period, a similar pattern can be seen. These data does not conclusively show that protease activity and bacterial load are linked, but that there is a similar peak at days 6 to 10 of life. To clarify the percentage protease activity has been derived from 1mg/ml of pure trypsin. Therefore, 1mg/ml of protein in the sample produces a percentage of the amount of activity in the same concentration of trypsin.

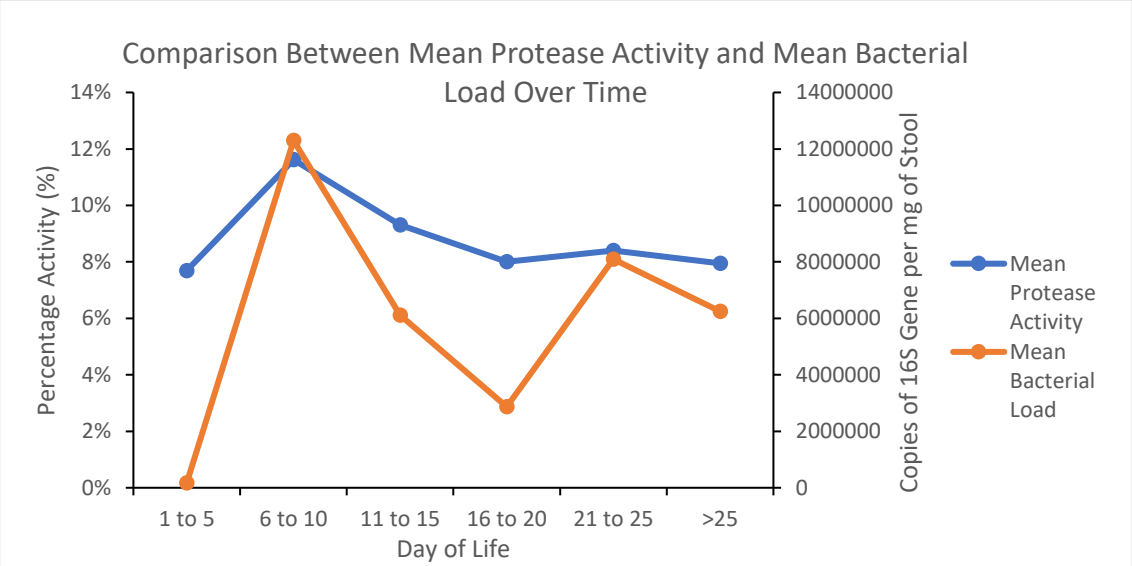

Figure 43. A Comparison between Mean Protease Activity and Mean Bacterial Load over time. The total number of samples included in this analysis was 199, 11 samples were excluded as they were collected from full term infants.

Once the collective protease activity was analysed it was imperative to investigate the change in protease activity on an individual infant basis. Figure 44, shows that the total protease activity differs between individuals. Furthermore, it shows that the patterns in different can be similar, infants 17 and 2 show the same increase. However, infant 6 shows a very different pattern of a sharp increase and decrease occurring at different times during the first month of life. Moreover, this infant shows the lowest amount of

activity at the end of the 30 days, whereas infant 2 has the highest activity at the start of the first month of life.

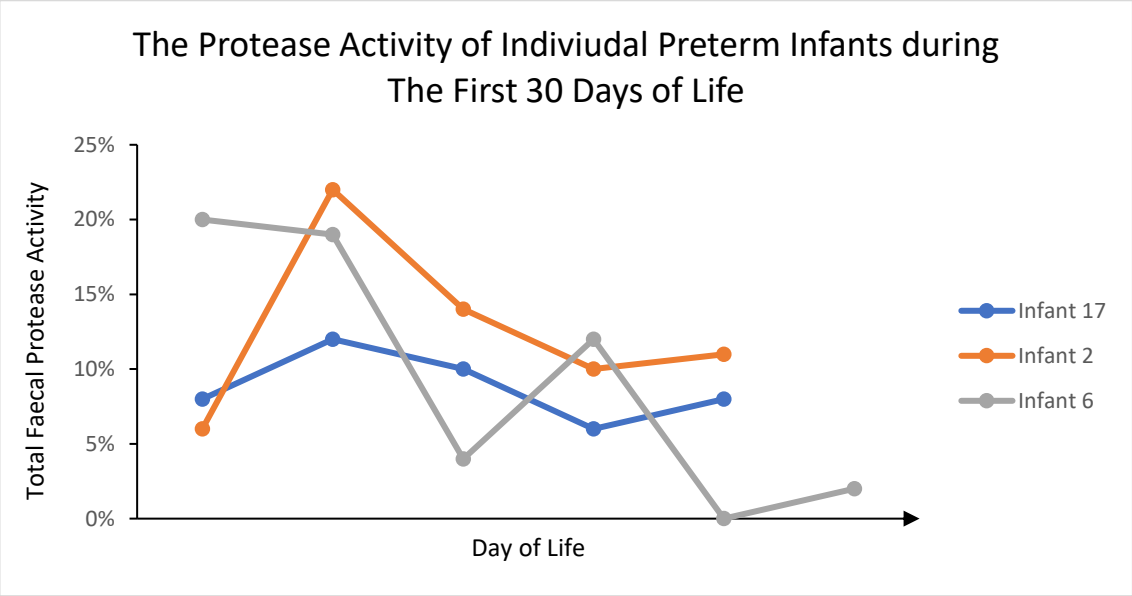

Figure 44. The Protease Activity of Individual Preterm Infants during the First 30 days of Life. The number of samples included during this analysis was 16. Day of life is given on the x axis as a continuous scale.

4.4.2 The Effect of Clinical Factors on the Total Protease Activity of Preterm Infant Stool

In a similar fashion to Chapter 3, the total protease activity of preterm infant stool was analysed in respect to several clinical factors. The first comparison investigated the role of proteases in the development of NEC. Figure 45, shows the mean protease activity for samples taken from infants who did and did not develop NEC. In addition, the samples taken from infants who did develop NEC have been grouped according to the severity of the disease. There are no significant differences here, but there is a definite decrease in protease activity detected in stool samples taken from infants who developed Grade 2 NEC. Overall, these results show that protease activity does not appear to influence the development or progression of NEC.

As discussed in detail in Chapter 4, antibiotics have a significant effect on the development of the preterm gut microbiome. Therefore, it was necessary to determine of there was a similar effect upon the protease activity of the stool. Interestingly, there was no difference in the mean protease activity between samples taken in the presence

of antibiotics compared to samples taken in the absence of antibiotics. The average total protease activity of the two groups was 9% compared to 1mg/ml of trypsin.

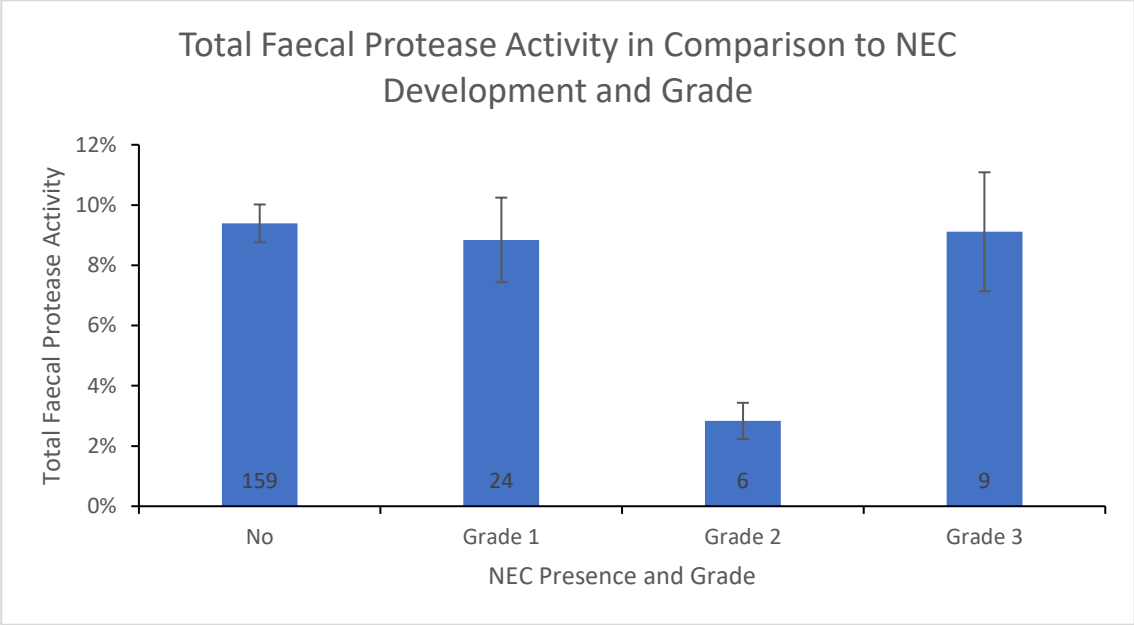

Figure 45. Total Faecal Protease Activity in Comparison to NEC Development and Grade. A total of 198 samples were included in this analysis, 12 were excluded as having been collected from full term infants. Each bar represents the mean protease activity for all samples included in their respective group. The numbers in each bar are the number of samples included in each group. The error bars for each bar are the standard error of the mean. NEC grades were defined according to Bell's staging criteria.

Figure 46 shows the effect of feeding routines on the total protease activity of preterm infant stool. There are no significant differences between the different feeding regimes, however there are some interesting changes. For example, the infants receiving no feeds show the lowest protease activity and the infants receiving formula feeds have the highest protease content. Moreover, donor and mother's own breast milk produce the same total faecal protease activity.

It was important to investigate the gender differences in total faecal protease activity, as there were significant differences in the microbial community. Moreover, the hypothesis is that bacterial proteases are contributing significantly to the total protease activity of the preterm infant gut. Therefore, it was necessary to investigate if these microbial differences translated into protease activity. The results of this analysis actually found no differences in faecal protease activity. The mean protease activity was 9% in both females and males.

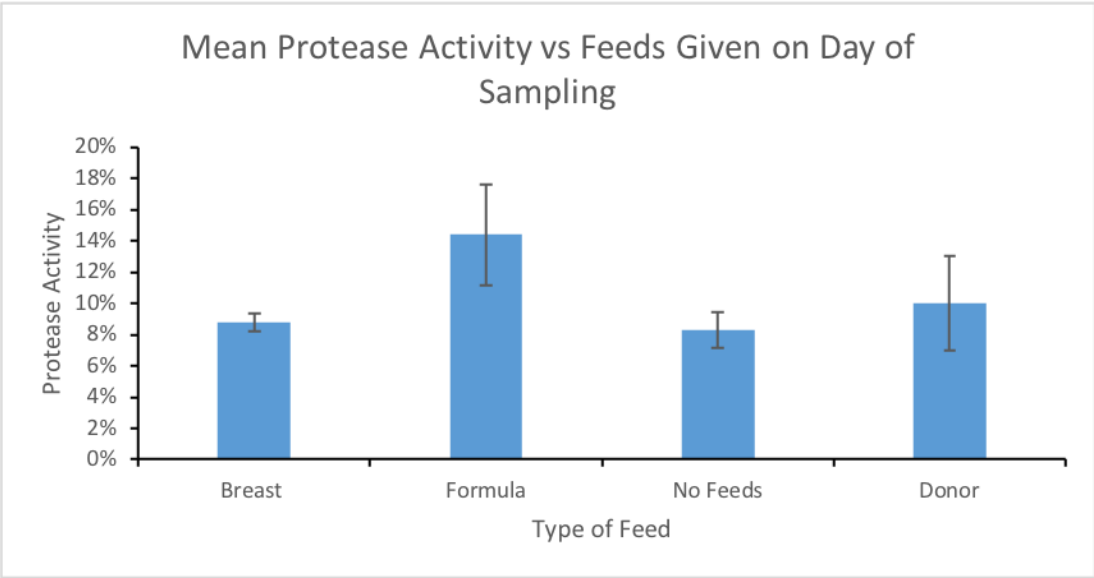

Figure 46. The Effect of Different Feeding Routines in the NICU on the Total Faecal Protease Activity. A total of 199 samples were included in this analysis, 11 were excluded as having been collected from full term infants. Each bar represents the mean protease activity. The error bars show the standard error of the mean.

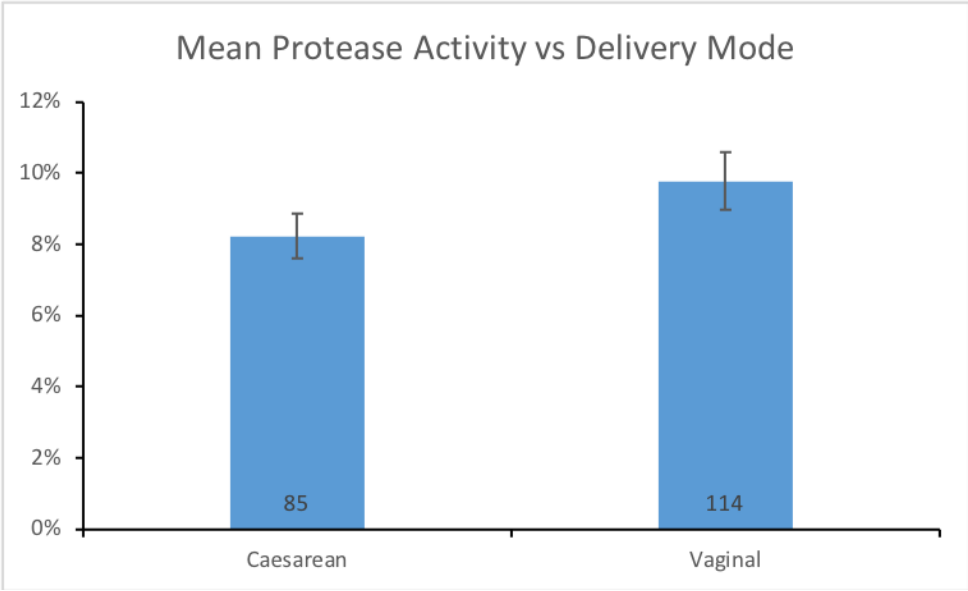

Figure 47. The Effect of Delivery Mode on the Total Faecal Protease Activity of the Preterm Infant Gut. The mean protease activity is shown by the bars. The error bars represent the standard error of the mean. The numbers at the base of the bar are the number of samples included in each group. In total 199 samples were used during this analysis.

Figure 47 shows the effect of delivery mode on the total faecal protease activity of the preterm infant gut. There are no significant differences as a result of delivery mode, however there is a slight increase in faecal protease activity from infants born vaginally.

4.4.3 The Families of Proteases Present in the Stool of Preterm Infants

In order to determine the families of proteases present in the stool of preterm infants, some initial optimization experiments were necessary. At first the ProteSeeker kit from G-Biosciences was used to determine the most dominant types of proteases present. Figure 48 shows the effect of 11 different protease inhibitors on the total faecal protease activity. To note this analysis was repeated on 5 more samples and the same results were obtained. From Figure 48 it can be determined that the inhibitors that produced the greatest inhibition were 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF), EDTA and Antipain. AEBSF is an irreversible serine inhibitor, EDTA is a metalloprotease inhibitor and Antipain is also a serine protease inhibitor. The results of this study showed that the dominant proteases present in the stool of preterm infants are serine and metalloproteases. Therefore, it was decided that further research into the types of proteases present in the stool would focus on the different proportions of serine and metalloproteases in the samples.

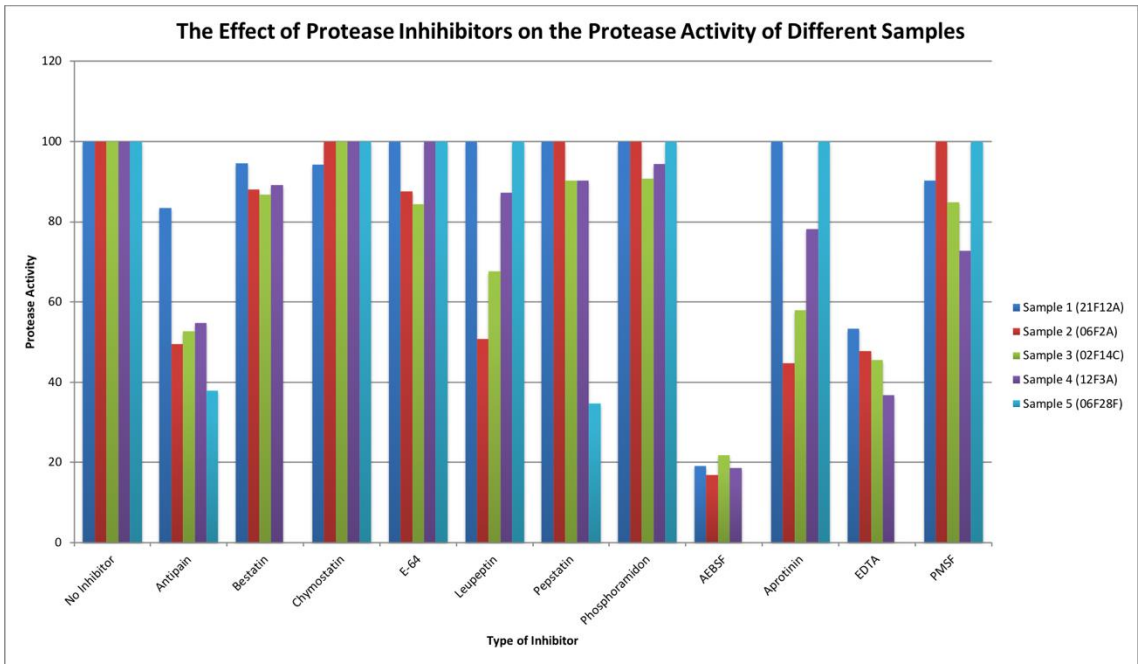

Figure 48. Investigation to Determine the Dominant Families of Proteases Present in the Stool of Preterm Infants. 5 samples were used during this analysis and 11 types of protease inhibitor were used to probe the samples. On the left the protease activity in the absence of an inhibitor is given.

Initial analysis found that protease profiles could be assigned to each sample based on the dominant family of protease present in that sample. Analysis of all the samples found that 35% of the samples were dominated by serine proteases, another 35% were

unaffected by the AEBSF and EDTA and therefore contained other proteases and were assigned a varied protease profile. In addition, 29% of the samples were equally inhibited by EDTA and AEBSF and therefore have a serine and metalloprotease dominated profile. Finally, only 3% of samples were solely dominated by metalloproteases.

4.4.4 Links between the Total Protease Activity and Types of Proteases Present in Preterm Infant Stool and the Microbial Community

Next it was important to investigate the links between protease activity, profile and microbial community. Figure 49, shows the results of the first of these investigations, to compare protease profile with protease activity over time. From the figure it can be found that there was no increase of a particular type of protease that coincided with the peak in protease activity between days 6 to 10 of life. The types of proteases present remain relatively stable during the first 30 days of life, however there is an increase in the serine and metallo protease profile at days 16 to 20 of life. Furthermore, the metalloprotease profile was only present at days 1 to 10 and 16 to 20 of life. Finally, the varied protease profile peaks at greater than 25 days of life.

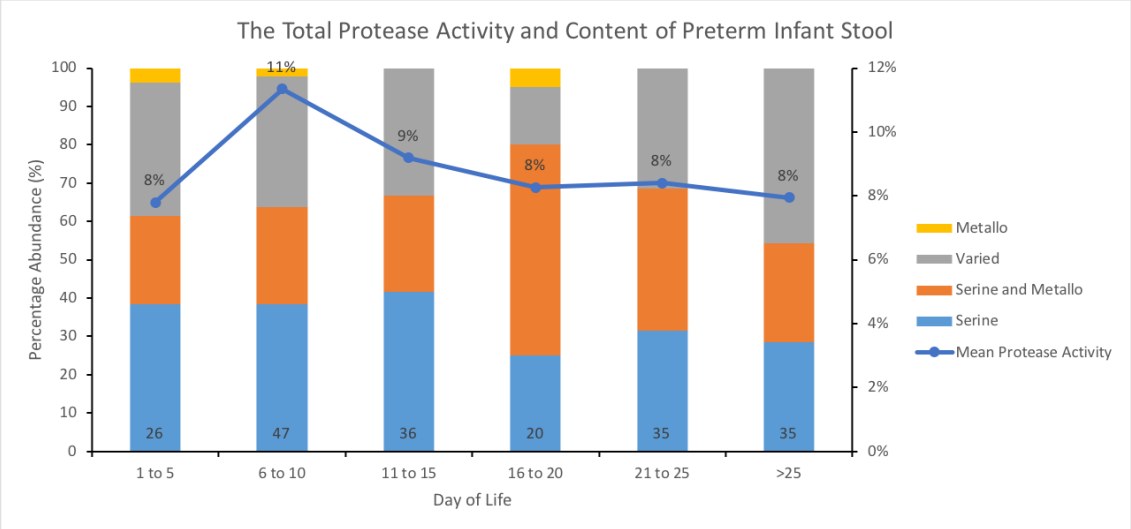

Figure 49. The Relationship between Total Faecal Protease Activity and the Protease Profile of the Preterm Infant Stool. The bars show the percentage abundance of each protease profile for samples during the development of the infant. The numbers at the bottom of each bar are the number of samples included in each days of life grouping. The line graph shows the mean percentage activity over time.

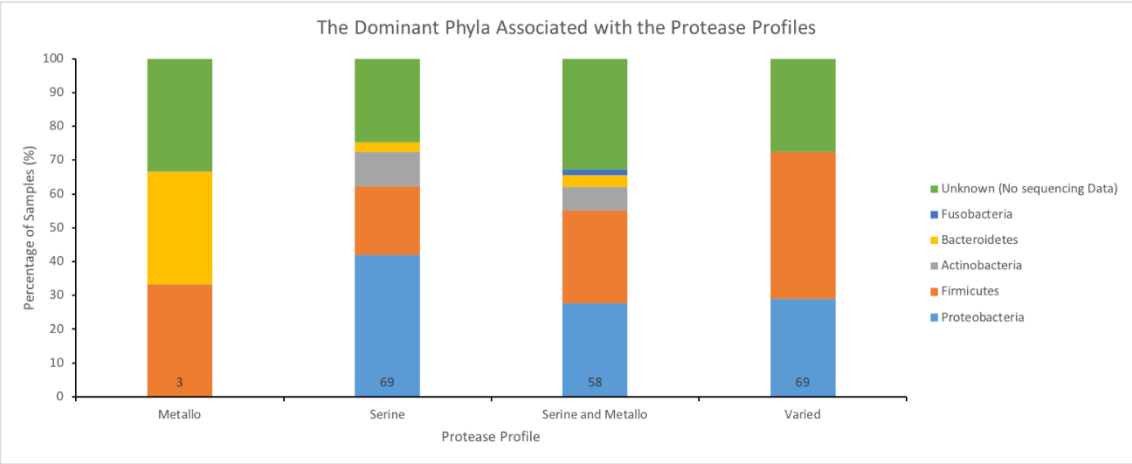

Figure 50. The Association between the Gut Microbial Community at the Phylum Level and the Protease Profile of the Preterm Infant Gut. The number of samples included in each protease profile group are shown at the bottom of each bar. Samples with an unknown phylum is because the sequencing was unsuccessful for these samples.

Figure 50, shows the relationship between the gut microbial community and the types of proteases present in the preterm infant gut. As shown, there are no significant differences in the protease profile between the gut microbial communities. However, the *Bacteroidetes* phylum is not associated with a varied protease profile. Furthermore, the highest level of serine proteases are seen in infants dominated by *Proteobacteria*.

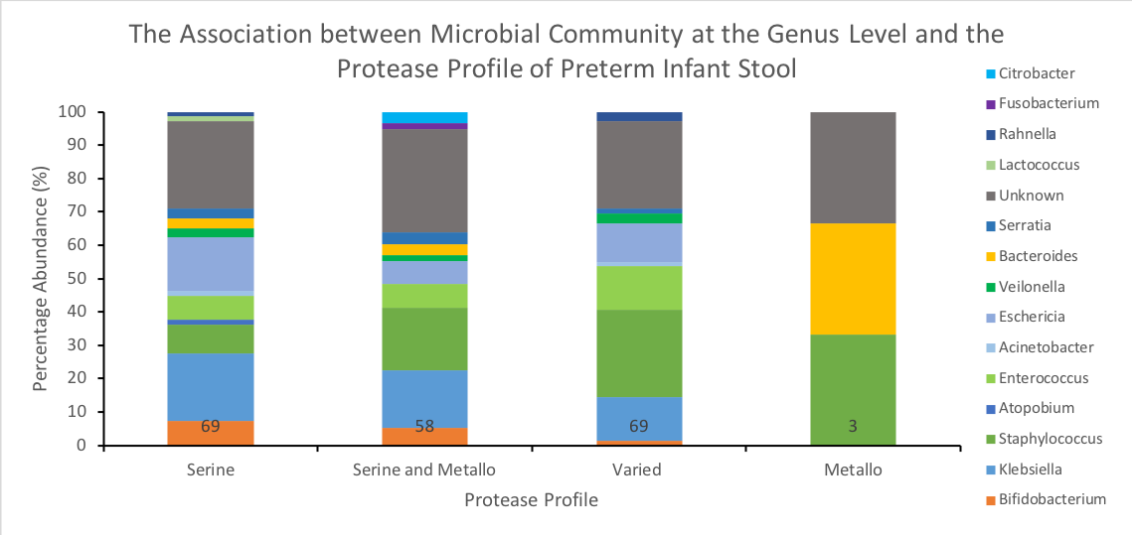

Figure 51. The Association between the Dominant Genus present and the Protease Profile of Preterm Infant Stool. The number of samples included in each protease profile group are shown at the bottom of each bar. Samples with an unknown phylum is because the sequencing was unsuccessful for these samples.

Figure 51, provides more detail on possible associations between the microbial communities in the preterm infant gut with the families of proteases present in the same

samples. To note there are no significant associations present and there are a large number of samples that were unable to provide sequencing. However, it is again worth mentioning that the *Bacteroides* genus was not present in samples that gave a varied protease profile. Furthermore, the greatest abundance of *Staphylococcus* was present alongside a varied protease profile.

In order to fully investigate the preterm infant microbiome, it was necessary to determine if an inflammatory response was present and could potentially be linked to the types of organisms and or proteases present. Furthermore, if protease activity could be linked to the inflammatory disease NEC.

4.4.5 The Inflammatory Response of Preterm Infant Stool

In order to investigate the presence of pro inflammatory cytokines in the stool of preterm infants a number of optimization experiments were conducted. This was to confirm that IL-6 and IL-8 could be detected in the stool of preterm infant, as there is no previous research on using ELISAs with preterm infant stool. The results of these experiments showed that IL-8 and IL-6 can be detected stool of preterm infants. However, when the ELISA was performed on 40 of the samples, there was no detectable presence of either IL-6 or IL-8. Therefore, the stool of preterm infants does not contain pro inflammatory cytokines, and as a result there is no detectable local inflammatory response in the gut of preterm infants. Finally, there was no difference in protease activity and inflammation between infants who did and did not develop NEC.

In addition, there was no association between protease activity and CRP and or WCC levels, taken as part of routine clinical care. Therefore, this provides further evidence that the protease activity or inflammation of the preterm infant gut is not resulting in significant systemic effects.

4.5 Discussion

4.5.1 The Total Protease Activity of Preterm Infant Stool

Overall, the results of this study show that there is no significant protease activity in the gut of preterm infants. This is not what was hypothesised as previous suggested that inflammatory driven gut diseases such as IBD, have a link with excessive protease activity. Therefore, the results of this study suggest that the preterm disease NEC does not have a contributory effect from proteases. Moreover, this is the first study to show this. There are several possible reasons for why the total protease activity of preterm infant stool is so low. Firstly, the total bacterial load of the preterm infant gut is significantly lower than that of adults. The average adult bacterial load is 10^{11} cells in the colon per ml, whereas the average found in this study of preterm infants was 10^{6} cells per mg of stool (Sender *et al.* 2016). This almost half of the levels seen in adults, therefore indicating that the guts of preterm infants are relatively sterile compared to that of adults. However, this is not unexpected as the gut microbiome is colonised throughout life and these infants are at the very beginning of their lives. Therefore, a significant reduction in the number of bacteria in the gut can explain a significant reduction in the total protease activity. Moreover, this only concerns the proteases produced by the bacteria themselves. The determination of the origin of the proteases has been a difficulty throughout this project. Therefore, it is more probable that the reduction in bacterial numbers accounts for the reduction in protease activity. There are other explanations for the low total protease activity, and these will be discussed later in this section.

Taking this into consideration, when comparing the pattern of activity alongside total bacterial load, even though it is not significant, bacterial load and protease activity peak the same time. Therefore, suggesting that total protease activity of the preterm infant gut is from the host.

Finally, the total protease activity of preterm infant stool shows individualistic patterns which is mirrored in the individualistic nature of the development of the microbial community. This is an interesting similarity in development pattern and adds further evidence to the individualistic nature of preterm gut microbiome development, previously noted. This individualism is a result of the differing treatments, feeding

routines, gestation, gender etc., and it is interesting to see this having a similar effect on the protease activity of the preterm infant gut.

In summary, there is not a significant amount of protease activity in the gut of preterm infants. Moreover, this activity is not linked to bacterial load, which indicates that the primary origin of these proteases is host. Finally, the development of the protease activity in the gut is unique to each infant, a pattern that is also seen in the development of the gut microbial community.

4.5.2 The Effect of Clinical Factors on the Total Protease Activity of Preterm Infant Stool.

One hypothesis of this project was that the development of NEC in preterm infants is linked to protease activity in the stool. As there is no research in protease activity of preterm infant stool, the hypothesis was based upon research in adults. Gut diseases in adults, such as IBD, have been linked to excessive protease activity found in the stool, therefore, it was important to investigate if the same links could be drawn. However, the results of this experiment showed that there were no statistically significant associations between stool protease activity and the development and progression of NEC. However, this was not the main focus of the PhD project and therefore the recruitment of infants was not centred on infants who developed NEC. As a result, the number of infants recruited who developed NEC was low. So even though no associations were found in this study, a focused investigation needs to be conducted to convincingly say that NEC is not associated with protease activity.

Antibiotics and their effect of the gut microbial community was discussed, at length, in Chapter 5. However, it was important to investigate the effect antibiotics could be having on the gut protease activity. Furthermore, it would also assist in determining the origin of the proteases. The results showed that antibiotics had no effect on the total protease activity of the preterm stool. This is in stark contrast to the significant effect that was seen on the microbial community. Therefore, this adds further evidence towards the host origin of the proteases as the total activity remains unaffected by changes in the bacterial community.

Similar to antibiotics, differences in gender resulted in significant changes to the microbial community, however, this did not translate to total protease activity. There was no change in total protease activity in the stool of male and female infants. This is another incidence indicating to the host origin of the proteases, because significant changes in microbial community did not result significant changes in total stool protease activity.

Although not significant feeding routines did cause a small change in protease activity. Previous research has shown that one of the key functions of proteases is in the breakdown of luminal contents, in other words digested food. Moreover, bacteria have been shown to assist in the breakdown of luminal contents to release molecules inaccessible by normal digestive processes. Therefore, this shows that diet in preterm infants has the potential to affect the total protease activity of preterm infant stool. However, the number of samples taken from infants receiving formula feeds was minimal and a larger study would need to be conducted to see if these results translate into significant differences.

Delivery mode also resulted in a minor change in protease activity, where on average a vaginal delivery resulted in a higher total protease activity. Due to the absence of research into this area it is hard to comment on the reasons why this difference might have occurred. Even though previously throughout this Chapter, results have been presented to show that the infant produces the proteases detected in the stool, the differences in delivery mode could be a result of changes in microbial communities.

In summary, several clinical factors did not have a significant effect on the total protease activity of preterm infant stool. This could be a result of the overall limited detectable activity. However, the evidence indicates that total protease activity of the preterm infant stool has a limited involvement in the development of the gut microbiome.

4.5.3 The Types of Proteases of the Preterm Infant Gut Microbiome.

The aim of these experiments was to determine the families of proteases present in the stool of preterm infants. Initially, it was necessary to determine if certain families of proteases dominated over others in the stool of preterm infants. The predominant

proteases were found to be serine and metallo proteases. Once again, due to the lack of research in this area, the role of these proteases can only be speculated upon. Serine and metalloprotease have been found to play significant roles in the permeability of the intestinal barrier and intestinal barrier inflammation. The preterm infant gut is known to be susceptible to excessive intestinal inflammation and permeability. Therefore, it is not unsurprising to see these types of proteases present in greater numbers than others.

Once it was found that serine and metalloproteases dominated the proteases present in the stool of preterm infants. All samples were then tested to determine the protease profile based on the percentage inhibition by a serine and metalloprotease inhibitor. The results found that the stool of preterm infants was very rarely dominated by metalloproteases alone. This indicates that serine proteases are necessary in the recruitment of metalloproteases, as previous research has shown proteases are often part of a complex activation and inhibition pathway.

Preliminary experiments showed that serine and metalloproteases were predominant in the stool of preterm infants. However, after testing all samples with only AEBSF and EDTA, a significant proportion of the samples showed a varied protease profile. This meant that serine or metalloproteases did not dominate that sample. This is interesting as it showed that the protease profile of the preterm infant gut was not as simple as initially thought. Ultimately, this showed that a more diverse suite of inhibitors should be used in future experiments in the determination of the types of proteases present in the stool of preterm infants.

In summary, the protease content of preterm infant stool appears to be more complex than the total activity. This indicates that in a proportion of the infants a variety of proteases are present even though the overall activity detected is low. This could be a result of the sample tested, stool. Therefore, the stool could only reflect the protease activity and types present in the colon. Whereas there may be more activity further up the GI tract. This suggest that the proteases in the gut of preterm infants participate in a variety of functions, contributing to the development of the infant.

4.5.4 The Links between Protease Type, Activity and Gut Microbial Community.

This analysis was used to determine the interactions between several components of the preterm gut microbiome. However, the results were inconclusive, and no significant observations could be concluded.

There was no relationship observed between total protease activity and the organisms present in the gut microbiome. This shows that there is not a significant contribution to total protease activity by a specific organism. This once again adds further evidence to the absence of bacterial proteases in the stool of preterm infants, as certain bacteria cannot be identified as contributing an amount of protease activity. Furthermore, there were no significant associations between the organisms present and the types of proteases present in the stool of preterm infants. This shows that one organism does not significantly contribute a single protease type to the total content of the stool. Overall, this suggests that disease processes occurring in the gut of preterm infants should not be contributed to one specific organism, as they do not contribute a significant amount of protease activity or a specific type of protease.

4.5.5 The Role of Inflammation in the Gut Microbiome of Preterm Infants.

The aim of these experiments was to investigate the pro inflammatory levels of IL-6 and IL-8 in the stool of preterm infants. There has been little evidence of the presence of these proteins in the stool of preterm infants (Moerch *et al.* 2008; Lusyati *et al.* 2013), and they were tested using high throughput sensitive methods. Therefore, an ELISA method was used to try and detect levels of IL-6 and IL-8 at concentrations indicative of localised inflammation.

After several optimisation experiments it was determined that IL-6 and IL-8 could be detected in the stool of preterm infants using an ELISA. However, no detectable IL-6 or IL-8 could be found. This indicates that the excessive inflammatory response associated with diseases such as NEC is a systemic response rather than a localised effect. Furthermore, the absence of an association between the microbial community in the gut and an inflammatory response it suggests that a specific bacterium is not associated with

the development of NEC. I was confident about the method used as levels of IL-6 and IL-8 could be detected in other samples collected from the same infants.

On the other hand, the IL-8 and IL-6 could have not been detected because they had degraded before the ELISA could be conducted. For instance, the faecal sample was extracted and processed in order to preserve the activity of the proteases present. This could have led to a breakdown of the cytokines. Furthermore, in order to normalise all samples to 1mg/ml total protein, the samples were diluted. This could have resulted in diluting the IL-6 and IL-8 to undetectable levels by ELISA. Future studies in this area should aliquot samples separately for protease and cytokine investigations.

4.6 Conclusion

In summary, the role of proteases in the development of the preterm infant gut microbiome does not appear to be significant. Including infants with and without NEC. The total protease activity of preterm infant stool is minimal and not associated with bacterial load or clinical factors such as antibiotics use. This substantially indicates at a host origin for these proteases. However, with the methods implemented during this PhD, a definitive conclusion cannot be drawn. Moreover, the predominant types of proteases present were serine and metalloproteases. However, a proportion of the samples tested produced a varied protease profile where serine and metalloproteases did not predominate in the sample. Therefore, the protease content of the preterm infant gut can be complex even in absence of a high degree of activity. When protease type and activity was compared to the microbial community no significant observations were recorded, again providing further evidence towards the host origin of these proteases. Finally, no detectable levels of IL-6 and IL-8 could be found in the stool of preterm infants. Overall, this indicates at a very minimal role of proteases in the development of the gut microbiome.

Chapter 5

The Effect of Antibiotics on the Developing Preterm Gut Microbiome.

Chapter 5. The Effect of Antibiotics on the Developing Preterm Neonatal Gut Microbiome.

5.1 Introduction

The microbiome, is a unique ecological niche, shared between the human host and the variety of organisms that call it home. Moreover, it comprises of all metabolites, from inhabitants and host, and their interactions. There are numerous microbiomes present on the human body, each very different from one another. The gut microbiome is one of the most complex environments and so has become a developing area of research over that past decade (Marchesi 2011).

In the UK, compared to the rest of Europe, approximately 15 defined daily doses (DDD) of antibiotics per 1000 inhabitants were consumed (Goossens *et al.* 2005). Therefore, discovering the effect of these drugs on the gut microbiome is of particular interest. Furthermore, an example of the widespread use of antibiotics is that the most commonly prescribed medication in the NICU was antibiotics (Hsieh *et al.* 2014), this indicates that preterm and vulnerable infants are the most susceptible to the effects of antibiotics.

In 2009, the effects of antibiotics on the developing microbiome was fist noted. It was found that and increase in *Enterococcus* and a decrease in *Bifidobacterium* during the first week of life, as a result of antibiotic consumption. Furthermore, they found an increase in the *Enterobacteriaceae* family to persist 1 month after treatment (Tanaka *et al.* 2009). A further study in 2011, found the abundance of *Bifidobacterium* to be reduced in infants who had received parenteral antibiotics compared to naïve infants (Hussey *et al.* 2011). A more recent longitudinal study observed the effects of antibiotics to remain during the first 3 years of life. They also found the gut microbiome to be less diverse and treated infants were more likely to display compositional change in consecutive samples (Yassour *et al.* 2016).

This microbiome change has been associated with long term outcomes. A study in Finnish children showed that antibiotic use between the ages of 2 and 7 years, resulted in a decrease of *Actinobacteria*, alongside an increase in *Proteobacteria* and *Firmicutes*.

Furthermore, the macrolide type of antibiotic resulted in an increase in asthma and weight gain in treated infants (Korpela *et al.* 2016). A population based study comprising of over 10,000 healthy children, median age of 24 months, concluded that antibiotic exposure in the first 6 months of life, or repeated exposure during infancy, significantly increased body mass (Saari *et al.* 2015). A further study found that antibiotic exposure was associated with obesity at 4 years of age, with more than 3 courses of antibiotics resulting in an increased risk at 2 years of age (Scott *et al.* 2016). Therefore, the effects of antibiotics can persist well past the time of administration.

Current research has been unable to determine a solid link between the preterm gut microbiome and disease (Abdulkadir *et al.* 2016b). However, an emerging link between antibiotic therapies in preterm infants, a disrupted microbiome, and an increased risk of disease development is becoming established. Furthermore, the most prolific drug used in NICUs is currently antibiotics (Clark *et al.* 2006). Therefore, investigations into the effect of this most common drug on the preterm neonatal population is of great importance.

The effects of parenteral antibiotic administration were examined in preterm infants. Both short- and long-term treatment caused a significant reduction in the number of *Bifidobacterium* for the first 3 weeks of life. Furthermore, the *Enterococcus* genus was significantly increased for up to 12 weeks after antibiotic treatment was stopped (Zwittink *et al.* 2018).

Several studies, over the past decade, have shown antibiotics to increase the incidence of common preterm disease, such as NEC and sepsis. More specifically, the empirical use of antibiotics has been demonstrated to increase the rates of NEC, LOS, mortality and hospital stay (Cotten *et al.* 2009; Alexander *et al.* 2011; Kuppala *et al.* 2011; Afjeh *et al.* 2016). Moreover, empiric antibiotic use significantly increased the abundance of *Enterobacter*, alongside NEC, sepsis and or death (Greenwood *et al.* 2014). In addition, antibiotic use in preterm infants with no evidence of infection lead to an increased in NEC diagnosis and or death (Alexander *et al.* 2011; Esaiassen *et al.* 2017). The research shows that empiric antibiotic use is more destructive than constructive in the NICU. To add, it has been shown that empirical antibiotic use accounts for 39% of inappropriate antibiotic use compared to only 4% as a result of antibiotic initiation (Patel *et al.* 2009).

A key preventative in the care of neonates is the use of perinatal antibiotics in women with the potential for preterm birth or with confirmed Group B *Streptococcus* (GBS) infection. Several studies have shown a significant increase in the *Enterobacteriaceae* family in infants born to mothers who received antibiotics. Furthermore, these effects have been shown to persist months after birth (Arboleya *et al.* 2015; Arboleya *et al.* 2016; Azad *et al.* 2016; Mazzola *et al.* 2016). Interestingly, recent research has shown that this antibiotic exposure before birth to be protective against neonatal disease. A very recent study showed a significant decrease in the incidence of NEC in preterm infants born to mothers who received antibiotics up to 72 hours before birth (Reed *et al.* 2018). Furthermore, prophylactic use of antibiotics for suspected GBS colonization did delay the colonization by *Bifidobacterium*, but this was normalized by 12 weeks of life (Stearns *et al.* 2017). In a recent study the effects of both prenatal and postnatal antibiotics were determined. The abundance of *Bacteroidetes* was significantly reduced in infants exposed to antibiotics before birth. Whereas, the abundance of *Bifidobacterium* was significantly reduced in infants exposed to a high level of antibiotics after birth (Zou *et al.* 2018).

In the SCOUT observational study there was no impact to outcomes such as NEC or survival between empirical and limited antibiotic use (Cantey *et al.* 2016). An earlier report published by the same group also showed that only 5% of antibiotic use was accountable by culture proven sepsis (Cantey *et al.* 2015). This and research from numerous NICUs around the world show that the use of antibiotics in the NICU is open to considerable variation, and, as discussed earlier, results in detrimental outcomes (De Keukeleire *et al.* 2016). In Australia and New Zealand it was found that empirical antibiotic use is minimal (Carr *et al.* 2017). In contrast, a study in the Netherlands found that antibiotic consumption varied from 130 to 360 DDD per 100 admissions. This is considerably higher than the doses given in the UK. Furthermore, 24 different antibiotics were used across the Dutch NICUs (Liem *et al.* 2010). A further study included 127 NICUs across California and over 50,000 infants. They discovered a 40 fold variation in antibiotic use across the centres, ranging from 2% to 97% of days sampled where antibiotics were prescribed (Schulman *et al.* 2015). Finally, in Canada it was reported that antibiotic use in infants with no proven infection was associated with higher rates of adverse outcomes, such as NEC (Ting *et al.* 2016).

As mentioned previously, the gut microbiome comprises the entire contents of the ecosystem including metabolites. In order to fulfil this definition, we investigated the activity of luminal proteases. More specifically, the effect of antibiotics of these microbiome components. Proteases have been associated with the development and progression of gastrointestinal disorders, such as IBD, and therefore should be investigated in a preterm cohort (Bustos *et al.* 1998; Dunlop *et al.* 2006; Cenac *et al.* 2007; Róka *et al.* 2007b; Gecse *et al.* 2008; Shulman *et al.* 2008). Furthermore, using antibiotics to determine an effect on protease activity may provide insights into the origin of these molecules, bacterial or host.

As a result, there is little consensus on the appropriate use of antibiotics in the NICU, and more research needs to be done in order to elucidate the full effect these drugs are having on such a vulnerable population (Shah and Sinn 2012). In addition, to advise future clinical practice on the use or misuse of these highly accessible drugs. Therefore, the aim of this investigation was to determine the effect of antibiotic administration on the gut microbiome of preterm neonates.

5.2 Aims

It can be concluded that the definitive effect of antibiotics on the preterm gut microbiome has yet to be decided. However, the use of these drugs without clinical proof of infection or sepsis can lead to common preterm disease such as NEC and or sepsis. Moreover, the effect antibiotics are having on the microbial community of the gut microbiome has yet to be investigated. There is a strong probability that antibiotics are causing a change in community structure within the gut microbiome, which is then leading to the increased incidence of preterm disease. In addition, ass mentioned previously, the protease activity of the preterm infant has yet to be investigated. Therefore, the effect of antibiotics on these proteins will provide insight into their role within the gut microbiome. Taking this into consideration, the aims of this study were:

1. To determine the effect of antibiotic administration on the preterm gut microbiome bacterial community.
2. To investigate the effect of antibiotics on the protease activity of the preterm gut. This will provide information on total protease activity, families of proteases present and organisms of origin.

5.3 Materials and Methods

Samples were collected as detailed in Section 2.2, and only samples from preterm infants were included in this study. These samples were then processed according to the methods detailed in Section 2.6. The resulting faecal slurry was used to determine the total protease activity using the protocol detailed in Section 2.12.3, and subsequently analysed as detailed in Section 2.12.4.

DNA was extracted on the resulting pellet from the methodology used in Section 2.6, as per Section 2.8.2. Preparation of the extracted DNA and the sequencing of the 16S rRNA gene was performed as listed in Section 2.9. The resulting sequencing data was analysed according to the methods detailed in Section 2.11. Furthermore, data on total bacterial load was performed by qPCR, according to Section 2.17, on the extracted DNA.

5.4 Results.

5.4.1 Patient Recruitment and Sample Demographics

Please refer to Table 14 as it shows the patient demographics for both the Study Cohort and Microbiome Cohort, the Microbiome Cohort contains samples taken from infants that produced more than 1000 reads during sequencing. All samples collected from infants admitted to the post-natal ward (PNW) did not sequence sufficiently, and were removed from the final data set, the Microbiome Cohort. There was no change in the average gestation and average maternal age between the two cohorts. However, the birthweight of the preterms increases slightly in the Microbiome Cohort. The length of hospital stay remains consistent between the two cohorts. Lastly, there was a decrease in positive sequencing results from the survival and NEC groups.

It was found that antibiotics had a significant effect on the positivity of sequencing results, $p < 0.001$. Therefore, antibiotics have a significant effect on the success of sequencing during microbiome research in preterm infants. This effect of antibiotics on the success of sequencing can be explained due to a decrease in bacterial load. Figure 52, demonstrates a significant reduction in total bacterial load during antibiotic administration. This is reinforced using a non-parametric Mann-Whitney U independent samples test, p value < 0.001. Therefore, antibiotics significantly decrease the total bacterial load and subsequently the success of gut microbiome sequencing.

5.4.2 The Preterm Gut Microbiome as a Result of Antibiotic Administration

Figure 53 A, shows an increase in *Proteobacteria* in the samples taken in the absence of antibiotics, compared to an increase in *Firmicutes* in the samples taken during a course of antibiotics. Furthermore, the abundance of *Actinobacteria* is greatly increased in the absence of antibiotics. In contrast, the abundance of *Bacteroidetes.*

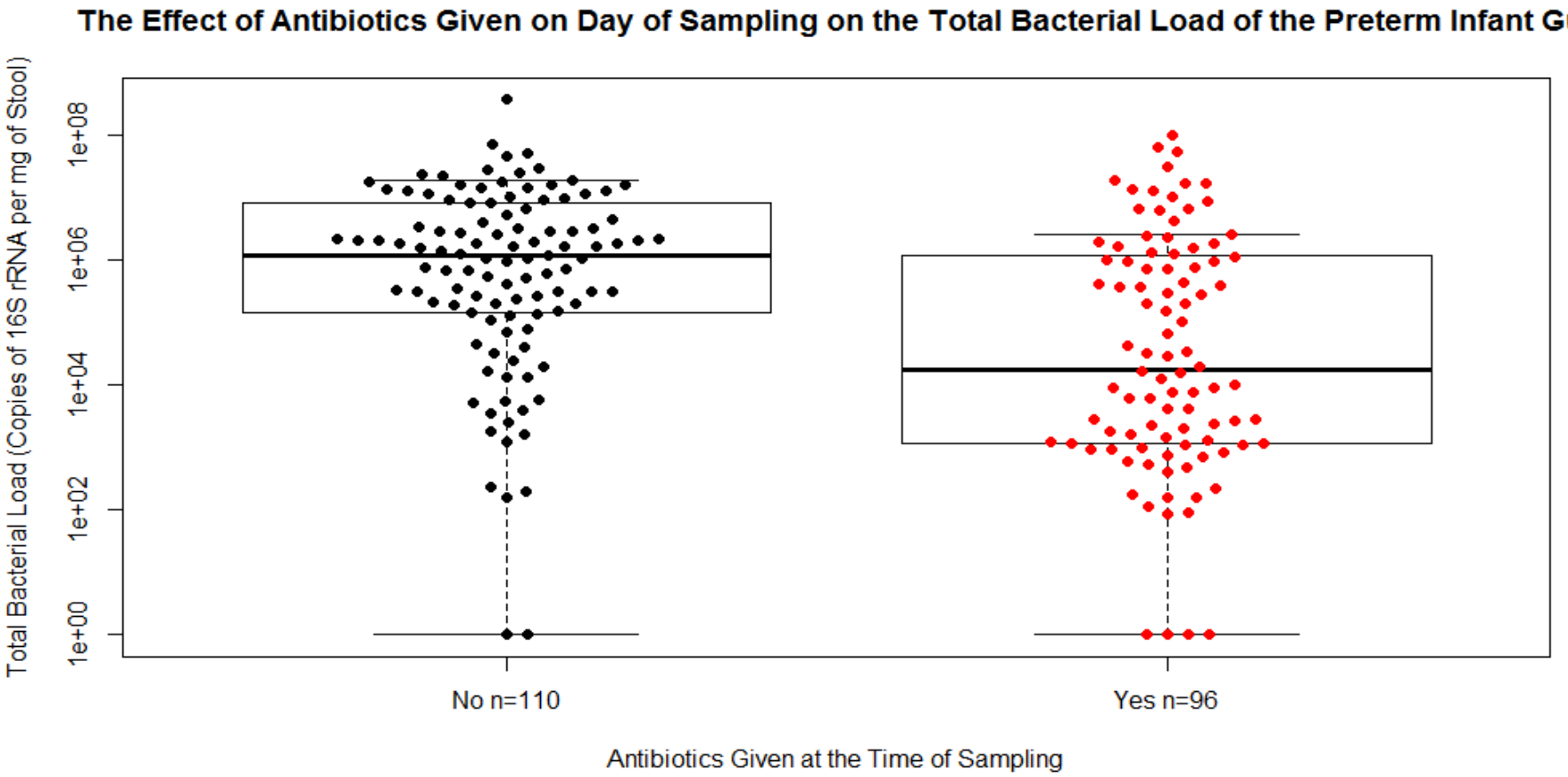

Figure 52. The Effect of Antibiotics on the Bacterial Load of the Preterm Infant Gut. Significance testing, non-parametric Mann-Whitney U independent samples test, gave a p value < 0.001, between the bacterial load of samples during a course of antibiotics, and those taken when antibiotics were not prescribed. A total of 4 samples were removed in order to construct this graph. Any sample that was taken ± days after a sample taken during antibiotics was removed. Each dot on the graph is a sample taken during an antibiotic present or free condition. The range is given by the whiskers either side of the box plot. The outer sides of the box show the upper and lower quartiles, with the mean shown by the bold black line in the middle of the box plot.

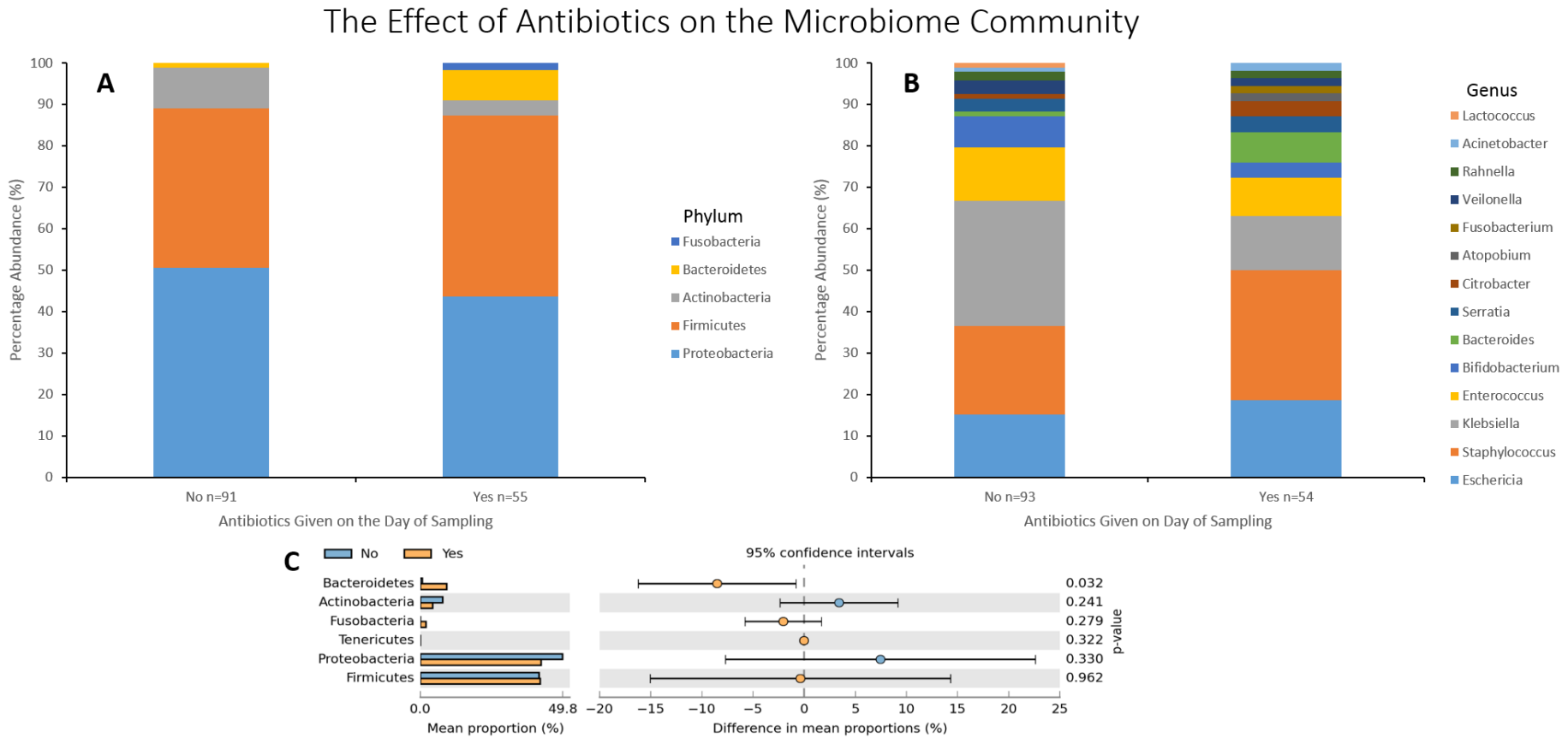

Figure 53. Changes in the Gut Microbiome of Preterm Infants as a Result of Antibiotic Administration. Panel A shows the relative abundance of the phyla present in the samples taken during antibiotic administration and those that were not. Panel B shows the relative abundance of the genera present in the samples taken during antibiotic administration and those that were not. On panels A and B the y axis refers to the percentage abundance of each organism in that group and the x axis shows is the samples were taken on and off antibiotics. The numbers on the x axis shows the number of samples in each group. Panel C is the results from a Welch's t-test using the STAMP software and shows the phylum level differences between the antibiotic present and absent groups. The confidence intervals are shown with the black lines and the bottom axis. The bars on the left of panel C shows the percentage abundance of that phyla within each group.

predominant in samples taken in the presence of antibiotics. Moreover, in Panel C this *Bacteroidetes* increase is significant, $p < 0.03$, therefore the presence of antibiotics allows the *Bacteroidetes* phylum to flourish. Overall, there are clear changes in the microbiome as a result of antibiotic use, however the full extent of these changes cannot be demonstrated at the phylum level. Panel B of Figure 53, shows the relative abundance of the genera present in the samples taken during antibiotic administration and those that were not. Firstly, there is a minimal difference in the common gut commensal *Escherichia* between the two groups. On the other hand, there is an observable difference in the abundance of *Staphylococcus*, *Klebsiella*, *Enterococcus*, *Bacteroides* and *Bifidobacterium*. The *Staphylococcus* and *Bacteroides* genera appear to be able to survive antibiotic administration compared to other genera.

In Figure 54, the gut microbiome communities of samples taken from infants who were continually exposed to antibiotics and those who were antibiotic naïve are shown. This figure reinforces the differences seen in Figure 53, and provides further evidence towards the effect of antibiotics on the gut microbiome community in preterm infants. Firstly, the *Bacteroidetes* phylum is only present in the samples taken from infants consistently exposed to antibiotics. Similarly, *Actinobacteria* are only present in samples taken from infants who never received antibiotics. The *Proteobacteria* and *Firmicutes* phylum are present in both groups, but in varying proportions. At the genus level, the first difference is the increase in *Escherichia* in the antibiotic naïve group, this is mirrored by a similar increase in *Staphylococcus* in the empirically exposed group. Furthermore, the *Bifidobacterium*, *Veillonella* and *Enterococcus* are only seen in the negative antibiotic group. In the same manner, only *Rahnella*, *Bacteroides* and *Citrobacter* are present in the individuals receiving multiple courses of antibiotics.

Once differences in the microbial community was determined, the changes in diversity were also investigated. The alpha diversity was measured using several different indices and no significant differences in alpha diversity was found between samples taken in the

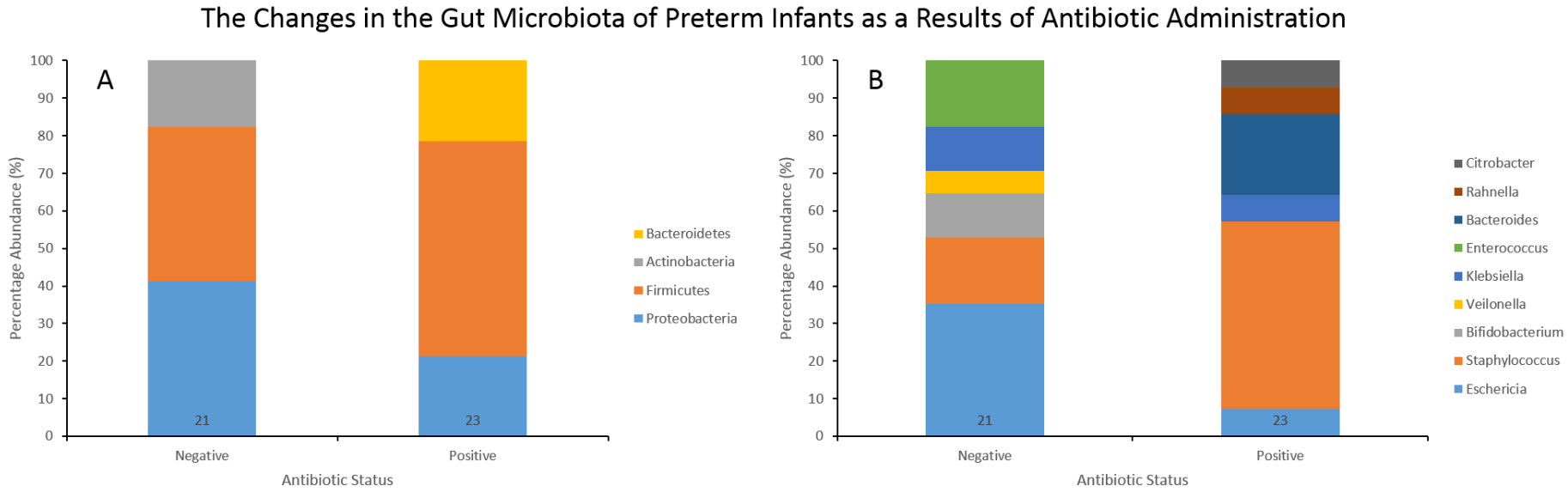

Figure 54. The Microbiome Community in Antibiotic Naïve and Empirically Exposed Infants. This figure was constructed using only data from infants who had multiple samples taken whilst all samples were either antibiotic positive or negative. The total number of samples used during the analysis was 44. Panel A shows the taxonomic differences between the groups at the phylum level. Panel B shows the taxonomic differences between the groups at the genus level.

presence or absence of antibiotics, Figure 55. Furthermore, the samples did not cluster according to principal co-ordinate analysis (PCoA), Figure 56. As a result, the preterm gut microbiome diversity remains unaffected by antibiotics, therefore the antibiotics cause a shift in predominance rather than an increase or decrease in the abundance of certain organisms.

Finally, the protease activity was measured in preterm infant stool taken in the presence and absence of antibiotics. There was no difference in protease activity as a result of antibiotic exposure. Furthermore, there was no change in the families of proteases present due to antibiotic exposure.

5.4.3 The Effect of Antibiotics on the Gut Microbiome of Samples taken from Recruitment Sites

With the data showing a decrease in the presence of *Actinobacteria* and *Bifidobacterium* in the presence of antibiotics, it was interesting to determine if this was a centre specific event. This is because only UHW administered the probiotic, Infloran, during the sampling period. Figure 57, clearly shows this dramatic effect antibiotics have on the gut microbiome community containing probiotics. Panels B and D show a substantial difference in *Actinobacteria* and *Bifidobacterium*, respectively, in the presence and absence of antibiotics.

The results of this analysis are twofold. It is important to find that the probiotics administered are becoming a stable member of the gut microbiome community in these infants. However, the administration of antibiotics is reducing the, if any, beneficial effects of these probiotics. This raises the question of whether probiotics should be given during a course of antibiotics.

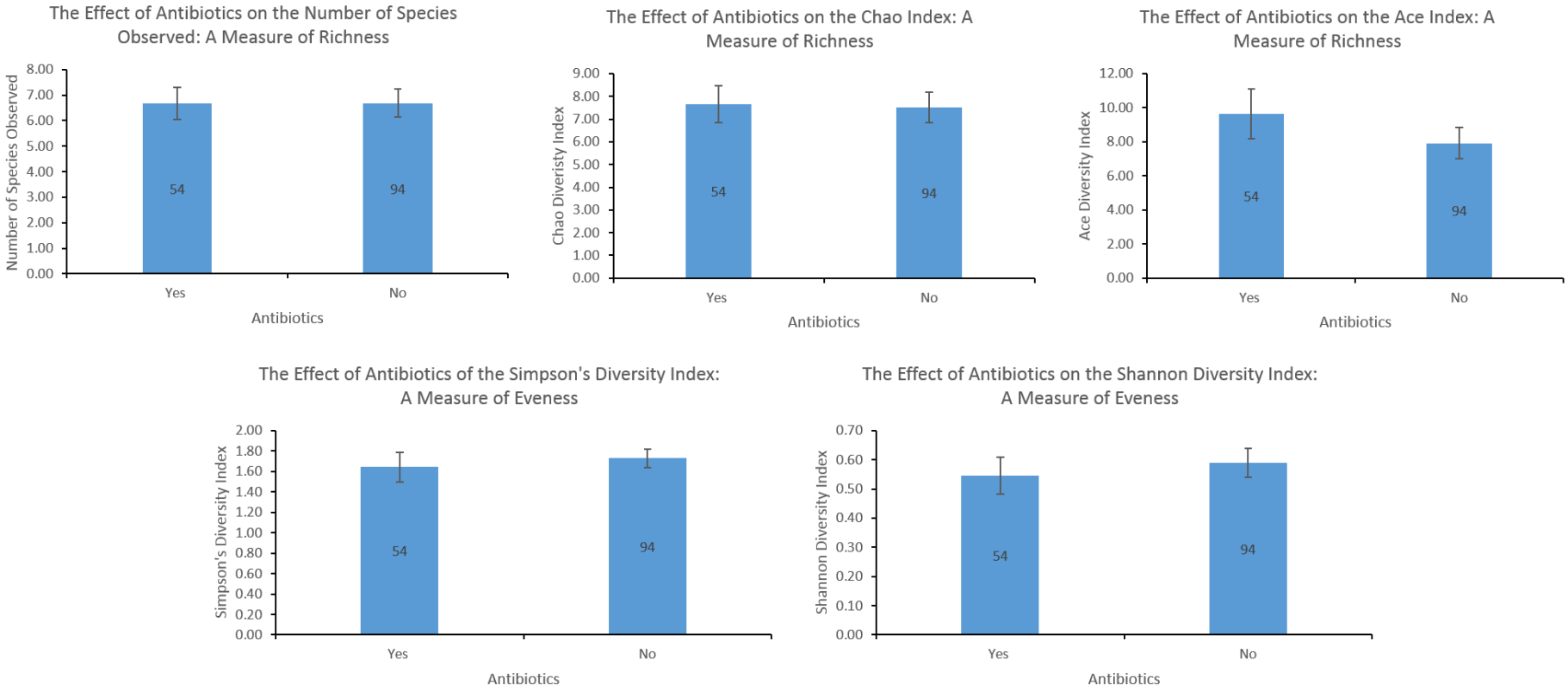

Figure 55. The Effect of Antibiotics on Several Alpha Diversity Indices. This analysis was conducted on 148 samples, including samples that sequenced successfully with samples removed if ± 3 days of a sample taken whilst on antibiotics.

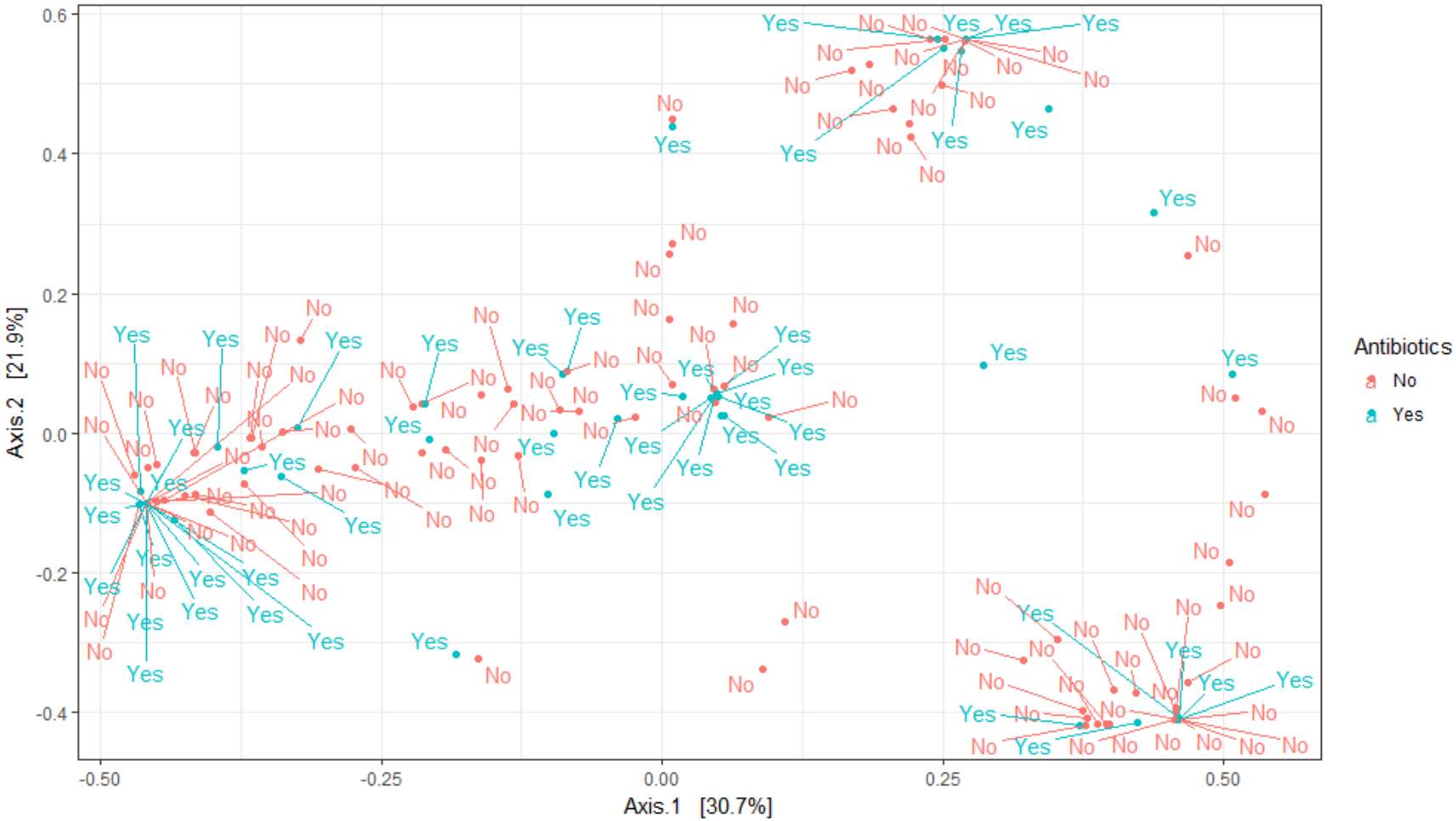

Figure 56. The Effect of Antibiotics on the Beta Diversity of the Preterm Infant Microbiome. This analysis was completed using 148 samples, including samples that sequenced successfully with samples removed on the basis of being ± 3 days of a sample taken in the presence of antibiotics. This is a Principal Co-ordinate Analysis (PCoA) using the Bray-Curtis dissimilarity index to ordinate the samples based on their dissimilarity to one another, in relation to taxonomic community. In other words, the samples positioned furthest apart on the graph are the most dissimilar to one another. In total the axis accounts for 53.8% of the difference between samples.

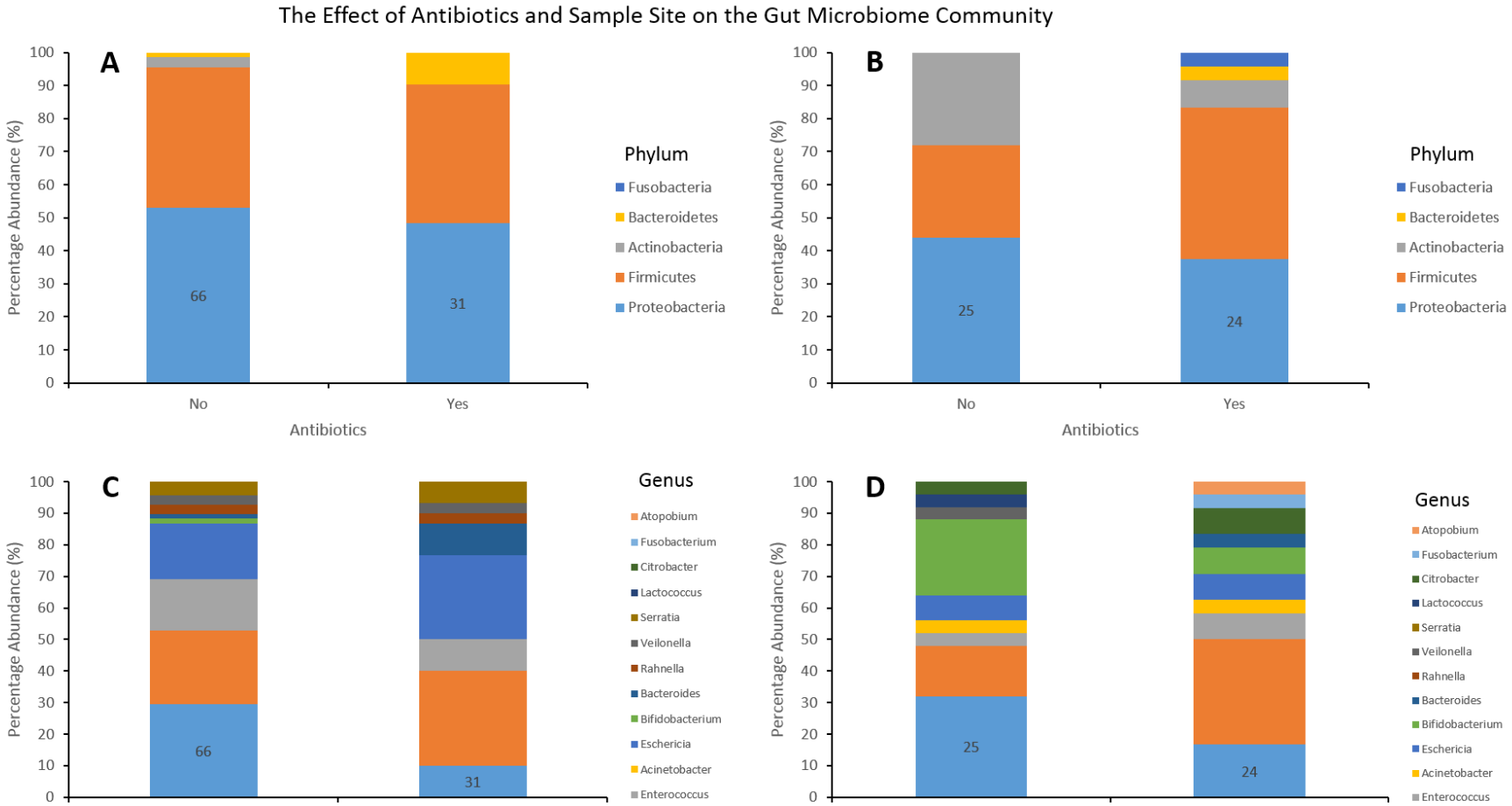

Figure 57. The Effect of Antibiotics on the Microbiome at Different Sampling Sites. Panel A and C shows the gut microbiome communities at the phylum and genus levels from North Bristol Trust (NBT). Panels B and D show the gut microbiome communities at the phylum and genus levels from University Hospital Wales (UHW). The y and x axis on all graphs show the percentage abundance of each organism and the presence of antibiotics, respectively. The numbers in each bar refer to the number of samples contributing to this community.

5.4.4 The Effect of Antibiotics on the Development of the Gut Microbiome in Preterm Infants

The most striking feature of Figure 58 are the large error bars. This is a result of the large variation in samples within the group. Firstly, there is a substantial peak in bacterial load at days 6 to 10 of life for samples taken in the absence of antibiotics. This peak does not occur in the samples taken in the presence of antibiotics. However, this group does peak, but it is much later at days 21 to 25 of life. Irrespective of antibiotic administration, both groups decrease in bacterial load at days 16 to 20 of life, and again rise at days 21 to 25. In addition, both groups start at a very low bacterial load. Finally, bacterial load is maintained after 25 days in the samples taken in the absence of antibiotics, where in the presence of antibiotics the bacterial load sharply decreases after 25 days of life.

Figure 59 shows the relative abundance of *Proteobacteria* is increased in the sample taken in the absence of antibiotics, compared to an increase in *Firmicutes* in the antibiotic exposed samples. There are more *Actinobacteria* present in samples taken during a course of antibiotics. Furthermore, there are more *Bacteroidetes* present in the sample taken in the presence of antibiotics, more specifically during the first 2 weeks of life. In contrast, there is only a minor appearance of the *Bacteroidetes* in the samples taken in the absence of antibiotics occur at greater than 25 days of life.

From Figure 60 it can be clearly seen that there is a greater abundance of *Klebsiella* in the samples without antibiotics, whereas the presence of *Staphylococcus* and *Escherichia* is much higher in the samples on antibiotics. In Panel A of Figure 60, a steady decline of *Enterococcus* overtime can be observed whereas this is much more sporadic in the presence of antibiotics. The *Bacteroides* are present from days 1 to 15 of life in the presence of antibiotics, whereas in the absence of antibiotics they only appear in small numbers between days 21 to 25 of life, this coincides with the highest bacterial load.

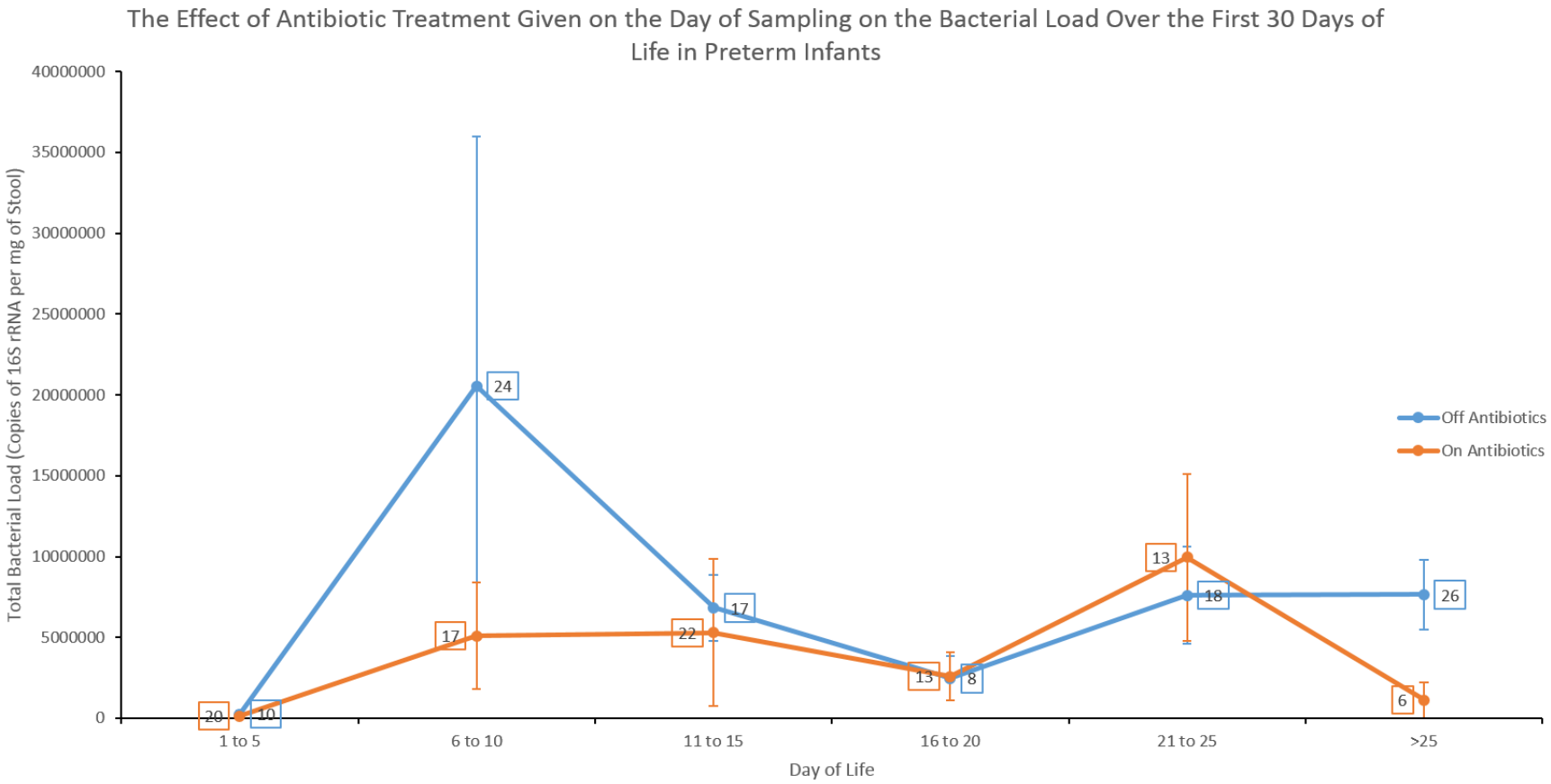

Figure 58. The Effect of Antibiotics at the Time of Sampling on the Development of the Gut Microbiome. A total of 200 samples were used during this analysis, as 10 were removed from the Study cohort for being from full term infants and the following reasons. If a sample was taken $\pm$ 3 days of another sample, this sample was removed. Secondly, if more than one sample from the same infant was given within the same day of life grouping e.g. days 6 to 10, if one of those samples was on antibiotics and the other not, this sample was taken and the other removed. However, if the 2 samples were either on or off antibiotics the mean bacterial load of the 2 samples was taken. This resulted in a final total of 188 samples used during this analysis. The number of samples in each group is given by the box next to the data, the box is outlined the same colour as the data series it represents. The same applies to the error bars which represent the standard error of the mean. Each plot for both data series represents the mean bacterial load from all the samples within that age group.

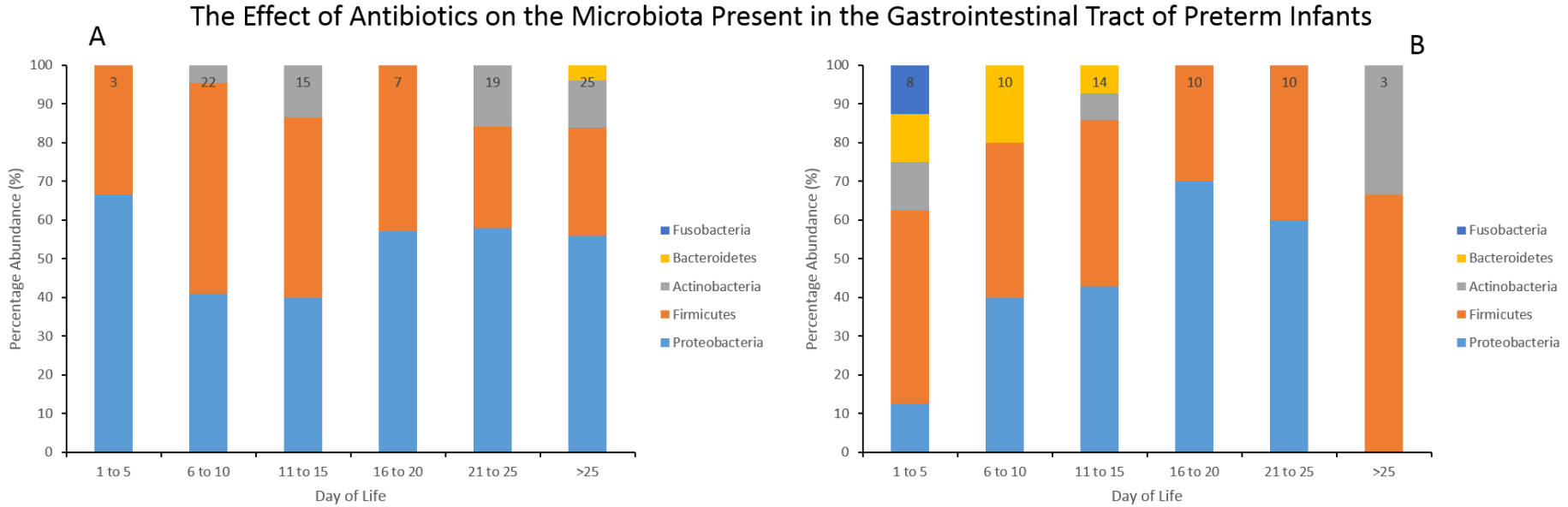

Figure 59. The Effect of Antibiotics on the Development of the Preterm Gut Microbiome. In total 146 samples were used during this analysis. This is a result of sequencing efficacy and the removal of samples taken ± 3 days of a sample taken when on antibiotics. The numbers included at the top of each bar are the number of samples included in each day of life. Panel A shows the samples taken in the absence of antibiotics and Panel B shows the samples taken in the presence of antibiotics. Both of the graphs shows the microbiome community at the phylum taxonomic level.

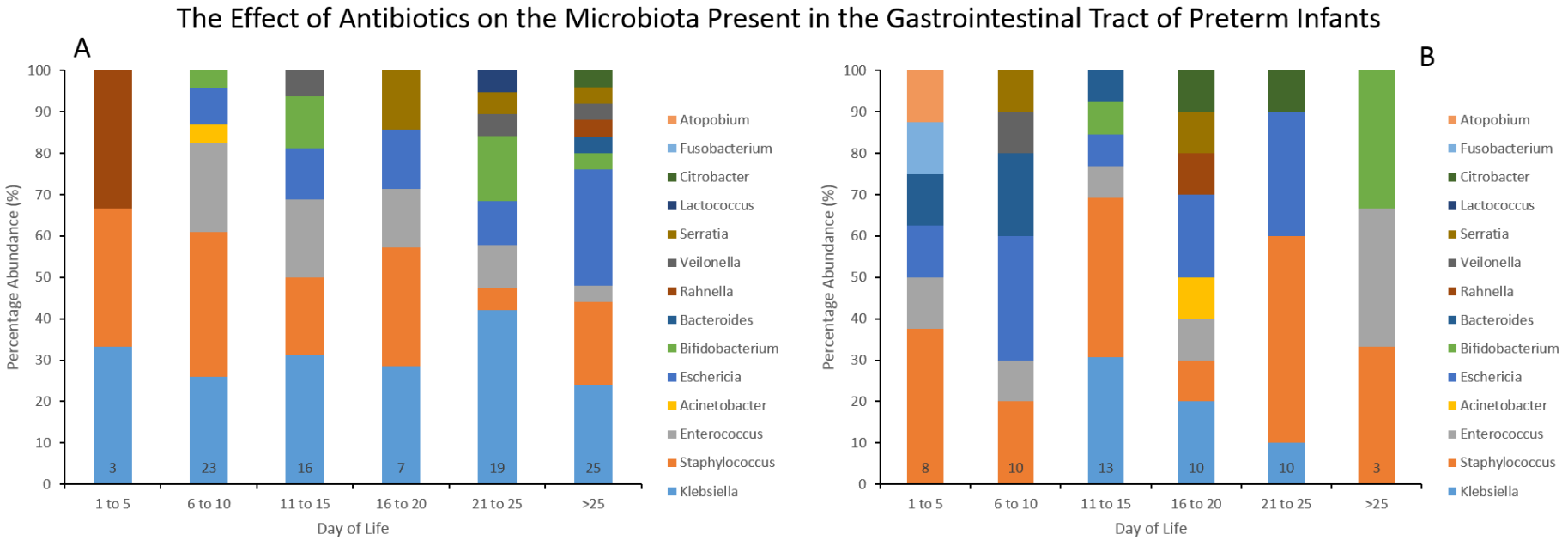

Figure 60. The Effect of Antibiotics on the Development of Preterm Gut Microbiome. In total 146 samples were used during this analysis. This is a result of the sequencing efficacy and the removal of samples taken $\pm$ 3 days of a sample taken when on antibiotics. The numbers included in the bottom of each bar on the graphs are the number of samples included in each day of life grouping. Panel A shows the genera present in samples taken over time in the absence of antibiotics. Panel B shows the genera present in samples taken over time in the presence of antibiotics.

Finally, the effect of antibiotics on the protease activity of preterm infant stool over the first 30 days of life is shown in Figure 61. The protease activity over time in both conditions follow a similar pattern until days 16 to 20 of life. After this time, the samples taken in the presence of antibiotics increase, compared to a further decrease during the same time in the absence of antibiotics. The reverse of the previous situation occurs between days 21 to 25 of life, then both groups decrease after 25 days of life.

In summary, new to the field of preterm research, antibiotics were found to have a significant effect on the positivity of sequencing results and bacterial load. This change in bacterial load was reflected in significant community changes in the microbiome, more so when the data is split by sampling site. Both the bacterial load and microbiome community developed differently over the first 30 days of life in the presence of antibiotics. Finally, antibiotics had no impact on protease activity.

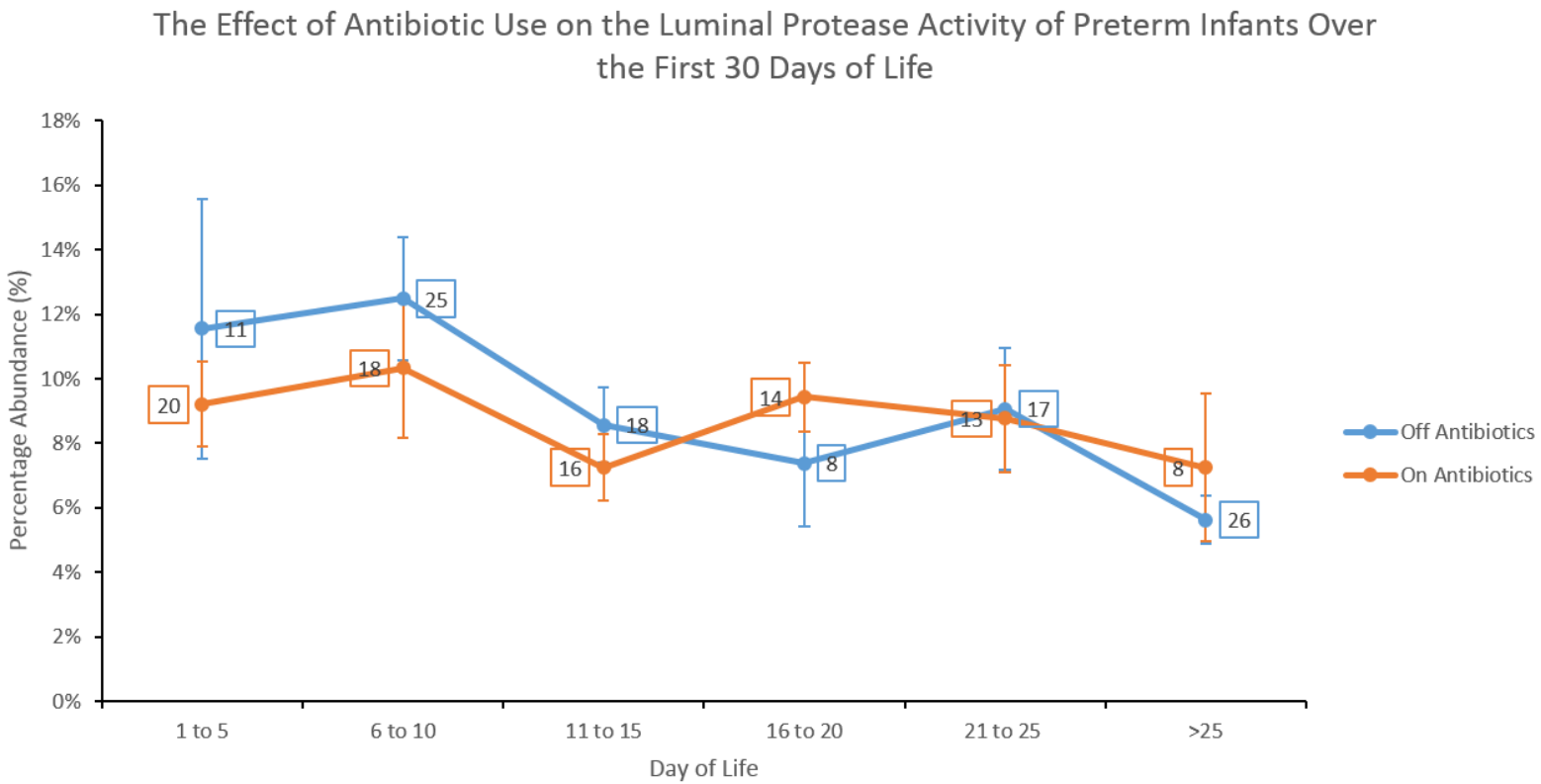

Figure 61. The Effects of Antibiotics on the Development of Protease Activity of Preterm Infant Stool. If a sample was taken whilst on a course of antibiotics and another sample was taken ± 3 days either side, the sample was removed. In total 10 samples were removed to leave a total of 200 samples. Secondly, if more than one sample form the same infant was given within the same day of life grouping e.g. days 6 to 10 of life, if one sample was on antibiotics and the other not, this sample was taken and the other removed. However, if the 2 were on or off antibiotics the mean of the protease activity of the 2 samples was taken. This resulted in a final total of 188 samples used during this analysis. The number of samples are included next to the mean protease activity for the day of life grouping and are outlined in the corresponding data series colour. Furthermore, the error bars are colour matched with the data series and are the standard error of the mean. The data point for each data series is the mean protease activity for each day of life group.

5.5 Discussion

In this investigation, it has been shown that antibiotics have a significant effect on the positivity of sequencing results during microbiome research in preterm infants. Furthermore, the administration of antibiotics clearly changes the development of the gut microbiome in these infants. However, the protease activity in preterm infant stool remained unaffected overall and over time by these drugs. Therefore, we propose the protease activity present to be primarily of host origin.

5.5.1 Patient Recruitment and Sample Demographics

Antibiotics were found to have a significant effect on the positivity of sequencing. Again this is not unexpected, as antibiotics have been shown to decrease the abundance of bacteria in the gut (Pérez-Cobas *et al.* 2012). This factor needs to be considered during future research, as the administration of antibiotics clearly reduces the efficacy of microbiome research in this field. Furthermore, it poses the question as to whether these drugs are skewing what would be considered normal microbiome development in preterm infants. On the other hand, antibiotics are critical in the routine care of these infants, that perhaps the antibiotic affected microbiome is more indicative of "normal" gut microbiome development in preterm infants.

As discussed previously, this was not the only explanation for the loss of samples during sequencing. There was a slight decrease in amount of stool provided by the negative samples compared to the positive samples, therefore a reduction in starting material is a plausible explanation for the reduction in sequencing positivity. Furthermore, the sample may have failed to amplify sufficiently during the Illumina sequencing process, resulting in data of insufficient quality. Finally, the majority of the samples that failed to sequence successfully were from the first week of life and will have a low biomass and a higher risk of sequence failure.

Factors that did not affect the positivity of sequencing results were, as follows. There was no change in average gestation between the two groups. This result shows that development is not a significant influence on gut microbiome colonisation. Furthermore, there was no change in the average maternal age. However, the birthweight of the infants who were positive for sequencing was slightly more than those negative for sequencing. This is in line with previously published research

showing that low birth weight infants harbour a less diverse, and therefore lower abundance of bacteria, microbiome compared to larger infants (LaTuga *et al.* 2011; Costello *et al.* 2013). Positivity of sequencing was unaffected by length of hospital stay, this indicates at a reduced effect of environmentally acquired organisms. However, hospital stay could be a proxy for age, as the longer an infant is in hospital, the older they become and gut microbiome colonisation increases. Finally, positive sequencing results was reduced in the infants who did not survive and who developed NEC. This could be explained two-fold. The more unwell infants, especially those with NEC, would find it difficult to pass stool, and therefore provide less samples. Secondly, it could have indicated that a threshold level of gut microbiome colonisation is required for normal or healthy development. This has been shown in gnotobiotic mice (Desbonnet *et al.* 2014).

In summary, the positivity of sequencing is influenced by numerous factors before and after birth. However, it can be concluded that the only significant factor to influence the success of sequencing is the use of antibiotics during the sampling period.

5.5.2 The Preterm Gut Microbiome as a Result of Antibiotic Administration

In this investigation it has been demonstrated that changes occur in the gut microbiome of infants exposed to antibiotics and those who do not. For instance, the numbers of *Bacteroides* are enriched in samples taken in the presence of antibiotics. This is an interesting finding as the *Bacteroides* genus is thought to contain beneficial organisms, as they were found in infants who did not develop NEC (Pammi *et al.* 2017). This was also seen in previously published data. These findings suggest that antibiotics are a good preventative measure against the development of NEC. Further research is needed to determine the method in which these organisms are protected from antibiotics, such as antibiotic resistance or the replacement of other organisms targeted by the antibiotics.

It was discovered during this study that antibiotics had no significant effect on the diversity of the gut microbiome in preterm infants, as seen previously (Yassour *et al.* 2016). Firstly, this shows that antibiotics are not affecting the predominance of organisms within this community. More specifically, they are either having no effect or are causing a shift from one dominant organism to another. Neither of which are beneficial to the health of the infant. As a result, the administration of antibiotics does

not promote a healthy diverse gut microbiome. The outcomes of which need further study.

However, antibiotics have been shown to have no effect on the microbiome. A very recent study concluded that the differences in the preterm microbiome between individuals was not as a result of antibiotic exposure (Dahl *et al.* 2018). This suggests that it is not antibiotics alone causing changes in the gut microbiome. Furthermore, it was concluded that weight gain was not significantly different between infants exposed to antibiotics during the first 6 months of life and those who did not (Gerber *et al.* 2016). This demonstrates that antibiotics do not impact infant health. Previously it was shown that empirical antibiotic use increases the risk of neonatal morbidity. However, it was found that the implementation of an Automatic Stop Order (ASO) on the empirical use of antibiotics, resulted in no change on observed outcomes, such as mortality and morbidity (Tolia *et al.* 2017). This conflict of results shows that the effect of antibiotic on the gut microbiome needs further investigation in order to fully elucidate the effects of these commonplace drugs.

Finally, there was no difference in protease activity observed between the antibiotic present or absent groups. This is an interesting finding as it indicates at the origin of the protease activity demonstrated. Therefore, if antibiotics have a significant effect on bacterial load but not on protease activity, this leads to the conclusion that the origin of these proteases to be host rather than bacterial.

5.5.3 The Effect of Antibiotics on the Gut Microbiome of Samples taken from Recruitment Sites

The results of this study provide a unique finding in the effect of antibiotics on the probiotic supplemented gut microbiome of preterm infants. It was clearly shown in Figure 6 that antibiotic administration dramatically reduces the levels of *Bifidobacterium*, a known component of the probiotic given during the sampling period. This was also demonstrated in previous research (Tanaka *et al.* 2009).

There is currently no consensus as to the efficacy of probiotics in preterm infant care (Costeloe *et al.* 2016; Hays *et al.* 2016; Olsen *et al.* 2016). Therefore, the results of this investigation demonstrate that organisms ingested as part of a probiotic do become integral members of the microbiome community. Moreover, they are affected by the

administration of antibiotics. However, further research is needed to determine if this is detrimental to the health of these infants. Furthermore, further evidence is needed to determine if probiotics are a viable therapeutic during antibiotic treatment.

5.5.4 The Effect of Antibiotics on the Development of the Gut Microbiome in Preterm Infants

In addition to the overall effects of antibiotics on the gut microbiome of preterm infants, these drugs can have a dramatic effect over time. At days 6 to 10 of life the presence of antibiotics dramatically decreases the total bacterial load in the gut of preterm infants. It is currently unknown as to the effects of this difference. However, this peak in bacterial load has been linked to infective process within the lung of preterm infants (Unpublished research). Therefore, the same process could be according in the gut, but needs a lot more research.

There is little evidence investigating the effects of antibiotic administration over time in the preterm infant population (Tanaka *et al.* 2009; Fouhy *et al.* 2012). Therefore, it makes interpreting the results of this study very difficult. In contrast to previously published research, this investigation found a decrease of *Proteobacteria* in the samples taken in the presence of antibiotics. Moreover, the presence of *Proteobacteria* as part of the community remained more stable in the absence of antibiotics. Furthermore, it was interesting to see that the *Bacteroidetes* phyla was only present in the community of infants exposed to antibiotics until 2 weeks of age. Therefore, the presence of antibiotics does reduce the numbers of *Bacteroidetes*, but it is a long process. These results add further evidence that the empirical use of antibiotics can be detrimental to the health of preterm infants. In summary, the development of the microbiome in the presence of antibiotics is distinctly different from the development in the absence of antibiotics. However, more research needs to be conducted in order to determine the acute and long-term health outcomes.

Finally, the effect of antibiotics on the protease activity of the preterm infant gut was examined and determined that not significant differences in protease activity over time occur as a result of antibiotics. Neither data set follows the same pattern in the protease activity and bacterial load graphs. This adds further evidence to the conclusions discussed above.

5.6 Conclusion

This study has provided several interesting findings to the area of preterm gut microbiome research. Firstly, antibiotics during the first 20 days of life has a significant and dramatic effect on the developing gut microbiome. Secondly, there is a significant effect of antibiotics on the positivity of sequencing results from preterm infants. This information will need to be considered when conducting further investigations into preterm infants. Furthermore, antibiotics have a significant effect on the bacterial load from the preterm infant gut. Also, there are noticeable changes in the gut microbiome community as a result of antibiotic use. However, antibiotic use does not have a significant effect on clinical parameters or protease activity. In respect to the development of the gut microbiome there are dramatic changes in bacterial load over the first 30 days of life, this coincides with community changes.

Chapter 6

The Gut Lung Axis of Preterm Infants.

Chapter 6. The Gut Lung Axis of Preterm Infants.

6.1 Introduction

The concept of the gut-lung axis was first proposed by Pugin and Chevrolet in 1991. Their original description was based upon the pathology of sepsis, whereby the translocation of bacteria and their endotoxins into the bloodstream would lead to a increase in the secretion of cytokines such as, TNF-α, IL-1β and IL-6. This would then result in adult respiratory distress syndrome and multiple system failure (Pugin and Chevrolet 1991). Later this definition was added to include the role of the immune system, and this is the definition is more appropriate as the translocation of bacteria between the gut and the lungs has yet to proven. Chen *et al* in 2011 showed that when the microbial community of the gut became depleted, the lung infection in the mice infected with *E. coli* became worse (Chen *et al.* 2011a).

As previously described the gut lung axis contains the following components: the lung and gut environment, the bacterial community of these organs, the immune system and transient components (Budden *et al.* 2017). Moreover, the interactions between these components has been linked to disease (Roussos *et al.* 2003; Keely *et al.* 2012; Rutten *et al.* 2014). In contrast, a study by Scuijt *et al* in 2017 has shown the gut microbial community to protective against lung disease, more specifically pneumococcal pneumonia (Schuijt *et al.* 2016). Therefore, the role of the gut-lung axis in development has yet to be investigated. More specifically, the gut lung axis of preterm infants has not been investigated or detailed in previous research. Therefore, it is important to investigate the links in preterm infants.

The lung microbiome of preterm infants is less understood than the gut but has been shown to be pivotal in the development of disease. It has previously been shown that the presence of predominant bacteria is associated with the development of chronic lung disease of prematurity (CLD), also called bronchopulmonary dysplasia (BPD) (Beeton *et al.* 2011). In addition, the progression of CLD is thought to occur as a result of inflammation as a result of an increased bacterial load or pathogenic bacteria (Davies *et al.* 2010; Beeton *et al.* 2011). The inflammatory system, both pulmonary and gastrointestinal, has been shown as the crucial link between the gut and the lungs (Marsland *et al.* 2015).

Currently the data from this project and previous research reports that the gut microbiome of preterm infants progresses from a *Firmicute* dominated environment to a *Proteobacteria* one. However, the research into the colonisation of the lower airways in preterm infants is less clear. A recent systematic review examines the results of several studies and reveals that colonisation of the lower airways can mirror that of the gut, from a *Firmicute* to *Proteobacteria* dominated. However, it can also be the reverse (Pammi *et al.* 2019). The data from the lower and upper airways used in this study, provided by Dr David Gallacher, shows that the lungs of preterm infants progresses from a *Firmicute* dominated community to a *Proteobacteria* one.

Proteases are thought to contribute to the pathology of disease via the breakdown of tight junctions in the gut epithelium. The breakdown in the tight junctions allows for the gut bacteria to pass from the intestinal lumen into the bloodstream, leading to the translocation of the bacteria to the lungs. The results of Chapter 4, showed that the overall protease activity of preterm infant stool was low, however the effect of the proteases on the gut epithelium was not tested. As yet inconclusive is the translocation of bacteria from one organ to another. This is because research into the gut-lung axis focuses more on the effect of the gut microbial community on the development of lung disease, via immune system crosstalk, rather than physical translocation of bacteria from one organ to another (Dang and Marsland 2019). Therefore, as part of this project it was important to provide more evidence to support or disprove the current hypothesis in research that there is no significant movement of bacteria form one location to the other.

6.2 Aims and Hypothesis

The overall aim of this chapter was to investigate the possible relationships of bacteria from different organs of the body. The two sites investigated were the lung, specifically NPAs for the upper airway, and TAF and BAL fluid to sample the lower airways. Secondly, the gut was sampled using stool. In order to study this hypothesis, the following aims were aims were developed:

1. To examine the change in bacterial load over the first 30 days of life from NPA, TAF, BAL and stool samples. This is to determine if colonisation of the different sites occurs at the same time and potentially to the same degree.
2. To investigate the microbial communities at the four sites to determine if similar organisms were colonising both sites. Moreover, community similarities at the four sites potentially indicates at the translocation of bacteria from one site to another.
3. Lastly, to determine if inflammatory programming is occurring comparisons between gut protease activity and lung inflammation will be analysed.

6.3 Materials and Methods

6.3.1 Data Acquisition

Microbiome data on nasopharyngeal aspirates (NPA), tracheal aspirate fluids (TAF) and bronchoalveolar lavages (BAL) from the same cohort of infants recruited to this project were obtained, with permission, from Dr David Gallacher. Data was provided in the form of .txt files. Meta data for all samples including bacterial load was also provided.

6.3.2 Data Analysis

Bacterial load was compared in Excel. Lung bacterial load was determined in copies per ml, whereas stool bacterial load was expressed in copies per mg of stool. Phylum and genus taxonomic analysis was also executed in Excel. Non-Metric Multidimensional Scaling (NMDS) was conducted in R.

The same method was used here as before where only reads greater than 1000 were used during this analysis,

6.4 Results

6.4.1 A Comparison of Total Bacterial Load from the Gut and Lung of Preterm Infants.

The aim of this analysis was to determine if the gut and lungs of preterm infants were being colonised by bacteria at a similar rate. This was achieved by comparing the average bacterial load at several time points during the first 30 days of life. Figure 62 below shows the results of this analysis.

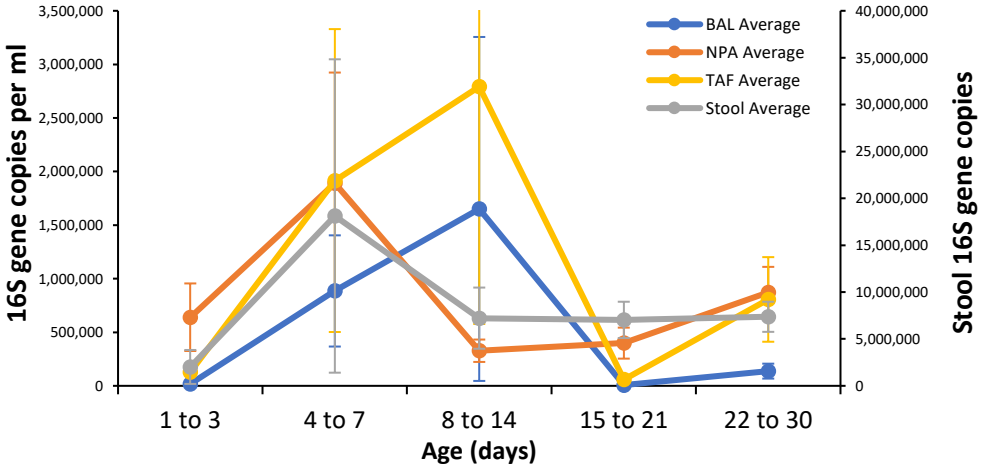

Figure 62. The Total Bacterial Load in the Gut and Lungs of Preterm Infants. The error bars are the standard error of the mean. The number of samples included in this analysis are as follows: BAL (36), NPA (124), Stool (151), and TAF (47). All lung samples, TAF, NPA and BAL, are shown on the primary Y axis, and the stool samples are presented on the secondary Y axis.

Firstly, it must be noted that the copies of the 16S rRNA gene in the stool are a factor of 10 greater than that of the lungs, so even though the TAF samples show the highest peak, this amount is lower than that of the highest stool samples. There is a clear pattern in bacterial load at the different sites. Firstly, the bacterial loads from NPA and stool samples peak at the sample time point, days 4 to 7 of life. Secondly, BAL and TAF samples peak at the same time, at days 8 to 14 of life. Furthermore, samples from all sites substantially decrease at the time point after the peak.

6.4.2 A Comparison of the Microbial Communities from the Gut and Lung of Preterm Infants.

The aim of this analysis was to compare the microbial communities at the four sites from the lung and gut to determine if there were any similarities. Moreover, the communities were compared during the first 30 days of life, this allowed for

comparisons in the development of the microbial community. The results of this analysis at the phylum level are shown in Figure 63. From observing the graphs in Figure 63, the microbial communities of NPA and stool samples most closely resemble one another, but are distinctly different. For instance, over time the abundance of Proteobacteria increased during the first 30 days of life. Furthermore, Actinobacteria are only present in these samples. The abundance of Tenericutes is the least in the NPA samples and not present in the stool samples.

In order to analyse the comparisons between the microbial communities at the four sites sampled from preterm infants, an NMDS was conducted. The results of this analysis can be found in Figure 64. Firstly, it can be seen that the TAF and BAL samples cluster the most closely together. Moreover, the NPA samples did cluster closely to the TAF and NPA samples but not as closely. Furthermore, the stool samples clustered separately from the lung samples but clustered to encompass all of the lung samples. Another interesting finding from this analysis was that the samples from each site did cluster together, showing that samples from each site are most similar to one another.

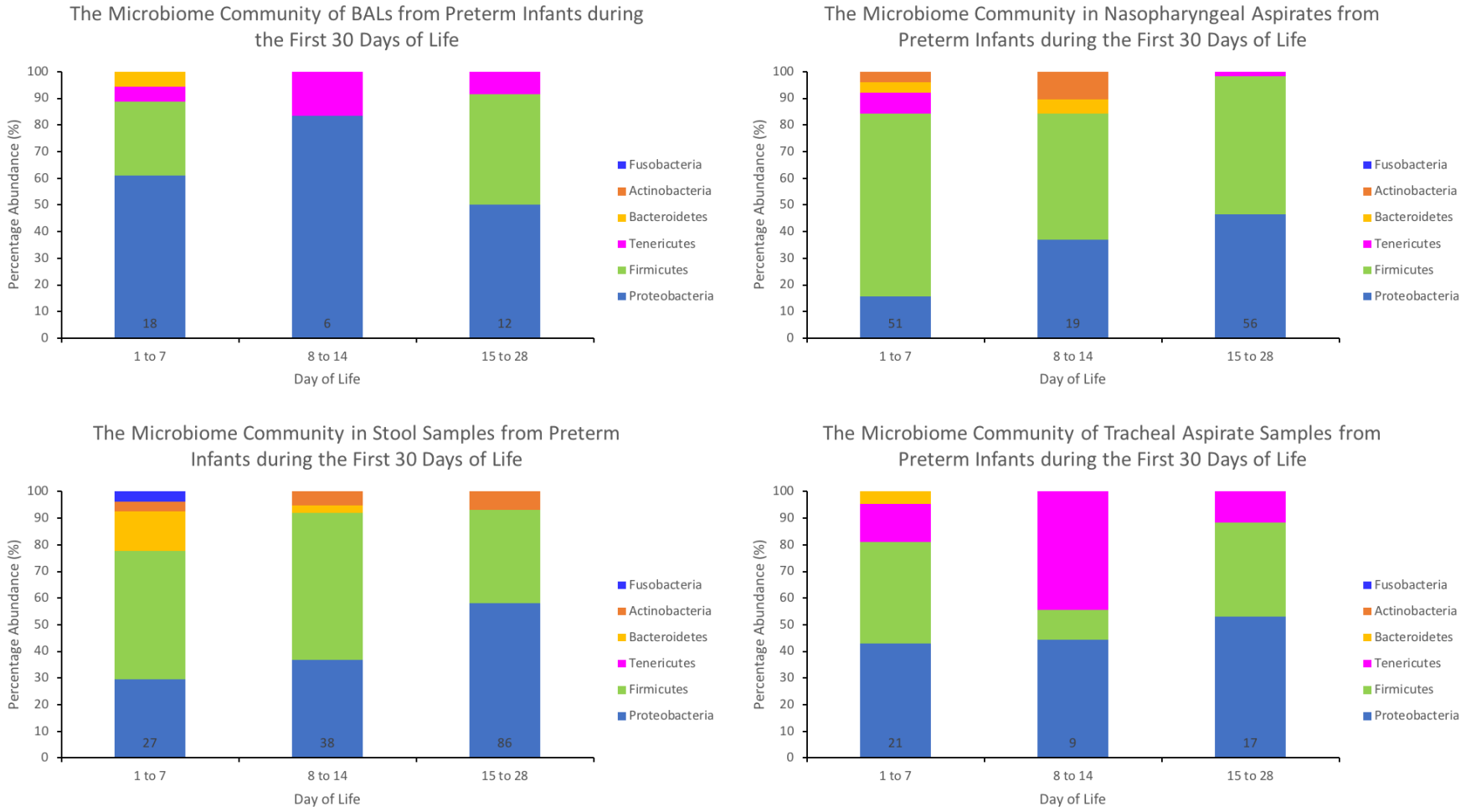

Figure 63. A Comparison of the Microbial Communities at the Gut and Lungs of Preterm Infants. The number of samples used in each day of life grouping is given at the bottom of each bar. The bars are the average percentage abundance of each phylum for that days of life grouping.

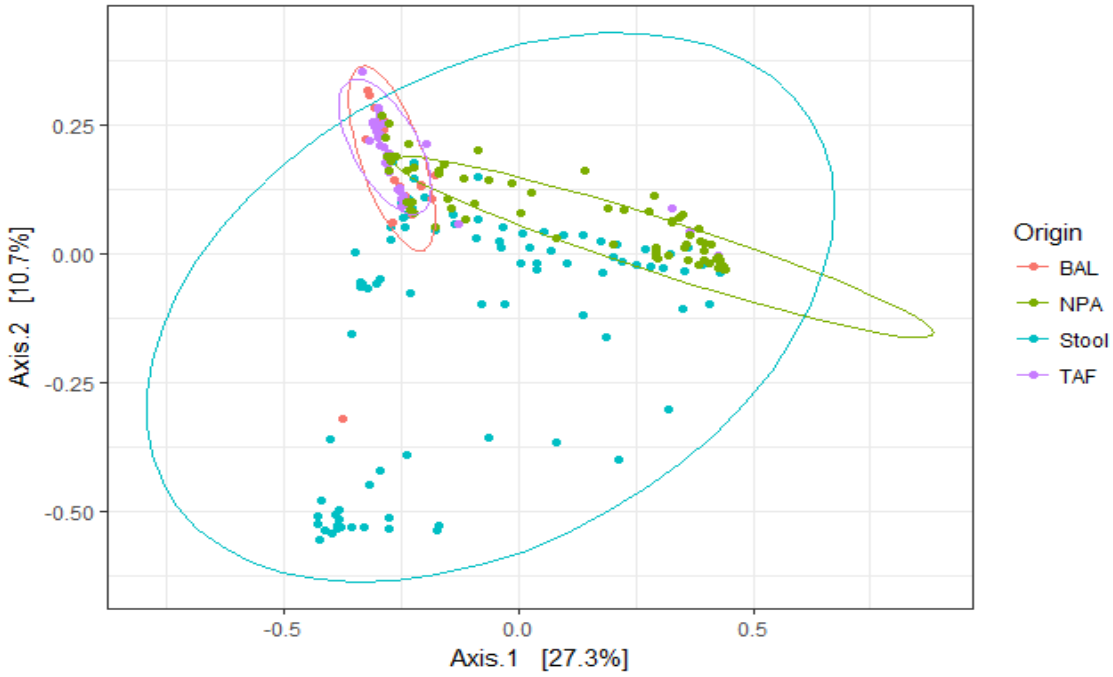

Figure 64. A Comparison of the Microbial Communities from Four Sites from the Gut and Lungs of Preterm Infants. This is a NMDS plot using the Bray-Curtis dissimilarity index to determine the relationships in bacterial communities between the samples. Samples are coloured according to site of origin. The ellipses were also calculated in R. The percentages on the X and Y axis when totalled, 38%, totals the amount of dissimilarity accounted for in this figure.

6.4.3 The Association between Stool Protease Activity and Lung Inflammation.

The aim of this analysis was to determine if the hypothesis of previous research, that the exposure to bacteria in the gut programmes the immune system to respond appropriately to bacterial colonisation in other organs of the body such as the lungs, was possible. In order to achieve this the levels of the cytokines IL-6 and IL-8 from the lungs were compared, over time, to the percentage of protease activity in the stool. The results of this analysis found that there was no association between protease activity in the stool and levels of IL-6 and IL-8 in BAL and TAF samples.

6.5 Discussion

6.5.1 The Comparison between the Bacterial Loads Sampled from the Gut and the Lungs of Preterm Infants.

The results of this analysis showed that overall, the bacterial load in the lungs is much lower than in the gut. In this population of infants, the predominant explanation for this is that all infants recruited to this study were ventilated, as shown by previous research (Armaforte *et al.* 2010; Kelly *et al.* 2016). The process of being ventilated means that the area is kept sterile and inhibits the growth of bacteria and therefore the colonization. In contrast, the gut is still open to the outside environment and therefore has the potential to be colonized by bacteria. There are potentially other contributing factors such as the gut being a more hospitable environment, such as the increased availability of food. In addition, during the ingestion of food, this introduces a substantial number of bacteria into the digestive system.

Another finding from this analysis is that the bacterial load pattern during the first 30 days of life is very similar in the TAF/BAL samples and the NPA/stool samples. This is a really interesting finding as it shows that the development of the lower airways is different to that of the upper airways. Moreover, it indicates that the colonisation of the upper airways is occurring at a similar rate to the gut. The TAF and BAL samples peaked at days 8 to 14 of life whereas the stool and NPA samples peaked at days 4 to 7 of life. This is the key result that separates them and potentially indicates at separate developmental process.

The lower bacterial load in the lungs and the similarity in bacterial load between the lower airways and the upper airways/gut indicates at a consequence of ventilation. In other words, the ventilation is potentially isolating the lower airways, the trachea and bronchus, from the outside environment. Whereas, the upper airways, the nose, and the gut are exposed to the environment either directly or indirectly. Therefore, it was important to investigate similarities in the bacterial community.

6.5.2 Comments on the Bacterial Communities of the Gut and Lungs in Preterm Infants, and their Similarities.

The aim of this analysis was to determine if the bacterial communities at the gut and lungs of preterm infants were similar or dissimilar. This was achieved by comparing the communities at the phylum taxonomic level and performing an NMDS. The results showed the NPA and BAL sites to one another and were very different from the NPA and stool samples. These results were confirmed by the NMDS as the TAF and BAL samples clustered closely together, with the NPA samples separate but close by. Then the stool samples also clustered separately but encompassed the samples taken from the lungs.

There are several possibilities that could explain the separation of the bacterial community from the NPA samples from the rest of the airways. However, there is one explanation that is most likely. The infants recruited to this study were ventilated due to respiratory distress after birth. As a result, they were intubated with a tube passing from the mouth into the lungs to deliver oxygen. This separated the lungs from the outside environment. In contrast, the nasopharynx was exposed to the outside environment and the gut received transient components from the outside environment. Furthermore, the intubation tube would be kept sterile and changed periodically, therefore further isolating the community there. As a result, the NPA and stool communities have developed with more similarities to one another rather than the NPA samples developing more closely to the rest of the airways. However, they ultimately remain very distinct from one another.

The translocation of bacteria from the gut to the lungs via the breakdown of the epithelial barrier by proteases or excessive inflammation, has been shown to not occur. Firstly, due to the lack of concordance in the microbial communities between the lower airways and the gut. Secondly, due to the lack of protease activity in the guts or preterm infants. Therefore, the minor similarities between the NPAs and stool samples is not a result of bacterial translocation.

However, the Actinobacteria phylum is only present in the NPA and stool samples and not the lower airways, but the translocation of bacteria is not happening in these infants from the results of this project. Therefore, the origin of these bacteria in the NPA

samples is unknown. The origin of the Actinobacteria in the gut is believed to be a result of the administration of probiotics, but this reasoning does not explain their presence in the NPA samples. However, it has been discussed that the NPA samples are different from the TAF and BAL samples because of the exposure of the area to the outside environment. Therefore, using this hypothesis the presence of Actinobacteria in the NPA samples is from the outside environment. Unfortunately, samples of the NICU were not taken as part of this study, therefore this theory cannot be proven. Moreover, this is something to consider for future studies, in order to determine the origin of the bacteria colonising preterm infants. There have been studies showing the correlation between bacteria present in the NICU and the microbial communities of preterm infants, and there is some correlation between these communities (Hewitt *et al.* 2013). Therefore, it is possible that the Actinobacteria present in the NPA samples is from the NICU environment. Furthermore, it is unknown how much of an effect possible NICU Actinobacteria are present in the guts of these infants.

Similar to Actinobacteria only being present in the NPA and stool samples, the Tenericutes phylum is only present in the TAF and BAL samples. This result, however, is easier to explain due to the isolation of the lower airways. Moreover, it further solidifies the lack of bacterial translocation due to the compromise of epithelial barrier integrity.

The NMDS produced an interesting result in showing that the samples from the airways, TAF, NPA and BAL, did not cluster together, and the stool samples clustered away from this group. Furthermore, the TAF and BAL samples clustered the closest together with the NPA samples close by but definitely separate. For reasons discussed previously the TAF and BAL samples cluster so closely together due to their isolation as a result of intubation. However, it is interesting to see that even due to the differences between the NPA and TAF, BAL samples that when analysing using a dissimilarity index the communities are similar and indicate at a shared origin. In addition, the stool samples appear to encompass or be shared with the lung samples rather than being completely separate. This is an unusual result. Firstly, it could demonstrate the low diversity present in the NICU due to the sterile conditions, therefore there is not a plethora of bacteria present to colonise the infant. Secondly, these infants are often

unable to feed enterally and therefore would not be ingesting bacteria from food, resulting in the communities being less diverse.

6.5.3 The Association between Stool Protease Activity and Inflammation.

The results of this study showed that due to the lack of inflammation and protease activity in the stool, no associations between the inflammation seen in the lung could be made. Previous research has suggested that colonisation of the gut prepares or matures the inflammatory system to respond appropriately to microbial colonisation in other areas of the body, such as the lungs. Therefore, it is reasonable to hypothesise that the high levels of inflammation seen in the lungs is a result of the lack of inflammatory programming in the gut, due to the low protease activity and undetectable IL-6 and IL-8 in the stool. However, more research is needed in order to determine if this is the case. In future research, it would be interesting to expose cultured gut epithelial cells, such as HT-29 cells, to the faecal water to investigate if the stool has the potential to illicit an inflammatory response. Moreover, blood samples could also be tested to determine if the inflammatory response is systemic.

Another key limitation of the study was the lack of samples from the same day from both the gut and the lungs. As a result, it was not possible to investigate associations between the gut microbial community and lung inflammation. Therefore, the current research community hypothesis on the gut lung axis could not be fully investigated as part of this study. However, as part of the study the lack of bacterial translocation was could be conclusively investigated.

6.6 Conclusion

In summary, the results of this study have shown that the gut and lungs of preterm infants do not have a similar microbial colonisation and that differences in this colonisation process could be a result of external influences, such as intubation. Furthermore, no association between stool protease activity and lung inflammation was found, this was an interesting result as it shows that gut inflammatory responses are not indicative of lung inflammatory responses. Conversely, the lack of immune response in the gut could lead to an excessive inflammatory response in the lungs, as the gut lung

axis hypothesis suggests that gut microbial colonisation “primes” the immune system at other sites to respond appropriately to bacterial colonisation (Budden *et al.* 2017; Dang and Marsland 2019). Ultimately, a lot more research is needed to fully elucidate the concordance in development of the microbial communities in preterm infants. Moreover, investigations into the role of these communities in the development of disease is needed also. This preliminary study points towards a vital role for the development of multiple microbial communities in the preterm infant. As this is the first study into the role of the gut lung axis in the development of preterm infants, there is no recent literature in which to comment on and place this study, regarding preterm infants. However, in conclusion the results of this study confirm the findings of previous studies in that there is no evidence of the translocation of bacteria from the gut and the lungs. In addition, due to the lack of concordant samples of the gut microbiome, it is unfortunate that comments on the influence of the gut microbiome on lung inflammation cannot be investigated.

Chapter 7

General Discussions and Conclusion

Chapter 7. General Discussion and Conclusions

A comprehensive study of the preterm gut microbiome and the protease potential of this bacterial community was undertaken. Overall, this research has further identified the developmental process of the preterm gut microbiome and, for the first time, examined the activity and types of proteases present in the guts of these infants. Moreover, it has investigated the preterm gut-lung axis for the first time and provided interesting findings. As a result, it has validated the presence of proteases in the gut of preterm infants, thus opening a door to further research within this area.

In this chapter, the central discussion points and conclusions of each chapter will be reiterated and how the aims of this project, Chapter 1, were met. Moreover, the implications of this study to the wider scientific community and future gut microbiome research in preterm infants will be discussed. In addition, the key areas for further research will be highlighted and the potential this has for the health of the preterm population.

7.1 The Gut Microbiome of the Preterm Infant

Previous research has shown the gut microbiome of preterm infants' progress from a *Firmicute* dominated community to a *Proteobacteria* one, and is affected by numerous factors after birth, such as feeding. As a result, the overall aim of this chapter was to provide further evidence on the developmental process of the gut bacterial community in this cohort. Moreover, clinical data collected at the time of sampling provided the opportunity to investigate the effect of gender and other factors, not previously reported upon. In addition, the data provided by this study allowed for the analysis of the microbiome in relation to the development of NEC.

The results of this study showed that this cohort of preterm infants is akin to that of other communities studied, as they progressed from a *Firmicute* dominated gut community to a *Proteobacteria* one. When these results are taken into consideration with the results of previous research it solidifies the hypothesis that the gut microbiome progresses towards a *Proteobacteria* dominated community from a *Firmicute* one during the first month of life in preterm infants. As a result, it can be said, with confidence, the development of the preterm gut microbiome differs significantly from that of full-term infants. Moreover, the individualistic nature of the preterm gut microbiome was further validated in this study. Therefore, the confirmation of these findings enables the next steps in this research to take place, such as the investigations into the manipulation of gut bacterial community. For example, maintaining the *Firmicutes* dominance and preventing the takeover by *Proteobacteria*, perhaps this could be achieved by probiotics, as later research has shown. Moreover, this is a possibility as the individual nature of the preterm gut microbiome indicates at a community that is easy to manipulate. In addition, the manipulation of the gut microbiome could lead to an increase in diversity, as this has been shown to be healthier, and the results of this study found the diversity of the preterm gut to be low and unchanging over the first 30 days of life.

The results of the analysis into the effect of clinical factors on the development of the preterm bacterial gut community showed that the community as a whole was not affected by the specific factors. However, significant shifts in the bacterial community occurred were shown to occur as a result of certain conditions. Firstly, gender caused a

significant increase in *Proteobacteria* in male infants. Reasons for this were discussed in Chapter 3, but a conclusion could not be drawn. Therefore, this is an area of research that needs further investigation as it implies that males are more at risk of developing a *Proteobacteria* dominated microbiome and therefore future complications. The effect of mode of delivery on the development of the bacterial community in the gut of preterm infants is another area of research that needs to be expanded upon, as the results from this study contracted the findings of previous studies.

The finding that the significant difference between the sampling sites used in this study was a result of probiotics, were two-fold. Firstly, that probiotics have a significant effect on the gut bacterial community. This is an important finding for future research as it shows that the gut microbiome of preterm infants can be manipulated. Secondly, it shows that large-scale studies recruiting from several NICU sites are a possibility as the data can be pooled and interpreted as a collective cohort.

Finally, in relation to future studies into the therapeutic manipulation of the preterm gut microbiome, the results of this study showed that diet could be another method. However, the results of this study in relation to the effect of diet on the bacterial community were limited. This was a result of the lack of infants receiving only formula feeds or donor breast milk. There needs to be conducted with a larger cohort and an equal number of infants receiving the variety of feeding methods.

Similar to the effect of mode of delivery, the results of this study contracted that of previous finding. The bacteria *Klebsiella* and *Clostridium* were not associated with the development of NEC in this study cohort, whereas they have been shown to be present in the development of this disease. This is an interesting result, but should be taken with a pinch of salt, as the study was not designed around recruiting a high number of infants who developed this disease. Therefore, this could be why the results of this study contradict that of previous research. On the other hand, this study did find the presence of *Bacteroidetes* was significantly increased in infants who did not develop NEC. These results are similar to previous findings. Therefore, the potential “protective” effect of the *Bacteroidetes* genus needs to be investigated further as it has the potential to save lives.

In summary, the results of this study demonstrate the need further research in this area. The data has shown that the gut microbiome is a key part during preterm infant development and can be manipulated by a number of factors. Therefore, the results of this study have been important in laying the foundation for future studies into the potential for therapeutic manipulation of the preterm gut microbiome. Moreover, there have been unique findings. For example, two cohorts of infants from separate NICUs could be combined into a single dataset and that the only significant difference was as a result of probiotics.

7.2 The Effect of Antibiotics on the Development of the Preterm Gut Microbiome.

The aim of this research was to determine the effect of antibiotic administration on the development of the preterm gut microbiome. Furthermore, this included investigations into the effect of antibiotics on the protease activity of the preterm infant gut. Previous research showed that antibiotic administration increased rates of sepsis and NEC but did not detail the effect it was having on the gut bacterial community. Moreover, the role of proteases in the preterm infant gut has yet to be commented on at all.

Firstly, antibiotics were found to have a significant effect on the positivity of sequencing. Antibiotics are made to target bacteria and eliminate them. However, for them to have such a significant effect on the bacterial community in the gut, is interesting. This was the first indication that the findings of previous studies, the increase in sepsis and NEC, could be related to the gut bacterial community.

The only organisms to change as a result of antibiotic administration were the *Bacteroidetes*, whose presence increased in samples taken on antibiotics. Taking this information into consideration with the findings from the previous chapter, that *Bacteroidetes* are more present in infants who do not develop NEC, it indicates that antibiotics could promote the growth of *Bacteroidetes*. However, this contradicts previous findings where antibiotics increase the development of NEC and sepsis. It seems odd that antibiotics would promote the growth of a bacteria that could be "protective" against NEC. As a result, further research is needed to fully elucidate the potential "protective" role of *Bacteroidetes*, and if there are specific types of antibiotics or dosage that is linked to the development of NEC and sepsis. This area of research could lead to a significant reduction in these preterm diseases.

One of the unique findings of this research was that there was no reduction in protease activity as a result of antibiotic use. Due to the limited previous research and methodology used during this study, the reasons for this can only be commented and not concluded upon. This result suggests that the proteases detected are of host origin. In order to fully determine the origin of these proteases, different methodologies need to be applied. For example, qPCR for transcription of human and bacterial proteases could

be employed. However, further studies into determining the exact types of proteases present would be needed first, in order to identify the genes of interest. This could be achieved by fractioning stool samples to contain proteases of different weight and then sending them for mass spectrometry to identify the molecular structure, this then leads to precise identification.

Another unique finding was the effect of antibiotics on the samples taken from infants at the different neonatal units. The antibiotics clearly reduced the presence of *Bifidobacterium* in the guts of infants from UHW. Therefore, antibiotics have the potential to reduce the efficacy of probiotics given at the same time of antibiotics. This an important finding to be taken forward into future research in this area. Or even a study dedicated to investigating the effect of antibiotics on the efficacy of probiotics would be useful to conduct. As the efficacy of probiotics within this community is not fully known and research is contradictory.

During the first 30 days of life antibiotics caused a dramatic decrease in bacterial load at days 6 to 10 of life, compared to infants not receiving antibiotics. This continuation of antibiotic therapy past the first few days life, results in the absence of a peak in load at the same time grouping. It is unknown what the effect this could be having on the development of the infant. However, considering previous research that suggests immune programming in the gut prevents inflammatory processes occurring at other sites within the body, such as the lungs. It could be suggested that this absence of bacteria could lead to increased inflammation in the lungs. On the other hand, it could also be protective for the infant, as minimizing exposure to potential pathogenic bacteria could reduce the risk of developing NEC.

In summary, antibiotics play a key role during the development of the gut microbiome in preterm infants. However, due to ethical issues of withdrawing antibiotic therapy, the full effects of these drugs will be very hard to study in this vulnerable population group. However, a more large-scale study, that could recruit enough infants to be able to examine the effects of a specific type of antibiotic or length of treatment, would provide some much-needed data on the role these drugs play. This study has provided preliminary evidence as to the effect they are having on the gut bacterial community, but more research is needed.

7.3 The Role of Proteases and Inflammation in the Development of the Preterm Gut Microbiome.

The aim of this study was to, for the first time, determine the total protease activity of preterm infant stool. Then investigate the content of this protease activity by finding out the types of proteases present and the origin of the proteases. In addition, the inflammatory content, levels of IL-6 and IL-8, was investigated and was used to compare to the protease activity. In relation to disease, it was necessary to determine of the preterm disease NEC, could be caused by an excess of protease activity, similar to adult gastrointestinal diseases such as IBD.

Overall the results of this study showed that there was no significant protease activity in the stool of preterm infants. Therefore, this suggests that preterm diseases such as NEC do not have a contributory effect from proteases. This was the first study into the total protease activity of preterm infant stool and the change in activity during the first 30 days of life. A disadvantage of this study was that there were no full-term samples to compare these results to. Therefore, no conclusions can be made as to whether this is "normal" for preterm infants, or even if the total protease activity of full-term infant stool is comparable. For future studies a full-term cohort needs to be recruited alongside preterm infants to investigate the role of proteases in full-term infants. Moreover, a further study with a different cohort of preterm infants will either confirm or disprove the findings of this study. This needs to be conducted to confirm the lack of protease activity. However, explanations for this reduction in protease activity also needs to be investigated because the lack of protease activity could be detrimental also. Ultimately, this research has generated a lot of questions and further avenues for future research, which is why it has been a successful study.

A further uncertainty of the results of this study is the origin of the proteases detected in the stool. The association with bacterial load suggests they are of bacterial origin but the lack of change in response to antibiotics suggests they are host derived. I have commented on possible future experiments to determine this and it should be investigated.

Another result of this study showed that there was no association between protease activity and the development of NEC. As mentioned previously, the recruitment of this study was not focused around recruiting infants who develop NEC and therefore the numbers used during analysis were not large enough to produce significance. Moreover, this is the first study in this area and further studies are needed to confirm or disprove these findings. For example, one of the biggest difficulties with obtaining samples from infants who develop NEC is that a symptom of the disease is the inability to pass stool, therefore making it a difficult population to study.

Interestingly, the results of this study showed that several clinical factors did not have effect on the total protease activity. This is an unusual finding as there are significant changes to the bacterial community, so the factors are having an effect in the preterm infant but not on the protease activity. Again, this is further evidence towards a host origin of the proteases detected.

The experiments used to determine the types of proteases present in the stool of preterm infants were limited and should be improved for future studies. For instance, all samples should be subjected to the full suite of protease inhibitors to determine all types of proteases present. Moreover, because the results could not be more specific than just identifying the family of proteases, it again was very difficult to identify the origin or comment on the role these proteases are having in the gut microbiome of preterm infants. The experiments used in this thesis were to identify if analysis of protease activity and type were viable in preterm infants. Therefore, this project was successful in showing that this avenue of research is viable and has the potential for providing a huge amount of information on the development of the gut microbiome in preterm infants.

Protease activity or family could not be linked to a group of organisms. Once again providing evidence towards to the host origin for these proteins. Ultimately, this shows that there are complex pathways and interactions occurring in the gut microbiome of preterm infants. Even those these infants may be preterm.

Finally, no detectable levels of IL-6 or IL-8 could be found in the stool of preterm infants. This was an unexpected finding and produced a lot of discussion, as it shows that there is little to none localised inflammatory signalling occurring in the gut of

preterm infants. Moreover, it suggests that inflammatory signals are being produced and taken away from the gut.

In summary, the protease and inflammatory content of the preterm gut microbiome was low and undetectable, respectively. This was an unexpected finding and shows the gut of preterm infants to be less complex than adults. In other words, the gut of preterm infants is underdeveloped in relation to total protease and inflammatory activity. This has the potential to explain the development of disease in these infants. However, a lot more research is needed to determine if this is the case and, as said previously, a cohort of full-term infants is needed to determine if the development of the gut microbiome in preterm infants is significantly different to that of full-term infants.

7.4 The Role of the Gut Lung Axis in the Development of Preterm Infants.

The aim of this chapter was to investigate the similarities in the microbial community between the gut and lungs of preterm infants. Moreover, due to previous research indicating at a role of inflammatory programming occurring in the gut, links between protease activity and lung inflammation were analysed.

Overall, the results of this analysis showed that the stool and NPA samples showed the most similarity to one another, but were distinctly different. Moreover, the TAF and BAL samples showed the most difference to the NPA and stool samples. This was an interesting finding and demonstrated the effect of intubation on the development of the lung and gut microbiome in preterm infants. Furthermore, these results confirmed the low protease activity observed in previous chapters. The lack of similarity between the lower airways and the gut demonstrated the integrity of the gut epithelial barrier in the infants. The strong epithelial barrier prevents the leakage of bacteria from gut, into the bloodstream, and then potentially to the lungs. Moreover, the undetectable levels of IL-6 and IL-8 in the stool and the high levels observed in the lungs provides evidence to the hypothesis of the gut being the source of immune programming for the other mucosal surfaces. Actually, this analysis was key in bringing together several observations and threads throughout this thesis. In addition, these results indicate a role for future research in the role of immune programming in the gut for the health of preterm infants.

7.5 Strengths and Limitations of this Project.

Overall, this project was successful, and this was a result of several contributing factors that will be discussed in this chapter. However, problems did arise during this project as with any scientific endeavour and if these challenges were not overcome, they will be discussed also in this chapter.

Firstly, a large number of infants were recruited to this project who provided a substantial number of samples. This added a substantial amount of power to the study. As a result, a high amount of confidence could be found in the results that they accurately reflected the parameter being tested. This was a great strength of the project as previous research into the gut microbiome of preterm infants did not have a large sample size (Jacquot *et al.* 2011; Arboleya *et al.* 2012; Barrett *et al.* 2013).

Secondly, the use of two recruitment sites was a positive for the recruitment of the large number of babies, but it also provided a unique opportunity to study the gut microbial communities of infants delivered and raised in different NICUs. To the benefit of this study it showed that the only significant difference was as a result of probiotic administration and therefore the data could be combined and studied as a whole. This information is now available for future studies, to enable even larger cohorts to be recruited across multiple sites.

Thirdly, multiple bodily sites were sampled from the infants recruited. This was another unique methodological technique for studies into the preterm gut microbiome and it yielded very interesting results. As a result, this study has shown that other ‘axis’ within the preterm infant can be investigated and should produce interesting findings to the scientific community.

As shown during this thesis, the methods implemented during this project were reliable and well known, and then found, to produce accurate results. In addition methods developed as part of this project were found to produce reliable results and allowed for insights not previously investigated in preterm infants.

Next, it was relatively easy to collect the stool samples as they would otherwise be considered waste products. Furthermore, the methods used to collect the samples were predominantly non-invasive. Therefore, the ethical considerations of the sample

collection methods for this study were minimal and it did not raise concerns with any of the parents of the participants.

Finally, a large amount of clinical data was collected alongside the samples which allowed for a detailed analysis during this project. This is an important strength of this study that is not as possible for smaller published studies. Moreover, it allowed for analysis to be conducted that hadn't been previously, such as the effect of gender on gut microbiome development. However as with all projects there were difficulties encountered and ways in which could be improved upon next time.

The total protease activity of the stool was low that potentially the method used was not sensitive enough. Firstly, the faecal water samples could have been normalised to a greater protein concentration, as these are samples from preterm infants. This would have resulted in larger numbers which may have been able to provide more insightful results. However, the ratio of sample activity to standard would have been the same so, ultimately this would not have solved the problem. Perhaps a manufactured kit such as the Pierce Colorimetric Protease assay would have been more appropriate, Pierce claim this assay is 1000 times more sensitive than unmodified casein assays.

Secondly, no specific proteases were identified, such as trypsin. This was a significant limitation when the ProteaseArrest assays from G-Biosciences could not identify bacterial and mammalian produced proteases. Without knowing the functionality of the proteases that were detected it was very difficult to assign any role for the proteases in the development of the preterm gut microbiome. Therefore, this is something that needs to be improved in future studies, such as using mass spectrometry would have been able to specifically identify the proteases present. In addition, methodology to investigate the origins of the proteases would be useful even if it was not successful in this study. For example, bacteria could be isolated from the faecal samples and cultured on lactose-free skimmed milk agar, an established method for identifying protease producing bacteria (Morris *et al.* 2012).

Thirdly, even though the results showed little or no local protease activity or inflammation to warrant epithelial barrier integrity experiments. It could have been a final experiment to conclusively determine if the epithelial barrier was compromised in these infants. Experiments that could have been conducted to investigate epithelial

barrier integrity would be trans-epithelial electric resistance (TEER) experiments. These experiments require the growth of an epithelial monolayer on a cell culture insert suspended in a well. The epithelial barrier needs to be fully differentiated for the tight junctions to form. After 21 days the resting resistance of the epithelial barrier would be measured. Then the layer would be exposed to faecal water from the preterm infant and the TEER remeasured after a given time. If the resistance of the layer drops then the proteases in the faecal water could be responsible for the destruction of the tight junctions (Srinivasan *et al.* 2015).

Finally, even though a plethora of data was collected on the infants, very little data was collected on the mother's therefore in future studies microbiome data from the mother's would be useful to identify the origin of the bacteria seeding the infant's microbiome. This was not a problem for this study as the multiple body sites tested and recruitment sites used, there was a lot of data. However, for future studies, particularly those investigating the protease content or gut-lung axis microbial content, data on the mothers would provide useful insights. For example, the origin of the *Actinobacteria* in the NPA samples was not found in this project and perhaps vaginal seeding is responsible but it is not possible to deduce without this information.

7.6 Future Directions.

A strength of the study not detailed in the previous section are the multiple avenues for future studies. This project has answered questions, but it has also generated many more and in this section these possible projects will be discussed.

Firstly, data from the mother's microbiome could be collected. Collecting this data would allow for the origins of colonising bacteria to be identified. Moreover, if combined with the techniques used in this thesis to study the gut-lung axis the seeding origins of multiple body sites on the preterm infant can be investigated. This has great potential. In addition, if the protease activity of the stool of the mothers could also be obtained then this could be compared to the infants. This would eliminate any genetic component ton the levels of protease activity discovered in the infants.

Secondly, furthermore in-depth methods could be used to precisely determine the families of proteases present in the guts of these infants. As previously mentioned, this

could be done by mass spectrometry or by using qPCR techniques to identify the transcription of protease genes within the bacteria and the infants themselves. Moreover, if the protease producing bacteria could be cultured then the proteases could be isolated this was and examined.

It would be beneficial to determine the origin of the proteases present, if they are bacterial or host. This would enable future researchers to determine if the bacteria colonising the gut of preterm infants are interacting with the host and fully exploiting this unique ecological niche. Moreover, investigations into the contribution of bacterial proteases on host health could be investigated.

Some of the future directions discussed in this section such as investigations into the types and origin of proteases present, could be conducted using other "-omic" techniques such as metatranscriptomics and metabolomics. These techniques would allow for studies that encompass the whole of the gut microbiome as discussed above. Moreover, the development of the microbial community during the first month of life has been thoroughly examined and now research into the contributions this community makes to host metabolism and disease needs to be investigated to the same degree. Using a metatranscriptomics technique, like a metagenomic technique, a huge amount od data on the active genes of both the host and bacteria can be investigated. Moreover, a metabolomics technique will again allow for research into the interactions between host and microbial community in the preterm gut microbiome. In addition, the transient diet components can also be taken into consideration using this technique and this will give future researchers access to study the ways the microbial community exploit the diet, or lack thereof, of the infant.

Finally, a previous limitation of the project, that can be taken into consideration with future projects, is the lack of a full-term cohort. Without this cohort it was impossible to comment if a lot of the findings were normal or not. Therefore, to include matched full-term cohorts in future studies would enable for the confirmation of several results of this study. For example, the finding that the lower airways develop independently from the upper airways and stool has been hypothesised to be a result of intubation. However, without results from full term infants this result cannot be confirmed.

7.7 Concluding Remarks

The key novel findings from this study are as follows. For this first time the significant effect of gender on the gut microbial community has been demonstrated. Previously, this effect has only been seen in full term neonates (Martin *et al.* 2016). Secondly, the investigations into the role of stool protease activity in the development of the preterm gut microbiome has been investigated. Unfortunately, the results were not found to be as informative as the role of proteases in the development of adult gut disease (Steck *et al.* 2013). However, it was still an important finding to show that proteases do not appear to have an important role in the development and progression of NEC. Thirdly, the *Bacteroidetes* genus was found by this project and in others to be "protective" against the development of NEC (Pammi *et al.* 2017). A novel finding of this study was that antibiotics significantly reduced the abundance of this phyla in the preterm infant gut. This is an important finding as previous studies have shown that antibiotics increase the incidence of disease in preterm infants, but have yet to identify specific bacteria to be affected by the antibiotics (Cotton 2010). Finally, for the first time the isolated development of the lower airways, upper airways and gut has been found in the preterm infant. Previous research has shown that the gut lung axis refers to the programming of the immune system by the gut, or the effect of gut bacteria on the development of lung diseases (Chen *et al.* 2011b).

The research presented in this thesis has confirmed the development of the preterm gut bacterial community during the first 30 days of life but has begun to shed light on the protease and inflammatory content of the stool. The future of this research has great implications for the health of preterm infants. Moreover, in contrast to the hypothesis, the lack of protease activity could have greater implications for health than the abundant. More targeted approaches need to be implemented, as discussed in this chapter, to isolate and characterize the proteases in the stool of preterm infants. In conclusion, the research provided by this thesis has shown that the gut microbiome of preterm infants is complex, significant to development and full of potential. Therefore, this thesis has provided a unique and significant contribution to scientific knowledge.

Chapter 8

References

Chapter 8. References

Aagaard, K., Ma, J., Antony, K. M., Ganu, R., Petrosino, J. and Versalovic, J. 2014. The placenta harbors a unique microbiome. *Science Translational Medicine* 6(237):237ra265.

Aagaard, K., Riehle, K., Ma, J., Segata, N., Mistretta, T.-A., Coarfa, C., Raza, S., Rosenbaum, S., Van den Veyver, I., Milosavljevic, A., Gevers, D., Huttenhower, C., Petrosino, J., and Versalovic, J. 2012. A Metagenomic Approach to Characterization of the Vaginal Microbiome Signature in Pregnancy. *PLoS ONE* 7(6):e36466.

Abdulkadir, B., Nelson, A., Skeath, T., Marrs, E. C. L., Perry, J. D., Cummings, S. P., Embleton, N.D., Berrington, J.E., and Stewart, C. J. 2016a. Routine Use of Probiotics in Preterm Infants: Longitudinal Impact on the Microbiome and Metabolome. *Neonatology* 109(4):239-247.

Abdulkadir, B., Nelson, A., Skeath, T., Marrs, E. C. L., Perry, J. D., Cummings, S. P., Embleton, N.D., Berrington, J.E., and Stewart, C. J. 2016b. Stool bacterial load in preterm infants with necrotising enterocolitis. *Early Human Development* 95:1-2.

Abrahamse-Berkeveld, M., Alles, M., Franke-Beckmann, E., Helm, K., Knecht, R., Köllges, R., Sander, B., Ben Amor, K., and Bufe, A. 2016. Infant formula containing galacto-and fructo-oligosaccharides and Bifidobacterium breve M-16V supports adequate growth and tolerance in healthy infants in a randomised, controlled, double-blind, prospective, multicentre study. *Journal of Nutritional Science* 5:e42.

Abrahamsson, T. R., Jakobsson, H. E., Andersson, A. F., Björkstén, B., Engstrand, L. and Jenmalm, M. C. 2014. Low gut microbiota diversity in early infancy precedes asthma at school age. *Clinical and Experimental Allergy: Journal of the British Society for Allergy and Clinical Immunology* 44(6):842-850.

Aceti, A., Gori, D., Barone, G., Callegari, M. L., Di Mauro, A., Fantini, M. P., Indrio, F., Maggio, L., Meneghin, F., Morelli, L., Zuccotti, G., Corvaglia, L., and Italian Society of Neonatology. 2015. Probiotics for prevention of necrotizing enterocolitis in preterm infants: systematic review and meta-analysis. *Italian Journal of Pediatrics* 41:89.

Adlerberth, I. and Wold, A. E. 2009. Establishment of the gut microbiota in Western infants. *Acta Paediatrica* 98(2):229-238.

Afjeh, S.-A., Sabzehei, M.-K., Fahimzad, S.-A.-R., Shiva, F., Shamshiri, A.-R. and Esmaili, F. 2016. Antibiotic Therapy for Very Low Birth Weigh Newborns in NICU. *Iranian Journal of Pediatrics* 26(2):e2612.

Akay, H. K., Bahar Tokman, H., Hatipoglu, N., Hatipoglu, H., Siraneci, R., Demirci, M., Borsa, B.A., Yuksel, P., Karakullukcu, A., Kangaba, A.A., Sirekbasan, S., Aka, S., Mamal Torun, M., and Kocazeybek, B. S. 2014. The relationship between bifidobacteria and allergic asthma and/or allergic dermatitis: A prospective study of 0–3 years-old children in Turkey. *Anaerobe* 28:98-103.

Al-Hosni, M., Duenas, M., Hawk, M., Stewart, L. A., Borghese, R. A., Cahoon, M., Atwood, L., Howard, D., Ferrelli, K., and Soll, R. 2012. Probiotics-supplemented feeding in extremely low-birth-weight infants. *Journal of Perinatology* 32(4):253-259.

Alcon-Giner, C., Caim, S., Mitra, S., Ketskemety, J., Wegmann, U., Wain, J., Belteki, G., Clarke, P., and Hall, L. J. 2017. Optimisation of 16S rRNA gut microbiota profiling of extremely low birth weight infants. *BMC Genomics* 18(1):841.

Alexander, V. N., Northrup, V. and Bizzarro, M. J. 2011. Antibiotic Exposure in the Newborn Intensive Care Unit and the Risk of Necrotizing Enterocolitis. *The Journal of pediatrics* 159(3):392-397.

Amann, R. I., Ludwig, W. and Schleifer, K. H. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews* 59(1):143-169.

Annaházi, A., Gecse, K., Dabek, M., Ait-Belgnaoui, A., Rosztóczy, A., Róka, R., Molnar, T., Theodorou, V., Wittmann, T., Bueno, L., and Eutamene, H. 2009. Fecal proteases from diarrheic-IBS and ulcerative colitis patients exert opposite effect on visceral sensitivity in mice. *PAIN* 144(1):209-217.

Antalis, T. M., Buzza, M. S., Hodge, K. M., Hooper, J. D. and Netzel-Arnett, S. 2010. The cutting edge: membrane-anchored serine protease activities in the pericellular microenvironment. *Biochemical Journal* 428(3):325-346.

Antalis, T. M. and Lawrence, D. A. 2004. Serpin mutagenesis. *Methods* 32(2):130-140.

Antalis, T. M., Shea-Donohue, T., Vogel, S. N., Sears, C. and Fasano, A. 2007. Mechanisms of disease: protease functions in intestinal mucosal pathobiology. *Nature Clinical Practice. Gastroenterology & Hepatology* 4(7):393-402.

Antonowicz, I. and Lebenthal, E. 1977. Developmental Pattern of Small Intestinal Enterokinase and Disaccharidase Activities in the Human Fetus. *Gastroenterology* 72(6):1299-1303.

Anvarian, A. H. P., Cao, Y., Srikumar, S., Fanning, S. and Jordan, K. 2016. Flow Cytometric and 16S Sequencing Methodologies for Monitoring the Physiological Status of the Microbiome in Powdered Infant Formula Production. *Frontiers in Microbiology* 7:968.

Appel, W. 1986. Chymotrypsin: Molecular and catalytic properties. *Clinical Biochemistry* 19(6):317-322.

Arboleya, S., Binetti, A., Salazar, N., Fernández, N., Solís, G., Hernández-Barranco, A., Margolles, A., de Los Reyes-Gavilan, C. G., and Gueimonde, M. 2012. Establishment and development of intestinal microbiota in preterm neonates. *FEMS microbiology ecology* 79(3):763-772.

Arboleya, S., Sánchez, B., Milani, C., Duranti, S., Solís, G., Fernández, N., de Los Reyes-Gavilan, C.G., Ventura, M., Margolles, A., and Gueimonde, M. 2015. Intestinal microbiota development in preterm neonates and effect of perinatal antibiotics. *The Journal of Pediatrics* 166(3):538-544.

Arboleya, S., Sánchez, B., Solís, G., Fernández, N., Suárez, M., Hernández-Barranco, A. M., Milani, C., Margolles, A., de Los Reyes-Gavilan, C.G., Ventura, M., and Gueimonde, M. 2016. Impact of Prematurity and Perinatal Antibiotics on the Developing Intestinal Microbiota: A Functional Inference Study. *International Journal of Molecular Sciences* 17(5):649.

Armaforte, E., Curran, E., Huppertz, T., Ryan, C. A., Caboni, M. F., O'Connor, P. M., Ross, R.P., Hirtz, C., Sommerer, N., Chevalier, F., and Kelly, A. L. 2010. Proteins and proteolysis in pre-term and term human milk and possible implications for infant formulae. *International Dairy Journal* 20(10):715-723.

Arrieta, M.-C., Stiemsma, L. T., Dimitriu, P. A., Thorson, L., Russell, S., Yurist-Doutsch, S., Kuzeljevic, B., Gold, M.J., Britton, H.M., Lefebvre, D.L., Subbarao, P., Mandhane, P., Becker, A., McNagny, K.M., Sears, M.R., Kollman, T., Child Study Investigators, Mohn, W.W., Turvey, S.E., and Brett Finlay, B. 2015. Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Science Translational Medicine* 7(307):307ra152.

Arslanoglu, S., Moro Guido, E., Ziegler Ekhard, E. and the, WAPM Working Group on Nutrition. 2010. Optimization of human milk fortification for preterm infants: new concepts and recommendations. *Journal of Perinatal Medicine* 38(3):223-228.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., Fernandes, G.R., Tap, J., Bruls, T., Batto, J-M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H.B., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E.G., Eang, J., Guarner, F., Pedersen, O., de Vos, W.M., Brunak, S., Dore, J., Meta HIT Consortium, Antolin, M., Artiguenave, F., Blottierre, H.M., Almeida, M., Brechot, C., Cara, C., Chevaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Foerstner, K.U., Friss, C., van de Guchte, M., Guedon, E., Haimet, F., Huber, W., van Hylckama-Vlieg, J., Jamet, A., Juste, C., Kaci, G., Melo Minardi, R., M'Rini, C., Muller, J., Oozeer, R., Parkhill, J., Renault, P., Rescigno, M., Sanchez, N., Sunagawa, S., Torrejon, A., Turner, K., Vandemeulebrouck, G., Varela, E., Winogradsky, Y., Zeller, G., Weissenback, J., Ehlrich, S.D., and Bork, P. 2011. Enterotypes of the human gut microbiome. *Nature* 473(7346):174-180.

Astrup, T. and Sterndorff, I. 1953. A Fibrinolytic System in Human iMilk. *Proceedings of the Society for Experimental Biology and Medicine* 84(3):605-608.

Aurélien, T., Eva, S. G., Koshika, Y., Anke, K. S., Norbert, S., Catherine, N.-B., Blanchard, C., Junt, T., Nicod, L.P., Harris, N.L., and Benjamin, J. M. 2014. Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. *Nature Medicine* 20(2):159.

Autran, C. A., Kellman, B. P., Kim, J. H., Asztalos, E., Blood, A. B., Spence, E. C. H., Patel, A.L., Hou, J., Lewis, N.E., and Bode, L. 2017. Human milk oligosaccharide composition predicts risk of necrotising enterocolitis in preterm infants. *Gut* 67(6):1064-1070.

Autran, C. A., Schoterman, M. H. C., Jantscher-Krenn, E., Kamerling, J. P. and Bode, L. 2016. Sialylated galacto-oligosaccharides and 2′-fucosyllactose reduce necrotising enterocolitis in neonatal rats. *British Journal of Nutrition* 116(2):294-299.

Avershina, E., Lundgård, K., Sekelja, M., Dotterud, C., Storrø, O., Øien, T., Johnsen, R., and Rudi, K. 2016. Transition from infant- to adult-like gut microbiota. *Environmental Microbiology* 18(7):22626-2236.

Avershina, E., Storrø, O., Øien, T., Johnsen, R., Pope, P. and Rudi, K. 2014. Major faecal microbiota shifts in composition and diversity with age in a geographically restricted cohort of mothers and their children. *FEMS Microbiology Ecology* 87(1):280-290.

Azad, M. B., Konya, T., Maughan, H., Guttman, D. S., Field, C. J., Chari, R. S., Sears, M.R., Becker, A.B., Scott, J.A., and Kozyrskyj, A. L. 2013. Gut microbiota of healthy

Canadian infants: profiles by mode of delivery and infant diet at 4 months. *Canadian Medical Association Journal* 185(5):385-394.

Azad, M. B., Konya, T., Persaud, R. R., Guttman, D. S., Chari, R. S., Field, C. J., Sears, M.J., Mandhane, P.J., Turvey, S.E., Subbarao, P., Becker, A.B., Scott, J.A., Kozyrskyj, A.L., and Child Study Investigators. 2016. Impact of maternal intrapartum antibiotics, method of birth and breastfeeding on gut microbiota during the first year of life: a prospective cohort study. *The journal of obstetrics and gynaecology of the british empire* 123(6):983-993.

Backert, S. and Hofreuter, D. 2013. Molecular methods to investigate adhesion, transmigration, invasion and intracellular survival of the foodborne pathogen Campylobacter jejuni. *Journal of Microbiological Methods* 95(1):8-23.

Bailey, A. and Keshav, S. 2012. *Satish Keshav and Adam Bailey. Gastrointestinal System at a Glance - Second Edition.* Wiley-Blackwell 2013.

Barrett, E., Kerr, C., Murphy, K., O'Sullivan, O., Ryan, C. A., Dempsey, E. M., Murphy, B.P., O'Toole, P.W., Cotter, P.D., Fitzgerald, G.F., Ross, P.R., and Stanton, C. 2013. The individual-specific and diverse nature of the preterm infant microbiota. *Archives of Disease in Childhood - Fetal and Neonatal Edition* 98(4):F334-340.

Barron, L. K., Warner, B. B., Tarr, P. I., Shannon, W. D., Deych, E. and Warner, B. W. 2017. Independence of gut bacterial content and neonatal necrotizing enterocolitis severity. *Journal of Pediatric Surgery* 52(6):1531-5037

Battersby, A. J. and Gibbons, D. L. 2013. The gut mucosal immune system in the neonatal period. *Pediatric Allergy and Immunology* 24(5):414-421.

Baugh, M. D., Perry, M. J., Hollander, A. P., Davies, D. R., Cross, S. S., Lobo, A. J., Taylor, C.J., and Evans, G. S. 1999. Matrix metalloproteinase levels are elevated in inflammatory bowel disease. *Gastroenterology* 117(4):814-822.

Baumgart, M., Dogan, B., Rishniw, M., Weitzman, G., Bosworth, B., Yantiss, R., Orsi, R.H., Wiedmann, M., McDonough, P., Kim, S.G., Berg, D., Schukken, Y., Scherl, E., and Simpson, K. W. 2007. Culture independent analysis of ileal mucosa reveals a selective increase in invasive Escherichia coli of novel phylogeny relative to depletion of Clostridiales in Crohn's disease involving the ileum. *The ISME journal* 1(5):403-418.

Beals, E. W. 1984. Bray-Curtis Ordination: An Effective Strategy for Analysis of Multivariate Ecological Data. *Advances in Ecological Research* 14(1984):1-55.

Beck, S., Wojdyla, D., Say, L., Betran, A. P., Merialdi, M., Requejo, J. H., Rubens, C., Menon, R., and Van Look, P. F. A. 2010. The worldwide incidence of preterm birth: a systematic review of maternal mortality and morbidity. *Bulletin of the World Health Organization* 88(1):31-38.

Beeton, M. L., Maxwell, N. C., Davies, P. L., Nuttall, D., McGreal, E., Chakraborty, Spiller, B.O., and M., Kotecha, S. 2011. Role of pulmonary infection in the development of chronic lung disease of prematurity. *European Respiratory Journal* 37(6):1424.

Bein, A., Lubetzky, R., Mandel, D. and Schwartz, B. 201). TIMP-1 inhibition of occludin degradation in Caco-2 intestinal cells: a potential protective role in necrotizing enterocolitis. *Pediatric Research* 77(5):649-655.

Benkoe, T., Reck, C., Gleiss, A., Kettner, S., Repa, A., Horcher, E. and Rebhandl, W. 2012. Interleukin 8 correlates with intestinal involvement in surgically treated infants with necrotizing enterocolitis. *Journal of Pediatric Surgery* 47(8):1548-1554.

Biagi, E., Nylund, L., Candela, M., Ostan, R., Bucci, L., Pini, E., Nikkila, J., Monti, D., Satokari, R., Franceschi, C., Brigidi, P., and De Vos, W. 2010. Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. *PloS One* 5(5):e10667.

Biancheri, P., Di Sabatino, A., Corazza, G. R. and MacDonald, T. T. 2013. Proteases and the gut barrier. *Cell and Tissue Research* 351(2):269-280.

Bister, V., Salmela, M. T., Heikkilä, P., Anttila, A., Rintala, R., Isaka, K., Andersson, S., and Saarialho-Kere, U. 2005. Matrilysins-1 and -2 (MMP-7 and -26) and metalloelastase (MMP-12), unlike MMP-19, are up-regulated in necrotizing enterocolitis. *Journal of Pediatric Gastroenterology and Nutrition* 40(1):60-66.

Biswas, S. and Biswas, I. 2005. Role of HtrA in surface protein expression and biofilm formation by Streptococcus mutans. *Infection and immunity* 73(10):6923-6934.

Blackwood, B. P., Hunter, C. J. and Grabowski, J. 2017. Variability in Antibiotic Regimens for Surgical Necrotizing Enterocolitis Highlights the Need for New Guidelines. *Surgical infections* 18(2):215-220.

Blakely, M. L., Lally, K. P., McDonald, S., Brown, R. L., Barnhart, D. C., Ricketts, R. R., Thompson, W.R., Scherer, L.R., Klein, M.D., Letton, R.W., Chwals, W.J., Toulokian, R.J., Kurkchubache, A.G., Skinner, M.A., Moss, L.R., Hilfiker, M.L., and

N. E. C. Subcommittee of the NICHD Neonatal Research Network. 2005. Postoperative outcomes of extremely low birth-weight infants with necrotizing enterocolitis or isolated intestinal perforation: a prospective cohort study by the NICHD Neonatal Research Network. *Annals of Surgery* 241(6):984-994.

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., and Shao, Y. 1997. The Complete Genome Sequence of Escherichia coli K-12. *Science* 277(5331):1453.

Boehm, M., Hoy, B., Rohde, M., Tegtmeyer, N., Bæk, K. T., Oyarzabal, O. A., Bronsted, L., Wessler, S., and Backert, S. 2012. Rapid paracellular transmigration of Campylobacter jejuni across polarized epithelial cells without affecting TER: role of proteolytic-active HtrA cleaving E-cadherin but not fibronectin. *Gut Pathogens* 4(1):3.

Bokulich, N. A., Mills, D. A. and Underwood, M. A. 2013. Surface Microbes in the Neonatal Intensive Care Unit: Changes with Routine Cleaning and over Time. *Journal of Clinical Microbiology* 51(8):2617-2624.

BorgstrÖM, B., Lindquist, B. and Lundh, G. 1960. Enzyme Concentration and Absorption of Protein and Glucose in Duodenum of Premature Infants. *A.M.A. Journal of Diseases of Children* 99(3):338-343.

Braegger, C. P., Spencer, J. and MacDonald, T. T. 1996. Ontogenetic aspects of the intestinal immune system in man. *International Journal of Clinical and Laboratory Research* 26(3):1-4.

Brannon, J. R., Burk, D. L., Leclerc, J.-M., Thomassin, J.-L., Portt, A., Berghuis, A. M., Gruenheid, S., and Le Moual, H. 2015. Inhibition of outer membrane proteases of the omptin family by aprotinin. *Infection and Immunity* 83(6):2300-2311.

Britton, J. R. and Koldovsky, O. 1989. Development of Luminal Protein Digestion: Implications for Biologically Active Dietary Polypeptides. *Journal of Pediatric Gastroenterology and Nutrition* 9(2):144-162.

Brock, J. H., Arzabe, F., Lampreave, F. and Piñeiro, A. 1976. The effect of trypsin on bovine transferrin and lactoferrin. *BBA - Protein Structure* 446(1):214-225.

Brooks, B., Firek, B. A., Miller, C. S., Sharon, I., Thomas, B. C., Baker, R., Morowitz, M.J., and Banfield, J. F. 2014. Microbes in the neonatal intensive care unit resemble those found in the gut of premature infants. *Microbiome* 2(1):1.

Bruewer, M., Luegering, A., Kucharzik, T., Parkos, C. A., Madara, J. L., Hopkins, A. M. and Nusrat, A. 2003. Proinflammatory cytokines disrupt epithelial barrier function by apoptosis-independent mechanisms. *Journal of Immunology* 171(11):6164-6172.

Buckel, W. 1986. Substrate stereochemistry of the biotin-dependent sodium pump glutaconyl-CoA decarboxylase from Acidaminococcus fermentans. *European Journal of Biochemistry* 156(2):259-263.

Budden, K. F., Gellatly, S. L., Wood, D. L. A., Cooper, M. A., Morrison, M., Hugenholtz, P. and Hansbro, P. M. 2017. Emerging pathogenic links between microbiota and the gut–lung axis. *Nature Reviews Microbiology* 15(1):55-63.

Bueno, L. and Fioramonti, J. 2008. Protease-activated receptor 2 and gut permeability: a review. *Neurogastroenterology and Motility: The Official Journal of the European Gastrointestinal Motility Society* 20(6):580-587.

Buffie, C. G. and Pamer, E. G. 2013. Microbiota-mediated colonization resistance against intestinal pathogens. *Nature reviews. Immunology* 13(11):790-801.

Bustos, D., Negri, G., De Paula, J. A., Di Carlo, M., Yapur, V., Facente, A. and De Paula, A. 1998. Colonic proteinases: increased activity in patients with ulcerative colitis. *Medicina* 58(3):262-264.

Buzza, M. S., Netzel-Arnett, S., Shea-Donohue, T., Zhao, A., Lin, C.-Y., List, K., Szabo, R., Fasano, A., Bugge, T.H., and Antalis, T. M. 2010. Membrane-anchored serine protease matriptase regulates epithelial barrier formation and permeability in the intestine. *Proceedings of the National Academy of Sciences* 107(9):4200-4205.

Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. and Gordon, J. I. 2005. Host-Bacterial Mutualism in the Human Intestine. *Science* 307(5717):1915-1920.

Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Li, Y., Xia, Y., Xie, H., Zhong, H., Khan, M.T., Zhang, J., Li, J., Xiao, L., Al-Aama, J., Zhang, D., Lee, Y.S., Kotowska, D., Colding, C., Tremaroli, V., Yin, Y., Bergmen, S., Xu, X., Madsen, L., Kristiansen, K., Dahlgren, J., and Wang, J. 2015. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host & Microbe* 17(6):852.

Cader, M. Z. and Kaser, A. 2013. Recent advances in inflammatory bowel disease: mucosal immune cells in intestinal inflammation. *Gut* 62(11):1653-1664.

Camilleri, M., Madsen, K., Spiller, R., Van Meerveld, B. G. and Verne, G. N. 2012. Intestinal barrier function in health and gastrointestinal disease. *Neurogastroenterology & Motility* 24(6):503-512.

Cantarel, B. L., Lombard, V. and Henrissat, B. 2012. Complex Carbohydrate Utilization by the Healthy Human Microbiome. *PLOS ONE* 7(6):e28742.

Cantey, J. B., Wozniak, P. S., Pruszynski, J. E. and Sánchez, P. J. 2016. Reducing unnecessary antibiotic use in the neonatal intensive care unit (SCOUT): a prospective interrupted time-series study. *The Lancet. Infectious Diseases* 16(10):1178-1184.

Cantey, J. B., Wozniak, P. S. and Sánchez, P. J. 2015. Prospective surveillance of antibiotic use in the neonatal intensive care unit: results from the SCOUT study. *The Pediatric Infectious Disease Journal* 34(3):267-272.

Carr, J. P., Burgner, D. P., Hardikar, R. S. and Buttery, J. P. 2017. Empiric antibiotic regimens for neonatal sepsis in Australian and New Zealand neonatal intensive care units. *Journal of Paediatrics and Child Health* 53(7):680-684.

Carroll, I. M. and Maharshak, N. 2013. Enteric bacterial proteases in inflammatory bowel disease- pathophysiology and clinical implications. *World Journal of Gastroenterology* 19(43):7531-7543.

Carroll, I. M., Ringel-Kulka, T., Ferrier, L., Wu, M. C., Siddle, J. P., Bueno, L. and Ringel, Y. 2013. Fecal Protease Activity Is Associated with Compositional Alterations in the Intestinal Microbiota. *PLoS ONE* 8(10):e78017.

Castanys-Muñoz, E., Martin, M. J. and Vazquez, E. 2016. Building a Beneficial Microbiome from Birth. *Advances in Nutrition* 7(2):323-330.

Cato, E. P., Johnson, J. L., Hash, D. E. and Holdeman, L. V. 1983. Synonomy of Peptococcus glycinophilus (Cardon and Barker 1946) Douglas 1957 with Peptostreptococcus micros (Prevot 1933) Smith 1957 and electrophoretic differentiation of Peptostreptococcus micros from Peptococcus magnus (Prevot 1933) Holdeman and Moore 1972. *International Journal of Systematic Bacteriology* 33(2):207-210.

Cenac, N., Andrews, C. N., Holzhausen, M., Chapman, K., Cottrell, G., Andrade-Gordon, P., Steinhoff, M., Barbara, G., Beck, P., Bunnett, N.W., Sharkey, K.A., Ferraz, J.G.P., Shaffer, E., and Vergnolle, N. 2007. Role for protease activity in visceral pain in irritable bowel syndrome. *The Journal of Clinical Investigation* 117(3):636-647.

Cenac, N., Coelho, A.-M., Nguyen, C., Compton, S., Andrade-Gordon, P., MacNaughton, W. K., Wallace, K., Wallace, J.L., Hollenberg, M.D., Bunnett, N.W., Garcia-Villar, R., Bueno, L., and Vergnolle, N. 2002. Induction of intestinal inflammation in mouse by activation of proteinase-activated receptor-2. *The American Journal of Pathology* 161(5):1903-1915.

Chakraborti, S. and Dhalla, N. S. 2017. *Proteases in Physiology and Pathology*. Singapore : Springer Singapore : Imprint: Springer.

Charles A Janeway, Jr., Travers, P., Walport, M. and Shlomchik, M. J. (2001). The mucosal immune system. *Jameway's Immunobiology 9th Edition*.

Chassaing, B., Rolhion, N., de Vallée, A., Salim, S. a. Y., Prorok-Hamon, M., Neut, C., Campbell, B.J., Soderholm, J.D., Hugot, J-P., Colombel, J-F., and Darfeuille-Michaud, A. 2011. Crohn disease--associated adherent-invasive E. coli bacteria target mouse and human Peyer's patches via long polar fimbriae. *The Journal of Clinical Investigation* 121(3):966-975.

Chen, L.-W., Chen, P.-H. and Hsu, C.-M. 2011a. Commensal Microflora Contribute to Host Defense Against Escherichia Coli Pneumonia Through Toll-Like Receptors. *Shock* 36(1)67-75.

Chen, L.-W., Chen, P.-H. and Hsu, C.-M. 2011b. Commensal microflora contribute to host defense against Escherichia coli pneumonia through Toll-like receptors. *Shock* 36(1):67-75.

Chen, Y., Hung, W.-T., Chen, S.-M. and Tseng, S.-H. 2002. Suppression of elevated plasma interleukin-8 levels due to total ischemia and reperfusion of the small intestine by luminal perfusion with fetal bovine serum. *Pediatric Surgery International* 18(2):107-109.

Chien, J., Ota, T., Aletti, G., Shridhar, R., Boccellino, M., Quagliuolo, L., Baldi, A., and Shridhar, V. 2009. Serine protease HtrA1 associates with microtubules and inhibits cell migration. *Molecular and Cellular Biology* 29(15):4177-4187.

Christensen, B., Schack, L., Klaning, E. and Sorensen, E. 2010. Osteopontin Is Cleaved at Multiple Sites Close to Its Integrin-binding Motifs in Milk and Is a Novel Substrate for Plasmin and Cathepsin D. *Journal Of Biological Chemistry* 285(11):7929-7937.

Christine, D., Rashmi, S., James, J. G., Zhiheng, P., Liying, Y., Richard, B. H. and Jiyoung, A. 2015. Sex, body mass index, and dietary fiber intake influence the human gut microbiome. *PLoS ONE* 10(4):e0124599.

Chu, D. M., Ma, J., Prince, A. L., Antony, K. M., Seferovic, M. D. and Aagaard, K. M. 2017. Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. *Nature Medicine* 23(3)314-326.

Claesson, M. J., Cusack, S., O'Sullivan, O., Greene-Diniz, R., de Weerd, H., Flannery, E., Marchesi, J.R., Falush, D., Dinan, T., Fitzgerald, G., Stanton, C., van Sinderen, D., O'Connor, M., Harendy, N., O'Connor, K., Henry, C., O'Mahoney, D., Fitzgerald, A.P., Shanahan, F., Twomey, C., Hill, C., Ross, P.R., and O'Toole, P. W. 2011. Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proceedings of the National Academy of Sciences of the United States of America* 108 Suppl 1:4586-4591.

Clark, R. H., Bloom, B. T., Spitzer, A. R. and Gerstmann, D. R. 2006. Reported medication use in the neonatal intensive care unit: data from a large national data set. *Pediatrics* 117(6):1979-1987.

Clarke, T. B. 2014. Early innate immunity to bacterial infection in the lung is regulated systemically by the commensal microbiota via nod-like receptor ligands. *Infection and immunity* 82(11):4596-4606.

Clausen, T., Kaiser, M., Huber, R. and Ehrmann, M. 2011. HTRA proteases: regulated proteolysis in protein quality control. *Nature Reviews. Molecular Cell Biology* 12(3):152-162.

Clausen, T., Southan, C. and Ehrmann, M. 2002. The HtrA family of proteases: implications for protein composition and cell fate. *Molecular Cell* 10(3):443-455.

Clayburgh, D. R., Shen, L. and Turner, J. R. 2004. A porous defense: the leaky epithelial barrier in intestinal disease. *Laboratory Investigation* 84(3):282.

Clemente, M. G., De Virgiliis, S., Kang, J. S., Macatagney, R., Musu, M. P., Di Pierro, M. R., Drago, S., Congia, M., and Fasano, A. 2003. Early effects of gliadin on enterocyte intracellular signalling involved in intestinal barrier function. *Gut* 52(2):218-223.

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., Brown, T.C., Porras-Alfaro, A., Kuske, C.R., and Tiedje, J. M. 2014. Ribosomal Database Project:

data and tools for high throughput rRNA analysis. *Nucleic Acids Research* 42(Database issue):D633-D642.

Coleman, J. L., Crowley, J. T., Toledo, A. M. and Benach, J. L. 2013. The HtrA protease of Borrelia burgdorferi degrades outer membrane protein BmpD and chemotaxis phosphatase CheX. *Molecular Microbiology* 88(3):619-633.

Colin, B., Clifford, S., Wu, P., Rathmanner, S. and Mengersen, K. 2017. Using Boosted Regression Trees and Remotely Sensed Data to Drive Decision-Making. *Open Journal of Statistics* 07:859-875.

Collado, M. C., Cernada, M., Neu, J., Pérez-Martínez, G., Gormaz, M. and Vento, M. 2015. Factors influencing gastrointestinal tract and microbiota immune interaction in preterm infants. *Pediatric Research* 77(6):726-731.

Cong, X., Judge, M., Xu, W., Diallo, A., Janton, S., Brownell, E. A., Mass, K., and Graf, J. 2017. Influence of Feeding Type on Gut Microbiome Development in Hospitalized Preterm Infants. *Nursing Research* 66(2):123-133.

Cong, X., Xu, W., Janton, S., Henderson, W. A., Matson, A., McGrath, J. M., .Maas, K., and Graf, J. 2016. Gut Microbiome Developmental Patterns in Early Life of Preterm Infants: Impacts of Feeding and Gender. *PloS One* 11(4):e0152751.

The Human Microbiome Consortium, Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., and White, O. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486(7402):207.

Costello, E. K., Carlisle, E. M., Bik, E. M., Morowitz, M. J. and Relman, D. A. 2013. Microbiome assembly across multiple body sites in low-birthweight infants. *mBio* 4(6):e00782-00713.

Costeloe, K., Hardy, P., Juszczak, E., Wilks, M. and Millar, M. R. 2016. Bifidobacterium breve BBG-001 in very preterm infants: a randomised controlled phase 3 trial. *The Lancet* 387(10019):649-660.

Cotten, C. M., Taylor, S., Stoll, B., Goldberg, R. N., Hansen, N. I., Sánchez, P. J., Ambalavanan, N., Benjamin, D.K., NICHD Neontal Research Network. 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):58-66.

Cotton, C. M. 2010. Early, prolonged use of postnatal antibiotics increased the risk of necrotising enterocolitis. *Archives of disease in childhood. Education and practice edition* 95(3):94.

Cottrell, G. S., Amadesi, S., Grady, E. F. and Bunnett, N. W. 2004. Trypsin IV, a novel agonist of protease-activated receptors 2 and 4. *The Journal of Biological Chemistry* 279(14):13532-13539.

Cottrell, G. S., Amadesi, S., Schmidlin, F. and Bunnett, N. 2003. Protease-activated receptor 2: activation, signalling and function. *Biochemical Society Transactions* 31(Pt 6):1191-1197.

Coughlin, S. R. 1999. How the protease thrombin talks to cells. *Proceedings of the National Academy of Sciences of the United States of America* 96(20):11023-11027.

Dahl, C., Stigum, H., Valeur, J., Iszatt, N., Lenters, V., Peddada, S., Bjornholt, J.V., Midtvedt, T., Mandal, S., and Eggesbø, M. 2018. Preterm infants have distinct microbiomes not explained by mode of delivery, breastfeeding duration or antibiotic exposure. *International Journal of Epidemiology* 47(5):1658-1669.

Dallas, D. C., Underwood, M. A., Zivkovic, A. M. and German, J. B. 2012. Digestion of Protein in Premature and Term Infants. *Journal of nutritional disorders & therapy* 2(3):112-112.

Dang, A. T. and Marsland, B. J. 2019. Microbes, metabolites, and the gut–lung axis. *Mucosal Immunology* 12(4):843-850.

Davenport, E. R., Mizrahi-Man, O., Michelini, K., Barreiro, L. B., Ober, C. and Gilad, Y. 2014. Seasonal variation in human gut microbiome composition. *PloS One* 9(3):e90731.

Davidson, L. A. and Lönnerdal, B. 1987. Persistence of Human Milk Proteins in the Breast-Fed Infant. *Acta Pædiatrica* 76(5):733-740.

Davies, P. L., Spiller, O. B., Beeton, M. L., Maxwell, N. C., Remold-O'Donnell, E. and Kotecha, S. 2010. Relationship of proteinases and proteinase inhibitors with microbial presence in chronic lung disease of prematurity. *Thorax* 65(3):246-251.

de Boer, R., Peters, R., Gierveld, S., Schuurman, T., Kooistra-Smid, M. and Savelkoul, P. 2010. Improved detection of microbial DNA after bead-beating before DNA isolation. *Journal of Microbiological Methods* 80(2):209-211.

De Keukeleire, S., Borrey, D., Decaluwe, W. and Reynders, M. 2016. Therapeutic Drug Monitoring of Meropenem in Neonate with Necrotizing Enterocolitis: A Challenge. *Case Reports in Infectious Diseases* 2016:6207487.

de la Cochetière, M.-F., Piloquet, H., des Robert, C., Darmaun, D., Galmiche, J.-P. and Rozé, J.-C. 2004. Early Intestinal Bacterial Colonization and Necrotizing Enterocolitis in Premature Infants: The Putative Role of Clostridium. *Pediatric Research* 56(3):366-370.

De Plaen, I. G. 2013. Inflammatory signaling in necrotizing enterocolitis. *Clinics in perinatology* 40(1):109-124.

Dean, R. S. and Grizzle, R. 2011. Prenatal Development. In: *Encyclopedia of Child Behavior and Development*. Springer, Boston, MA, pp. 1151-1152.

Del Chierico, F., Vernocchi, P., Petrucca, A., Paci, P., Fuentes, S., Praticò, G., Capuani, G., Masotti, A., Reddel, S., Russo, A., Vallone, C., Salvatori, G., Buffone, E., Signore, F., Rigon, G., Dottta, A., Miccheli, A., de Vos, W.M., Dallapiccola, B., and Putignani, L. 2015. Phylogenetic and Metabolic Tracking of Gut Microbiota during Perinatal Development. *PLoS ONE* 10(9):e0137347.

Denizot, J., Sivignon, A., Barreau, F., Darcha, C., Chan, C., P Stanners, C., Hofman, P., Darfeuille-Michaud, A., and Barnich, N. 2011. Adherent-Invasive Escherichia coli Induce Claudin-2 Expression and Barrier Defect in CEABAC10 Mice and Crohn's Disease Patients. *Inflammatory Bowel Disease* 18(2): 294-304.

Derré, I. 2015. Chlamydiae interaction with the endoplasmic reticulum: contact, function and consequences. *Cellular microbiology* 17(7):959-966.

Desbonnet, L., Clarke, G., Shanahan, F., Dinan, T. G. and Cryan, J. F. 2014. Microbiota is essential for social development in the mouse. *Molecular Psychiatry* 19(2):146-148.

Dethlefsen, L., McFall-Ngai, M. and Relman, D. A. 2007. An ecological and evolutionary perspective on human-microbe mutualism and disease. *Nature* 449(7164):811-818.

Dethlefsen, L. and Relman, D. A. 2011. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proceedings of the National Academy of Sciences of the United States of America* 108 Suppl 1:4554-4561.

Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C. R., Yu, W.-H., Lakshmanan, A, and Wade, W. G. 2010. The Human Oral Microbiome. *Journal of Bacteriology* 192(19):5002-5017.

Dickson, R. P., Singer, B. H., Newstead, M. W., Falkowski, N. R., Erb-Downward, J. R., Standiford, T. J. and Huffnagle, G. B. 2016. Enrichment of the lung microbiome with gut bacteria in sepsis and the acute respiratory distress syndrome. *Nature microbiology* 1(10):16113-16113.

DiGiulio, D. B. 2012. Diversity of microbes in amniotic fluid. *Seminars in Fetal & Neonatal Medicine* 17(1):2-11.

DiGiulio, D. B., Callahan, B. J., McMurdie, P. J., Costello, E. K., Lyell, D. J., Robaczewska, A., Relman, D. A. 2015. Temporal and spatial variation of the human microbiota during pregnancy. *Proceedings of the National Academy of Sciences* 112(35):11060-11065.

Dominguez-Bello, M. G., Costello, E. K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N. and Knight, R. 2010. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences of the United States of America* 107(26):11971-11975.

Downard, C. D., Renaud, E., St Peter, S. D., Abdullah, F., Islam, S., Saito, J. M., Blakely, M.L., Huang., E.Y., Arca., M.J., Cassidy, L., Aspelund, G., and American Pediatric Surgical Association Outcomes Clinical Trials, C. 2012. Treatment of necrotizing enterocolitis: an American Pediatric Surgical Association Outcomes and Clinical Trials Committee systematic review. *Journal of Pediatric Surgery* 47(11):2111-2122.

Drago, S., El Asmar, R., Di Pierro, M., Grazia Clemente, M., Tripathi, A., Sapone, A., Thakar, M., Iacono, G., Carroccio, A., D'Agate, C., Not, T., Zampini, L., Catassi, C., and Fasano, A. 2006. Gliadin, zonulin and gut permeability: Effects on celiac and non-celiac intestinal mucosa and intestinal cell lines. *Scandinavian Journal of Gastroenterology* 41(4):408-419.

Duci, M., Fascetti-Leon, F., Erculiani, M., Priante, E., Cavicchiolo, M. E., Verlato, G. and Gamba, P. 2018. Neonatal independent predictors of severe NEC. *Pediatric Surgery International* 34(6):663-669.

Duffy, L., Zielezny, M., Carrion, V., Griffiths, E., Dryja, D., Hilty, M., Rook, C., and Morin, F. 1997. Concordance of Bacterial Cultures with Endotoxin and Interleukin-6 in Necrotizing Enterocolitis. *Digestive Diseases and Sciences* 42(2):359-365.

Dulon, S., Candé, C., Bunnett, N. W., Hollenberg, M. D., Chignard, M. and Pidard, D. 2003. Proteinase-activated receptor-2 and human lung epithelial cells: disarming by neutrophil serine proteinases. *American Journal of Respiratory Cell and Molecular Biology* 28(3):339-346.

Dulon, S., Leduc, D., Cottrell, G. S., D'Alayer, J., Hansen, K. K., Bunnett, N. W., Hollenberg, M.D., Pidard, D., and Chignard, M. 2005. Pseudomonas aeruginosa elastase disables proteinase-activated receptor 2 in respiratory epithelial cells. *American Journal of Respiratory Cell and Molecular Biology* 32(5):411-419.

Dunlop, S. P., Hebden, J., Campbell, E., Naesdal, J., Olbe, L., Perkins, A. C. and Spiller, R. C. 2006. Abnormal intestinal permeability in subgroups of diarrhea-predominant irritable bowel syndromes. *The American Journal of Gastroenterology* 101(6):1288-1294.

Déry, O., Corvera, C. U., Steinhoff, M. and Bunnett, N. W. 1998. Proteinase-activated receptors: novel mechanisms of signaling by serine proteases. *The American Journal of Physiology* 274(6 Pt 1):C1429-1452.

Dürre, P., Spahr, R. and Andreesen, J. 1983. Glycine fermentation via a glycine reductase in Peptococcus glycinophilus and Peptococcus magnus. *Archives of Microbiology* 134(2):127-135.

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E., and Relman, D. A. 2005. Diversity of the human intestinal microbial flora. *Science* 308(5728):1635-1638.

Edelson, M. B., Bagwell, C. E. and Rozycki, H. J. 1999. Circulating Pro- and Counterinflammatory Cytokine Levels and Severity in Necrotizing Enterocolitis. *Pediatrics* 103(4):766.

Edgar, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19):2460-2461.

El Asmar, R., Panigrahi, P., Bamford, P., Berti, I., Not, T., Coppa, G. V., Catassi, C., Fasano, A., and El Asmar, R. 2002. Host-dependent zonulin secretion causes the impairment of the small intestine barrier function after bacterial exposure. *Gastroenterology* 123(5):1607-1615.

Elmi, A., Nasher, F., Jagatia, H., Gundogdu, O., Bajaj-Elliott, M., Wren, B. W. and Dorrell, N. 2015. Campylobacter jejuni outer membrane vesicle-associated proteolytic

activity promotes bacterial invasion by mediating cleavage of intestinal epithelial cell E-cadherin and occludin. *Cellular Microbiology* 18(4):561-72.

Esaiassen, E., Fjalstad, J. W., Juvet, L. K., van den Anker, J. N. and Klingenberg, C. 2017. Antibiotic exposure in neonates and early adverse outcomes: a systematic review and meta-analysis. *The Journal of Antimicrobial Chemotherapy* 72(7):1858-1870.

Fabbri, G., Panico, M., Dallolio, L., Suzzi, R., Ciccia, M., Sandri, F. and Farruggia, P. 2013. Outbreak of Ampicillin/Piperacillin-Resistant Klebsiella Pneumoniae in a Neonatal Intensive Care Unit (NICU): Investigation and Control Measures. *International Journal of Environmental Research and Public Health* 10(3):808-815.

Fagundes, C. T., Amaral, F. A., Vieira, A. T., Soares, A. C., Pinho, V., Nicoli, J. R., Vieira, L.Q., Teixeira, M.M., and Souza, D. G. 2012. Transient TLR Activation Restores Inflammatory Response and Ability To Control Pulmonary Bacterial Infection in Germfree Mice. *The Journal of Immunology* 188(3):1411.

Fasano, A., Not, T., Wang, W., Uzzau, S., Berti, I., Tommasini, A. and Goldblum, S. E. 2000. Zonulin, a newly discovered modulator of intestinal permeability, and its expression in coeliac disease. *Lancet* 355(9214):1518-1519.

Favier, C. F., Vaughan, E. E., De Vos, W. M. and Akkermans, A. D. L. 2002. Molecular monitoring of succession of bacterial communities in human neonates. *Applied and Environmental Microbiology* 68(1):219-226.

Ferranti, P., Traisci, M. V., Picariello, G., Nasi, A., Boschi, V., Siervo, M., Falconi, C., Chianese, L., and Addeo, F. 2004. Casein proteolysis in human milk: tracing the pattern of casein breakdown and the formation of potential bioactive peptides. *Journal of Dairy Research* 71(1):74-87.

Fitzgibbons, S. C., Ching, Y., Yu, D., Carpenter, J., Kenny, M., Weldon, C., Lillehei, C., Valim, C., Horbar, J.D., and Jaksic, T. 2009. Mortality of necrotizing enterocolitis expressed by birth weight categories. *Journal of Pediatric Surgery* 44(6):1072-1075; discussion 1075-1076.

Fleisch, F., Zimmermann-Baer, U., Zbinden, R., Bischoff, G., Arlettaz, R., Waldvogel, K., Nadal, D., and Christian, R. 2002. Three Consecutive Outbreaks of Serratia marcescens in a Neonatal Intensive Care Unit. *Clinical Infectious Diseases* 34(6):767-773.

Flokas, M. E., Karageorgos, S. A., Detsis, M., Alevizakos, M. and Mylonakis, E. 2017. Vancomycin-resistant enterococci colonisation, risk factors and risk for infection among

hospitalised paediatric patients: a systematic review and meta-analysis. *International Journal of Antimicrobial Agents* 49(5):565-572.

Fouhy, F., Guinane, C. M., Hussey, S., Wall, R., Ryan, C. A., Dempsey, E. M., Murphy, B., Roxx, P.R., Fitzgerald, G.F., Stanton, C., and Cotter, P. D. 2012. High-throughput sequencing reveals the incomplete, short-term recovery of infant gut microbiota following parenteral antibiotic treatment with ampicillin and gentamicin. *Antimicrobial Agents and Chemotherapy* 56(11):5811-5820.

Fox, P. F. 1981. *Proteinases in dairy technology*. pp. 233-253.

Fujimoto, T., Imaeda, H., Takahashi, K., Kasumi, E., Bamba, S., Fujiyama, Y. and Andoh, A. 2013. Decreased abundance of Faecalibacterium prausnitzii in the gut microbiota of Crohn's disease. *Journal of Gastroenterology and Hepatology* 28(4):613-619.

Gallacher, D., Mitchell, E., Alber, D., Wach, R., Klein, N., Marchesi, J.R., and Kotecha, S., 2020. Dissimilarity of the gut-lung axis and dysbiosis of the lower airways in ventilated preterm infants. *European Respiratory Journal* 0:1901909 [https://doi.org/10.1183/ 13993003.01909-2019]

Gauguet, S., D'Ortona, S., Ahnger-Pier, K., Duan, B., Surana, N. K., Lu, R., Cywes-Bentley, C., Gadjeva, M., Shan, Q., Priebe, G.P., and Pier, G. B. 2015. Intestinal Microbiota of Mice Influences Resistance to Staphylococcus aureus Pneumonia. *Infection and immunity* 83(10):4003-4014.

Gecse, K., Róka, R., Ferrier, L., Leveque, M., Eutamene, H., Cartier, C., Ait-Belgnaoui, A., Rosztoczy, A., Izbeki, F., Fiormonti, J., Wittmann, T., and Bueno, L. 2008. Increased faecal serine protease activity in diarrhoeic IBS patients: a colonic lumenal factor impairing colonic permeability and sensitivity. *Gut* 57(5):591-599.

Gerber, J. S., Bryan, M., Ross, R. K., Daymont, C., Parks, E. P., Localio, A. R., Grundmeier, R.W., Stallings, V.A., and Zaoutis, T. E. 2016. Antibiotic Exposure During the First 6 Months of Life and Weight Gain During Childhood. *JAMA* 315(12):1258-1265.

Gibson, M. K., Wang, B., Ahmadi, S., Burnham, C.-A. D., Tarr, P. I., Warner, B. B. and Dantas, G. 2016. Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome. *Nature microbiology* 7(1):16024.

Gibson, S. A., McFarlan, C., Hay, S. and MacFarlane, G. T. 1989. Significance of microflora in proteolysis in the colon. *Applied and Environmental Microbiology* 55(3):679-683.

Gill, C. J., Mantaring, J. B. V., Macleod, W. B., Mendoza, M., Mendoza, S., Huskins, W. C., Goldmann, D.A., and Hamer, D. H. 2009. Impact of Enhanced Infection Control at 2 Neonatal Intensive Care Units in The Philippines. *Clinical Infectious Diseases* 48(1):13-21.

Gill, S. R., Pop, M., DeBoy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., Gordron, J.I., Relman, D.A., Fraser-Liggett, C.M., and Nelson, K. E. 2006. Metagenomic Analysis of the Human Distal Gut Microbiome. *Science* 312(5778):1355-1359.

Giongo, A., Gano, K. A., Crabb, D. B., Mukherjee, N., Novelo, L. L., Casella, G., Drew, J.C., Ilonen, J., Knip, M., Hyoty, H., Veijolsa, R., Simell, T., Simelli, I., Neu, J., Wasserfall, C.H., Achatz, D., Atkinson, M.A., and Triplett, E. W. 2010. Toward defining the autoimmune microbiome for type 1 diabetes. *The Isme Journal* 5(1):82-91.

Giovannoni, S. J., Britschgi, T. B., Moyer, C. L. and Field, K. G. 1990. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* 345(6270):60-63.

Goldenberg, R. L., Culhane, J. F., Iams, J. D. and Romero, R. 2008. Epidemiology and causes of preterm birth. *Lancet (London, England)* 371(9606):75-84.

Golovkine, G., Faudry, E., Bouillot, S., Voulhoux, R., Attrée, I. and Huber, P. 2014. VE-cadherin cleavage by LasB protease from Pseudomonas aeruginosa facilitates type III secretion system toxicity in endothelial cells. *PLoS pathogens* 10(3):e1003939.

Gomez, A., Luckey, D. and Taneja, V. 2015. The gut microbiome in autoimmunity: Sex matters. *Clinical Immunology* 159(2):154-162.

Goodrich, Julia K., Waters, Jillian L., Poole, Angela C., Sutter, Jessica L., Koren, O., Blekhman, R., Beaumont, M., Van Treuren, W., Knight, R., Bell, J.T., Spector, T.D., Clark, A.G., and Ley, Ruth E. 2014. Human Genetics Shape the Gut Microbiome. *Cell* 159(4):789-799.

Goossens, H., Ferech, M., Vander Stichele, R., Elseviers, M. and Group, E. P. 2005. Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. *Lancet (London, England)* 365(9459):579-587.

Gosalbes, M. J., Vallès, Y., Jiménez-Hernández, N., Balle, C., Riva, P., Miravet-Verde, S., de Vries, L.E., Llop, S., Agerso, Y., Sorensen, S.J., Ballester, F., and Francino, M. P. 2016. High frequencies of antibiotic resistance genes in infants' meconium and early fecal samples. *Journal of Developmental Origins of Health and Disease* 7(1):35-44.

Gottesman, S., Wickner, S. and Maurizi, M. R. 1997. Protein quality control: triage by chaperones and proteases. *Genes & Development* 11(7):815-823.

Greenwood, C., Morrow, A. L., Lagomarcino, A. J., Altaye, M., Taft, D. H., Yu, Z., Newburg, D.S., Ward, D.V., and Schibler, K. R. 2014. Early empiric antibiotic use in preterm infants is associated with lower bacterial diversity and higher relative abundance of Enterobacter. *The Journal of Pediatrics* 165(1):23-29.

Gregory, K. E., DeForge, C. E., Natale, K. M., Phillips, M. and Van Marter, L. J. 2011. Necrotizing Enterocolitis in the Premature Infant. *Advances in neonatal care : official journal of the National Association of Neonatal Nurses* 11(3):155-166.

Grodberg, J. and Dunn, J. J. 1988. ompT encodes the Escherichia coli outer membrane protease that cleaves T7 RNA polymerase during purification. *Journal of Bacteriology* 170(3):1245-1253.

Groer, M. W., Luciano, A. A., Dishaw, L. J., Ashmeade, T. L., Miller, E. and Gilbert, J. A. 2014. Development of the preterm infant gut microbiome: a research priority. *Microbiome* 13(2):38.

Grönlund, M. M., Lehtonen, O. P., Eerola, E. and Kero, P. 1999. Fecal microflora in healthy infants born by different methods of delivery: permanent changes in intestinal flora after cesarean delivery. *Journal of Pediatric Gastroenterology and Nutrition* 28(1):19-25.

Gschwind, A., Zwick, E., Prenzel, N., Leserer, M. and Ullrich, A. 2001. Cell communication networks: epidermal growth factor receptor transactivation as the paradigm for interreceptor signal transmission. *Oncogene* 20(13):1594-1600.

Guasch, A., Coll, M., Avilés, F. X. and Huber, R. 1992. Three-dimensional structure of porcine pancreatic procarboxypeptidase A: A comparison of the A and B zymogens and their determinants for inhibition and activation. *Journal of Molecular Biology* 224(1):141-157.

Halpern, M. D. and Denning, P. W. 2015. The role of intestinal epithelial barrier function in the development of NEC. *Tissue Barriers* 3(1-2): e1000707.

Hans, N. 1999. Proteolytic enzymes, past and future. *Proceedings of the National Academy of Sciences of the United States of America* 96(20):10962-10963.

Hansen, K. K., Sherman, P. M., Cellars, L., Andrade-Gordon, P., Pan, Z., Baruch, A., Wallace, J.L., Hollenberg, M.D., and Vergnolle, N. 2005. A major role for proteolytic activity and proteinase-activated receptor-2 in the pathogenesis of infectious colitis. *Proceedings of the National Academy of Sciences of the United States of America* 102(23):8363-8368.

Haque, K. N. 2016. Necrotizing enterocolitis - Some things old and some things new: A comprehensive review. *Journal of Clinical Neonatology* 5(2):79-90.

Harding, R. and Bocking, A. D. 2001. *Fetal Growth and Development*. Cambridge University Press, p. 300.

Haro, C., Rangel-Zúñiga, O. A., Alcalá-Díaz, J. F., Gómez-Delgado, F., Pérez-Martínez, P., Delgado-Lista, J., Quintana-Navarro, G.M., Landa, B.B., Navas-Cotes, J.A., Tena-Sempere, M., Clemente, J.C., Lopez-Miranda, J., Perez-Jimenez, F., and Camargo, A. 2016. Intestinal Microbiota Is Influenced by Gender and Body Mass Index. *PLOS ONE* 11(5):e0154090.

Harris, M. C., Costarino, A. T., Sullivan, J. S., Dulkerian, S., McCawley, L., Corcoran, L., Butler, S., and Kilpatrick, L. 1994. Cytokine elevations in critically ill infants with sepsis and necrotizing enterocolitis. *The Journal of Pediatrics* 124(1):105-111.

Hascoët, J.-M., Hubert, C., Rochat, F., Legagneur, H., Gaga, S., Emady-Azar, S. and Steenhout, P. G. 2011. Effect of formula composition on the development of infant gut microbiota. *Journal of Pediatric Gastroenterology and Nutrition* 52(6):756-762.

Hays, S., Jacquot, A., Gauthier, H., Kempf, C., Beissel, A., Pidoux, O., Jumas-Black, E., Decullier, E., Lachambre, E., Beck, L., Cambione, G., Putet, G., Claris, O., and Picaud, J.-C. 2016. Probiotics and growth in preterm infants: A randomized controlled trial, PREMAPRO study. *Clinical Nutrition* 35(4):802-811.

Hering, N. A., Luettig, J., Krug, S. M., Wiegand, S., Gross, G., van Tol, E. A., Schulzke, J.D., and Rosenthal, R. 2017. Lactoferrin protects against intestinal inflammation and bacteria-induced barrier dysfunction in vitro. *Annals of the New York Academy of Sciences* 1405(1):177-188.

Hermon-Taylor, J., Perrin, J., Grant, D. A., Appleyard, A., Bubel, M. and Magee, A. I. 1977. Immunofluorescent localisation of enterokinase in human small intestine. *Gut* 18(4):259-265.

Herz, J. and Strickland, D. K. 2001. LRP: a multifunctional scavenger and signaling receptor. *The Journal of clinical investigation* 108(6):779-784.

Hevia, A., Milani, C., López, P., Donado, C. D., Cuervo, A., González, S., Suarez, A., Turroni, F., Gueimonde, M., Ventura, M., Sanchez, B., and Margolles, A. 2016. Allergic Patients with Long-Term Asthma Display Low Levels of Bifidobacterium adolescentis. *PloS one* 11(2):e0147809-e0147809.

Hewitt, K. M., Mannino, F. L., Gonzalez, A., Chase, J. H., Caporaso, J. G., Knight, R. and Kelley, S. T. 2013. Bacterial Diversity in Two Neonatal Intensive Care Units (NICUs). *PLOS ONE* 8(1):e54703.

Heyndrickx, G. V. 1963. Further investigations on the enzymes in human milk. *Pediatrics* 31:1019-1023.

Hill, T. C. J., Walsh, K. A., Harris, J. A. and Moffett, B. F. 2003. Using ecological diversity measures with bacterial communities. *FEMS Microbiology Ecology* 43(1):1-11.

Holland, J. L., Louie, L., Simor, A. E. and Louie, M. 2000. PCR detection of Escherichia coli O157:H7 directly from stools: evaluation of commercial extraction methods for purifying fecal DNA. *Journal of Clinical Microbiology* 38(11):4108-4113.

Hollenberg, M. D. and Compton, S. J. 2002. International Union of Pharmacology. XXVIII. Proteinase-activated receptors. *Pharmacological Reviews* 54(2):203-217.

Holman, R. C., Stoll, B. J., Curns, A. T., Yorita, K. L., Steiner, C. A. and Schonberger, L. B. 2006. Necrotising enterocolitis hospitalisations among neonates in the United States. *Paediatric and Perinatal Epidemiology* 20(6):498-506.

Hooper, J. D., Clements, J. A., Quigley, J. P. and Antalis, T. M. 2001. Type II transmembrane serine proteases. Insights into an emerging class of cell surface proteolytic enzymes. *The Journal of Biological Chemistry* 276(2):857-860.

Horbar, J. D., Carpenter, J. H., Badger, G. J., Kenny, M. J., Soll, R. F., Morrow, K. A. and Buzas, J. S. 2012. Mortality and neonatal morbidity among infants 501 to 1500 grams from 2000 to 2009. *Pediatrics* 129(6):1019-1026.

Hoy, B., Löwer, M., Weydig, C., Carra, G., Tegtmeyer, N., Geppert, T., Schroder, P., Sewald, N., Backert, S., Schneider, G., and Wessler, S. 2010. Helicobacter pylori HtrA is a new secreted virulence factor that cleaves E-cadherin to disrupt intercellular adhesion. *EMBO reports* 11(10):798-804.

Hritonenko, V. and Stathopoulos, C. 2007. Omptin proteins: an expanding family of outer membrane proteases in Gram-negative Enterobacteriaceae. *Molecular Membrane Biology* 24(5-6):395-406.

Hsieh, E. M., Hornik, C. P., Clark, R. H., Laughon, M. M., Benjamin, D. K., Smith, P. B. and Best Pharmaceuticals for Children Act—Pediatric Trials, N. 2014. Medication use in the neonatal intensive care unit. *American Journal of Perinatology* 31(9):811-821.

Hua, X., Goedert, J. J., Pu, A., Yu, G. and Shi, J. 2015. Allergy associations with the adult fecal microbiota: Analysis of the American Gut Project. *EBioMedicine* 3:172-179.

Huddy, C. L. J., Johnson, A. and Hope, P. L. 2001. Educational and behavioural problems in babies of 32–35 weeks gestation. *Archives of Disease in Childhood - Fetal and Neonatal Edition* 85(1):F23-F28.

Huesgen, P. F., Schuhmann, H. and Adamska, I. 2009. Deg/HtrA proteases as components of a network for photosystem II quality control in chloroplasts and cyanobacteria. *Research in Microbiology* 160(9):726-732.

Hunter, C. J., Podd, B., Ford, H. R. and Camerini, V. 2008a. Evidence vs experience in neonatal practices in necrotizing enterocolitis. *Journal of Perinatology: Official Journal of the California Perinatal Association* 28 Suppl 1:S9-S13.

Hunter, C. J., Singamsetty, V., Chokshi, N., Boyle, P., Camerini, V., Grishin, A., Upperman, J, Ford, H., and Prasadarao, N. 2008b. Enterobacter sakazakii enhances epithelial cell injury by inducing apoptosis in a rat model of necrotizing enterocolitis. *Journal Of Infectious Diseases* 198(4):586-593.

Huse, S. M., Ye, Y., Zhou, Y. and Fodor, A. A. 2012. A Core Human Microbiome as Viewed through 16S rRNA Sequence Clusters. *PLOS ONE* 7(6):e34242.

Hussey, S., Wall, R., Gruffman, E., O'Sullivan, L., Ryan, C. A., Murphy, B., Fitzgerald, G., Stanton, C., and Ross, R. P. 2011. Parenteral Antibiotics Reduce Bifidobacteria Colonization and Diversity in Neonates. *International Journal of Microbiology* pii: 130574. doi: 10.1155/2011/130574.

Huston, W. M. 2010. Bacterial proteases from the intracellular vacuole niche; protease conservation and adaptation for pathogenic advantage. *FEMS immunology and medical microbiology* 59(1):1-10.

Ichinohe, T., Pang, I. K., Kumamoto, Y., Peaper, D. R., Ho, J. H., Murray, T. S. and Iwasaki, A. 2011. Microbiota regulates immune defense against respiratory tract influenza A virus infection. *Proceedings of the National Academy of Sciences of the United States of America* 108(13):5354-5359.

Ingmer, H. and Brøndsted, L. 2009. Proteases in bacterial pathogenesis. *Research in Microbiology* 160(9):704-710.

Jacob, C., Yang, P.-C., Darmoul, D., Amadesi, S., Saito, T., Cottrell, G. S., Coelho, A-M., Singh, P., Grady, E.F., Perdue, M., and Bunnett, N. W. 2005. Mast cell tryptase controls paracellular permeability of the intestine. Role of protease-activated receptor 2 and beta-arrestins. *The Journal of Biological Chemistry* 280(36):31936-31948.

Jacquot, A., Neveu, D., Aujoulat, F., Mercier, G., Marchandin, H., Jumas-Bilak, E. and Picaud, J.-C. 2011. Dynamics and clinical evolution of bacterial gut microflora in extremely premature patients. *The Journal of Pediatrics* 158(3):390-396.

Jakobsson, H. E., Abrahamsson, T. R., Jenmalm, M. C., Harris, K., Quince, C., Jernberg, C., Jernberg, C., Bjorksten, B., Engstranbd., and Andersson, A. F. 2014. Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by caesarean section. *Gut* 63(4):559-566.

Janvier, A., Malo, J. and Barrington, K. J. 2014. Cohort study of probiotics in a North American neonatal intensive care unit. *The Journal of Pediatrics* 164(5):980-985.

Jayasinghe, T. N., Hilton, C., Tsai, P., Apple, B., Shepherd, P., Cutfield, W. S. and O'Sullivan, J. M. 2017. Long-term stability in the gut microbiome over 46 years in the life of Billy Apple®. *Human Microbiome Journal* 5-6(Supplement C):7-10.

Jepson, M. A., Collares-Buzato, C. B., Clark, M. A., Hirst, B. H. and Simmons, N. L. 1995. Rapid disruption of epithelial barrier function by Salmonella typhimurium is associated with structural modification of intercellular junctions. *Infection and Immunity* 63(1):356-359.

Jiménez, E., Fernández, L., Marín, M. L., Martín, R., Odriozola, J. M., Nueno-Palop, C., Narbad, A., Olivares, M., Xaus, J., and Rodríguez, J. M. 2005. Isolation of commensal bacteria from umbilical cord blood of healthy neonates born by cesarean section. *Current Microbiology* 51(4):270-274.

Johnson, J. L., Jones, M. B. and Cobb, B. A. 2015. Bacterial capsular polysaccharide prevents the onset of asthma through T-cell activation. *Glycobiology* 25(4):368-375.

Jung, C., Meinzer, U., Montcuquet, N., Thachil, E., Château, D., Thiébaut, R., Roy, M., Alnanhani, Z., Berrebi, D., Dussailliant, M., Pedruzzi, E., Thenet, S., Cerf-Benussan, N., Hugot, J-P., and Barreau, F. 2012. Yersinia pseudotuberculosis disrupts intestinal barrier integrity through hematopoietic TLR-2 signaling. *The Journal of clinical investigation* 122(6):2239-2251.

Kalischuk, L. D., Inglis, G. D. and Buret, A. G. 2009. Campylobacter jejuni induces transcellular translocation of commensal bacteria via lipid rafts. *Gut Pathogens* 1(1):2.

Kang, S., Denman, S. E., Morrison, M., Yu, Z., Dore, J., Leclerc, M. and McSweeney, C. S. 2010. Dysbiosis of fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray. *Inflammatory Bowel Diseases* 16(12):2034-2042.

Karunasena, E., McMahon, K. W., Chang, D. and Brashears, M. M. 2014. Host Responses to the Pathogen Mycobacterium avium subsp. paratuberculosis and Beneficial Microbes Exhibit Host Sex Specificity. *Applied and Environmental Microbiology* 80(15):4481-4490.

Katz, J., Sambandam, V., Wu, J. H., Michalek, S. M. and Balkovetz, D. F. 2000. Characterization of Porphyromonas gingivalis-induced degradation of epithelial cell junctional complexes. *Infection and Immunity* 68(3):1441-1449.

Kawabata, A. 2003. Gastrointestinal functions of proteinase-activated receptors. *Life Sciences* 74(2-3):247-254.

Kawahara, T., Takahashi, T., Oishi, K., Tanaka, H., Masuda, M., Takahashi, S., Takano, M., Kawakami, T., Fukushima, K., Kanazawa, H., and Suzuki, T. 2015. Consecutive oral administration of Bifidobacterium longum MM-2 improves the defense system against influenza virus infection by enhancing natural killer cell activity in a murine model. *Microbiology and Immunology* 59(1):1-12.

Keely, S., Talley, N. J. and Hansbro, P. M. 2012. Pulmonary-intestinal cross-talk in mucosal inflammatory disease. *Mucosal immunology* 5(1):7-18.

Kelly, B. J., Imai, I., Bittinger, K., Laughlin, A., Fuchs, B. D., Bushman, F. D. and Collman, R. G. 2016. Composition and dynamics of the respiratory tract microbiome in intubated patients. *Microbiome* 4(1):7.

Kim, Y. S., Birtwhistle, W. and Kim, Y. W. 1972. Peptide hydrolases in the bruch border and soluble fractions of small intestinal mucosa of rat and man. *The Journal of clinical investigation* 51(6):1419-1430.

Klappenbach, J. A., Saxman, P. R., Cole, J. R. and Schmidt, T. M. 2001. rrndb: the Ribosomal RNA Operon Copy Number Database. *Nucleic Acids Research* 29(1):181-184.

Klein, C. J. 2002. Nutrient requirements for preterm infant formulas. *Journal Of Nutrition* 132(6):1395S-1577S.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M. and Glöckner, F. O. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research* 41(1):e1.

Klymiuk, I., Bilgilier, C., Stadlmann, A., Thannesberger, J., Kastner, M.-T., Högenauer, C., Puspok, A., Biowski-Frotz, S., Schrutka-Kolbl, C., Thallinger, G.G., and Steininger, C. 2017. The Human Gastric Microbiome Is Predicated upon Infection with Helicobacter pylori. *Frontiers in Microbiology* 14(8):2508.

Knight, J. M., Davidson, L. A., Herman, D., Martin, C. R., Goldsby, J. S., Ivanov, I. V., Donovan, S.M., and Chapkin, R. S. 2014. Non-invasive analysis of intestinal development in preterm and term infants using RNA-Sequencing. *Scientific Reports* 4:5453.

Koch, K. N., Hartung, M. L., Urban, S., Kyburz, A., Bahlmann, A. S., Lind, J., Backert, S., Taube, C., and Müller, A. 2015. Helicobacter urease-induced activation of the TLR2/NLRP3/IL-18 axis protects against asthma. *The Journal of clinical investigation* 125(8):3297-3302.

Koenig, J. E., Spor, A., Scalfone, N., Fricker, A. D., Stombaugh, J., Knight, R., Angenet, L.T., and Ley, R. E. 2011a. Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences of the United States of America* 108 Suppl 1:4578-4585.

Koenig, J. E., Spor, A., Scalfone, N., Fricker, A. D., Stombaugh, J., Knight, R., Angenet, L.T., and Ley, R. E. (2011b). Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences* 108(Supplement 1):4578-4585.

Kolacek, S., Puntis, J. W., Lloyd, D. R., Brown, G. A. and Booth, I. W. 1990. Ontogeny of pancreatic exocrine function. *Archives of Disease in Childhood* 65(2):178.

Kollef, M. H. 2008. Update on the Appropriate Use of Meropenem for the Treatment of Serious Bacterial Infections. *Clinical Infectious Diseases* 47(Supplement_1):S1-S2.

Koren, O., Goodrich, Julia K., Cullender, Tyler C., Spor, A., Laitinen, K., Kling Bäckhed, H., Gonzalez, A., Werner, J.J., Angenet, L.T., Knight, R., Backhead, F., Isolauri, E., Salimen, S., and Ley, Ruth E. 2012. Host Remodeling of the Gut Microbiome and Metabolic Changes during Pregnancy. *Cell* 150(3):470-480.

Korpela, K., Blakstad, E. W., Moltu, S. J., Strømmen, K., Nakstad, B., Rønnestad, A. E., Baekke, K., Iversen, P.O., Drevon, C.A., and Vos, W. 2018. Intestinal microbiota development and gestational age in preterm neonates. *Scientific Reports* 8(1):2453.

Korpela, K., Salonen, A., Virta, L. J., Kekkonen, R. A., Forslund, K., Bork, P., and de Vos, W. M. 2016. Intestinal microbiome is related to lifetime antibiotic use in Finnish pre-school children. *Nature Communications* 7:10410.

Kosloske, A. M. 1994. Epidemiology of necrotizing enterocolitis. *Acta Paediatrica (Oslo, Norway: 1992). Supplement* 396:2-7.

Kramer, R. A., Brandenburg, K., Vandeputte-Rutten, L., Werkhoven, M., Gros, P., Dekker, N. and Egmond, M. R. 2002. Lipopolysaccharide regions involved in the activation of Escherichia coli outer membrane protease OmpT. *European Journal of Biochemistry* 269(6):1746-1752.

Kukkonen, M. and Korhonen, T. K. 2004. The omptin family of enterobacterial surface proteases/adhesins: from housekeeping in Escherichia coli to systemic spread of Yersinia pestis. *International journal of medical microbiology: IJMM* 294(1):7-14.

Kuppala, V. S., Meinzen-Derr, J., Morrow, A. L. and Schibler, K. R. 2011. Prolonged initial empirical antibiotic treatment is associated with adverse outcomes in premature infants. *The Journal of Pediatrics* 159(5):720-725.

La Rosa, P. S., Warner, B. B., Zhou, Y., Weinstock, G. M., Sodergren, E., Hall-Moore, C. M., Stevens, H.J., Bennett, W.E., Shaikh, N., Linneman, L.A., Hoffmann, J.A., Hasmvas, A., Deych, E., Shands, B.A., Shannon, W.D., and Tarr, P. I. 2014. Patterned progression of bacterial populations in the premature infant gut. *Proceedings of the National Academy of Sciences of the United States of America* 111(34):12522-12527.

Laskowska, E., Kuczyńska-Wiśnik, D., Skórko-Glonek, J. and Taylor, A. 1996. Degradation by proteases Lon, Clp and HtrA, of Escherichia coli proteins aggregated in vivo by heat shock; HtrA protease action in vivo and in vitro. *Molecular Microbiology* 22(3):555-571.

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

Lawn, J. E., Cousens, S., Zupan, J. and Lancet Neonatal Survival Steering, T. 2005. 4 million neonatal deaths: when? Where? Why? *Lancet* 365(9462):891-900.

Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M., Arumungam, M., Batto, J-M., Kennedy, S., Leonard, P., Li, J., Burgdorf, K., Grarup, N., Jorgensen, T., Brandslund, N., brunak, S., Rasmussen, S., Sunagawa, S., Tap, J., Tims, S., Zoetendeal, E.G., Clement, K., Dore, J., Kleerebezem, M., Kristiansen, K., Remualt, P., Sicheritz-Ponten, T., de Vos, W.M., Zucker, J-D., Raes, J., Hansen, T., Meta HIT Concortium, Bork, P., Wang, J., Ehrlich, S.D., and Pedersen, O. (2013). Richness of human gut microbiome correlates with metabolic markers. *Nature* 500(7464):541-546.

Lebenthal, E. and Lee, P. C. 1980. Development of functional response in human exocrine pancreas. *Pediatrics* 66(4):556-560.

Lee, E., Kim, B. J., Kang, M. J., Choi, K. Y., Cho, H. J., Kim, Y., Yang Song, I., Jung, Y.H., Kim, H.Y., Seo, J.H., Kwon, J.W., Kim, H.B., Lee, S.Y., and Hong, S. J. 2016. Dynamics of Gut Microbiota According to the Delivery Mode in Healthy Korean Infants. *Allergy, Asthma & Immunology Research* 8(5):471-477.

Leiros, H. K. S. 2004. Trypsin specificity as elucidated by LIE calculations, X-ray structures, and association constant measurements. *Protein Science* 13(4):1056-1070.

Li, J., Jia, H., Cai, X., Zhong, H., Feng, Q., Sunagawa, S., Manimozhiyan, A., Jens, R.K., Edi, P., Nielsen, T., Juncker, A.S., Manichanh, C., Chen, B., Zhang, W., Levenez, F., Wang, J., Xu, X., Xiao, L., Lianmg, S., Zhang, D., Zhang, Z., Chen, W., Zhao, H., Al-Aama, J.Y., Edris, S., Yang, H., Wang, J., Hansen, T., Nielsen, H.B., Brunak, S., Kristiansen, K., Guarner, F., Pedersen, O., Dore, J., Ehrlich, S.D., MetaHIT Consortium, Bork, P., and Wang, J. 2014. An integrated catalog of reference genes in the human gut microbiome. *Nature Biotechnology* 32:834-841.

Li, K., Bihan, M. and Methé, B. A. 2013. Analyses of the Stability and Core Taxonomic Memberships of the Human Microbiome. *PLOS ONE* 8(5):e63139.

Li, M., Gong, J., Cottrill, M., Yu, H., de Lange, C., Burton, J. and Topp, E. 2003. Evaluation of QIAamp® DNA Stool Mini Kit for ecological studies of gut microbiota. *Journal of Microbiological Methods* 54(1):13-20.

Liem, T. B. Y., Krediet, T. G., Fleer, A., Egberts, T. C. G. and Rademaker, C. M. A. 2010. Variation in antibiotic use in neonatal intensive care units in the Netherlands. *The Journal of Antimicrobial Chemotherapy* 65(6):1270-1275.

Liepke, C., Adermann, K., Raida, M., Mägert, H. J., Forssmann, W. G. and Zucht, H. D. 2002. Human milk provides peptides highly stimulating the growth of bifidobacteria. *European Journal of Biochemistry* 269(2):712-718.

Liu, C. M., Aziz, M., Kachur, S., Hsueh, P.-R., Huang, Y.-T., Keim, P. and Price, L. B. 2012. BactQuant: an enhanced broad-coverage bacterial quantitative real-time PCR assay. *BMC microbiology* 12:56.

Lloyd-Price, J., Abu-Ali, G. and Huttenhower, C. 2016. The healthy human microbiome. *Genome Medicine* 8(1):51.

Lorey, S., Faust, J., Mrestani-Klaus, C., Kähne, T., Ansorge, S., Neubert, K. and Bühling, F. 2002. Transcellular proteolysis demonstrated by novel cell surface-associated substrates of dipeptidyl peptidase IV (CD26). *The Journal of Biological Chemistry* 277(36):33170-33177.

Lu, Q., Cheng, S., Zhou, M. and Yu, J. 2017. Risk Factors for Necrotizing Enterocolitis in Neonates: A Retrospective Case-Control Study. *Pediatrics and Neonatology* 58(2):165-170.

Lubetzky, R., Mandel, D., Mimouni, F. B., Herman, L., Reich, R. and Reif, S. 2010. MMP-2 and MMP-9 and their tissue inhibitor in preterm human milk. *Journal of Pediatric Gastroenterology and Nutrition* 51(2):210-212.

Lucas, A. and Cole, T. J. 1990. Breast milk and neonatal necrotising enterocolitis. *Lancet (London, England)* 336(8730):1519-1523.

Lusyati, S., Hulzebos, C., Zandvoort, J., Sukandar, H. and Sauer, P. J. J. 2013. Cytokines patterns in newborn infants with late onset sepsis. *Journal of neonatal perinatal medicine* 6(2):153-163.

MacCallum, A., Hardy, S. P. and Everest, P. H. 2005. Campylobacter jejuni inhibits the absorptive transport functions of Caco-2 cells and disrupts cellular tight junctions. *Microbiology* 151(Pt 7):2451-2458.

Macfarlane, G. and Allison, C. 1986. Utilisation of protein by human gut bacteria. *FEMS Microbiology Ecology* 2(1):19-24.

Macfarlane, G. T., Allison, C., Gibson, S. A. and Cummings, J. H. 1988. Contribution of the microflora to proteolysis in the human large intestine. *The Journal of Applied Bacteriology* 64(1):37-46.

Macfarlane, R. G. 1964. An Enzyme Cascade in the Blood Clotting Mechanism, and its Function as a Biochemical Amplifier. *Nature* 202(4931):498-499.

Mackie, R. I., Sghir, A. and Gaskins, H. R. 1999. Developmental microbial ecology of the neonatal gastrointestinal tract. *The American Journal of Clinical Nutrition* 69(5):1035S-1045S.

MacNaughton, W. K. 2005. Epithelial effects of proteinase-activated receptors in the gastrointestinal tract. *Memorias Do Instituto Oswaldo Cruz* 100 Suppl 1:211-215.

Madan, J. C., Salari, R. C., Saxena, D., Davidson, L., O'Toole, G. A., Moore, J. H., Sogiun, M.J., Foster, J.A, Edwards, W.H., Palumbo, P., and Hibberd, P. L. 2012. Gut microbial colonisation in premature neonates predicts neonatal sepsis. *Archives of Disease in Childhood. Fetal and Neonatal Edition* 97(6):F456-462.

Magne, F., Abély, M., Boyer, F., Morville, P., Pochart, P. and Suau, A. 2006. Low species diversity and high interindividual variability in faeces of preterm infants as revealed by sequences of 16S rRNA genes and PCR-temporal temperature gradient gel electrophoresis profiles. *FEMS microbiology ecology* 57(1):128-138.

Maheshwari, A., Lu, W., Lacson, A., Barleycorn, A. A., Nolan, S., Christensen, R. D. and Calhoun, D. A. 2002. Effects of interleukin-8 on the developing human intestine. *Cytokine* 20(6):256-267.

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

Maidak, B. L., Olsen, G. J., Larsen, N., Overbeek, R., McCaughey, M. J. and Woese, C. R. 1997. The RDP (Ribosomal Database Project). *Nucleic Acids Research* 25(1):109-111.

Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., Nalin, R., Jarrin, C., Chardon, P., Marteau, P., Roca, J., and Dore, J. 2006. Reduced

diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* 55(2):205-211.

Mankertz, J., Tavalali, S., Schmitz, H., Mankertz, A., Riecken, E. O., Fromm, M. and Schulzke, J. D. 2000. Expression from the human occludin promoter is affected by tumor necrosis factor alpha and interferon gamma. *Journal of Cell Science* 113 (Pt 11):2085-2090.

Marchesi, J. R. 2011. Human distal gut microbiome. *Environmental Microbiology* 13(12):3088-3102.

Marchesi, J. R., Adams, D. H., Fava, F., Hermes, G. D. A., Hirschfield, G. M., Hold, G., Quiraishi, M.N., Kinross, J., Smidt, H., Tuohy, K.M., Thomas, L.V., Zoetendal, E.G., and Hart, A. 2015. The gut microbiota and host health: a new clinical frontier. *Gut* 65(2):330-339.

Marchesi, J. R. and Ravel, J. 2015. The vocabulary of microbiome research: a proposal. *Microbiome* 3:31.

Marsland, B. J., Trompette, A. and Gollwitzer, E. S. 2015. The Gut–Lung Axis in Respiratory Disease. *Annals of the American Thoracic Society* 12(Supplement 2):S150-S156.

Martin, C. R. and Walker, W. A. 2006. Intestinal immune defences and the inflammatory response in necrotising enterocolitis. *Seminars in Fetal & Neonatal Medicine* 11(5):369-377.

Martin, R., Makino, H., Cetinyurek Yavuz, A., Ben-Amor, K., Roelofs, M., Ishikawa, Kubota, H., Swinkels, S., Sakai, T., Oishi, K., Kushiro, A., and E., Knol, J. 2016. Early-Life Events, Including Mode of Delivery and Type of Feeding, Siblings and Gender, Shape the Developing Gut Microbiota. *PLOS ONE* 11(6):e0158498.

Mazzola, G., Murphy, K., Ross, R. P., Di Gioia, D., Biavati, B., Corvaglia, L. T., Faldella, G., and Stanton, C. 2016. Early Gut Microbiota Perturbations Following Intrapartum Antibiotic Prophylaxis to Prevent Group B Streptococcal Disease. *PloS One* 11(6):e0157527.

McGuire, W. and Anthony, M. Y. 2003. Donor human milk versus formula for preventing necrotising enterocolitis in preterm infants: systematic review. 88(1):F11-F14.

McMurdie, P. J. and Holmes, S. 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE* 8(4):e61217.

McMurdie, P. J. and Holmes, S. 2014. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Comput Biol* 10(4):e1003531.

McOrist, A. L., Jackson, M. and Bird, A. R. 2002. A comparison of five methods for extraction of bacterial DNA from human faecal samples. *Journal of Microbiological Methods* 50(2):131-139.

Medina, C. and Radomski, M. W. 2006. Role of Matrix Metalloproteinases in Intestinal Inflammation. *Journal of Pharmacology and Experimental Therapeutics* 318(3):933-938.

Meinzen-Derr, J., Poindexter, B., Wrage, L., Morrow, A. L., Stoll, B. and Donovan, E. F. 2009. Role of human milk in extremely low birth weight infants' risk of necrotizing enterocolitis or death. *Journal of Perinatology* 29(1):57-62.

Meinzer, U., Barreau, F., Esmiol-Welterlin, S., Jung, C., Villard, C., Léger, T., Ben-Mkaddem, S., Berrebi, D., Dussaillant, M., Alnabhani, Z., Roy, M., Bonacorsi, S., Wolf-Watz, H., Perroy, J., Ollendorf, V., and Hugot, J.-P. 2012. Yersinia pseudotuberculosis Effector YopJ Subverts the Nod2/RICK/TAK1 Pathway and Activates Caspase-1 to Induce Intestinal Barrier Dysfunction. *Cell Host & Microbe* 11(4):337-351.

Melville, J. M. and Moss, T. J. M. 2013. The immune consequences of preterm birth. *Frontiers in Neuroscience* 7:79.

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

Milner, J. M., Patel, A. and Rowan, A. D. 2008. Emerging roles of serine proteinases in tissue turnover in arthritis. *Arthritis and Rheumatism* 58(12):3644-3656.

Mirsepasi, H., Persson, S., Struve, C., Andersen, L. O. B., Petersen, A. M. and Krogfelt, K. A. 2014. Microbial diversity in fecal samples depends on DNA extraction method: easyMag DNA extraction compared to QIAamp DNA stool mini kit extraction. *BMC Research Notes* 7(1):50.

Moerch, E., Westerbeek, E. A. M., Knol, F. R., Kok, A., Fetter, W. P. F., Lafeber, H. N. and vanElburg, R. M. 2008. Calprotectin and il-8: markers of intestinal inflammation in (pre)term infants? *Archives of Disease in Childhood* 93:ps314.

Moles, L., Gómez, M., Heilig, H., Bustos, G., Fuentes, S., de Vos, W., Fernandez, L., Rodriguez, J.M., and Jiménez, E. 2013. Bacterial Diversity in Meconium of Preterm Neonates and Evolution of Their Fecal Microbiota during the First Month of Life. *PLOS ONE* 8(6):e66986.

Molodecky, N. A., Soon, I. S., Rabi, D. M., Ghali, W. A., Ferris, M., Chernoff, G., Benchimol, E.I., Panaccione, R., Ghosh, S., Barkema, H.W., and Kaplan, G. G. 2012. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology* 142(1):46-54.

Momose, Y., Hirayama, K. and Itoh, K. 2008. Competition for proline between indigenous Escherichia coli and E. coli O157:H7 in gnotobiotic mice associated with infant intestinal microbiota and its contribution to the colonization resistance against E. coli O157:H7. *Antonie van Leeuwenhoek* 94(2):165-171.

Morecroft, J. A., Spitz, L., Hamilton, P. A. and Holmes, S. J. K. 1994. Plasma cytokine levels in necrotizing enterocolitis. *Acta Paediatrica* 83(s396):18-20.

Morgan, X. C., Tickle, T. L., Sokol, H., Gevers, D., Devaney, K. L., Ward, D. V., Reyes, J.A., Shah, S.A., LeLeiko, N., Snapper, S.B., Bousvaros, A., Korzenik, J., Sands, B.E., Xaveir, R.J., and Huttenhower, C. 2012. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biology* 13(9):R79.

Morris, A., Beck, J. M., Schloss, P. D., Campbell, T. B., Crothers, K., Curtis, J. L., Flores, S.C., Fontenot, A.P., Gedin, E., Huang, L., Jablonski, K., Kleerup, E., Lynch, S.V., Sodergen, E., Twigg, H., Young, V.B., Bassis, C.M., Venkstarman, A., Schmidt, T.M., Weinstock, G.M., Lung, H. I. V. Microbiome Project. 2013. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *American journal of respiratory and critical care medicine* 187(10):1067-1075.

Morris, L. S., Evans, J. and Marchesi, J. R. 2012. A robust plate assay for detection of extracellular microbial protease activity in metagenomic screens and pure cultures. *Journal of Microbiological Methods* 91(1):144-146.

Mounier, J., Vasselon, T., Hellio, R., Lesourd, M. and Sansonetti, P. J. 1992. Shigella flexneri enters human colonic Caco-2 epithelial cells through the basolateral pole. *Infection and Immunity* 60(1):237-248.

Mozaffarian, D., Hao, T., Rimm, E. B., Willett, W. C. and Hu, F. B. 2011. Changes in Diet and Lifestyle and Long-Term Weight Gain in Women and Men. *New England Journal of Medicine* 364(25):2392-2404.

Mueller, N. T., Whyatt, R., Hoepner, L., Oberfield, S., Dominguez-Bello, M. G., Widen, E. M., Hassouen, A., Perera, F., and Rundle, A. 2015. Prenatal exposure to antibiotics, cesarean section and risk of childhood obesity. *International journal of obesity* 39(4):665-670.

Mulder, D. J., Noble, A. J., Justinich, C. J. and Duffin, J. M. 2014. A tale of two diseases: The history of inflammatory bowel disease. *Journal of Crohn's and Colitis* 8(5):341-348.

Műzes, G., Molnár, B., Tulassay, Z. and Sipos, F. 2012. Changes of the cytokine profile in inflammatory bowel diseases. *World journal of gastroenterology* 18(41):5848-5861.

Nakanishi-Matsui, M., Zheng, Y. W., Sulciner, D. J., Weiss, E. J., Ludeman, M. J. and Coughlin, S. R. (2000. PAR3 is a cofactor for PAR4 activation by thrombin. *Nature* 404(6778):609-613.

Nanninga, H. J., Drent, W. J. and Gottschal, J. C. 1986. Major differences between glutamate-fermenting species isolated from chemostat enrichments at different dilution rates. *FEMS microbiology letters* 38(6):321-329.

Nanthakumar, N. N., Fusunyan, R. D., Sanderson, I. and Walker, W. A. 2000. Inflammation in the developing human intestine: A possible pathophysiologic contribution to necrotizing enterocolitis. *Proceedings of the National Academy of Sciences of the United States of America* 97(11):6043-6048.

Nash, A. K., Auchtung, T. A., Wong, M. C., Smith, D. P., Gesell, J. R., Ross, M. C., Stewart, C.J., Metcalf, G.A., Muzny, D.M., Gibbs, R.A., Ajami, N.J., and Petrosino, J. F. 2017. The gut mycobiome of the Human Microbiome Project healthy cohort. *Microbiome* 5(1):153.

Nechvatal, J. M., Ram, J. L., Basson, M. D., Namprachan, P., Niec, S. R., Badsha, K. Z., Matherly, L.H>, Majumdar, A.P.N., and Kato, I. 2008. Fecal collection, ambient preservation, and DNA extraction for PCR amplification of bacterial and human markers from human feces. *Journal of Microbiological Methods* 72(2):124-132.

Netzel-Arnett, S., Hooper, J. D., Szabo, R., Madison, E. L., Quigley, J. P., Bugge, T. H. and Antalis, T. M. 2003. Membrane anchored serine proteases: a rapidly expanding

group of cell surface proteolytic enzymes with potential roles in cancer. *Cancer Metastasis Reviews* 22(2-3):237-258.

Netzel-Arnett, S., Mitola, D. J., Yamada, S. S., Chrysovergis, K., Holmbeck, K., Birkedal-Hansen, H. and Bugge, T. H. 2002. Collagen dissolution by keratinocytes requires cell surface plasminogen activation and matrix metalloproteinase activity. *The Journal of Biological Chemistry* 277(47):45154-45161.

Neu, J. 2007a. Gastrointestinal development and meeting the nutritional needs of premature infants. *The American Journal of Clinical Nutrition* 85(2):629S-634S.

Neu, J. 2007b. Gastrointestinal maturation and implications for infant feeding. *Early Human Development* 83(12):767-775.

Neurath, H. 1989. Proteolytic processing and physiological regulation. *Trends in Biochemical Sciences* 14(7):268-271.

Neurath, H. and Walsh, K. A. 1976. Role of proteolytic enzymes in biological regulation (a review). *Proceedings of the National Academy of Sciences of the United States of America* 73(11):3825-3832.

Nikolaev, N., Schlessinger, D. and Wellauer, P. K. 1974. 30 S pre-ribosomal RNA of Escherichia coli and products of cleavage by ribonuclease III: Length and molecular weight. *Journal of Molecular Biology* 86(4):741-748.

Nordahl Petersen, T., Rasmussen, S., Hasman, H., Carøe, C., Bælum, J., Charlotte Schultz, A., Bergmark, L., Sevendsen, C.A., Lund, O., Sicheritz-Ponten, T., and Aarestrup, F. M. 2015. Meta-genomic analysis of toilet waste from long distance flights; a step towards global surveillance of infectious diseases and antimicrobial resistance. *Scientific Reports* 5:11444.

Normann, E., Fahlén, A., Engstrand, L. and Lilja, H. E. 2013. Intestinal microbial profiles in extremely preterm infants with and without necrotizing enterocolitis. *Acta Paediatrica* 102(2):129-136.

O Cróinín, T. and Backert, S. 2012. Host epithelial cell invasion by Campylobacter jejuni: trigger or zipper mechanism? *Frontiers in Cellular and Infection Microbiology* 2:25.

Odamaki, T., Kato, K., Sugahara, H., Hashikura, N., Takahashi, S., Xiao, J.-Z., Abe, F., and Osawa, R. 2016. Age-related changes in gut microbiota composition from newborn to centenarian: a cross-sectional study. *BMC microbiology* 16(1):90.

Oh, J. Z., Ravindran, R., Chassaing, B., Carvalho, F. A., Maddur, M. S., Bower, M., Hakipour, P., Gill, K.P., Kakaya, H.I., Yarovinsky, F., Sartor, R.B., Gewirtz, A.T., and Pulendran, B. 2014. TLR5-mediated sensing of gut microbiota is necessary for antibody responses to seasonal influenza vaccination. *Immunity* 41(3):478-492.

Ohtsu, H., Dempsey, P. J. and Eguchi, S. 2006. ADAMs as mediators of EGF receptor transactivation by G protein-coupled receptors. *American Journal of Physiology. Cell Physiology* 291(1):C1-10.

Olsen, G. J., Lane, D. J., Giovannoni, S. J., Pace, N. R. and Stahl, D. A. 1986. Microbial ecology and evolution: a ribosomal RNA approach. *Annual Review of Microbiology* 40:337-365.

Olsen, R., Greisen, G., Schrøder, M. and Brok, J. 2016. Prophylactic Probiotics for Preterm Infants: A Systematic Review and Meta-Analysis of Observational Studies. *Neonatology* 109(2):105-112.

Olszak, T., An, D., Zeissig, S., Vera, M. P., Richter, J., Franke, A., Glickman, J.N., Siebert, R., Baron, R.M., Kaspar, D.L., and Blumberg, R. S. 2012. Microbial exposure during early life has persistent effects on natural killer T cell function. *Science* 336(6080):489-493.

Oppong, G. O., Rapsinski, G. J., Newman, T. N., Nishimori, J. H., Biesecker, S. G. and Tükel, Ç. 2013. Epithelial cells augment barrier function via activation of the Toll-like receptor 2/phosphatidylinositol 3-kinase pathway upon recognition of Salmonella enterica serovar Typhimurium curli fibrils in the gut. *Infection and Immunity* 81(2):478-486.

Orivuori, L., Mustonen, K., de Goffau, M. C., Hakala, S., Paasela, M., Roduit, C., Dalphin, J.C., Genuneit, J., Launener, R., Riedler, J., Weber, J., von Mutis, E., Pekkanen, J., Harmsen, H.J.M., Vaarala, O., The Pasture Study Group. 2015. High level of fecal calprotectin at age 2 months as a marker of intestinal inflammation predicts atopic dermatitis and asthma by age 6. *Clinical & Experimental Allergy* 45(5):928-939.

Ott, S. J., Musfeldt, M., Wenderoth, D. F., Hampe, J., Brant, O., Fölsch, U. R., Timmis, K., and Schreiber, S. 2004. Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. *Gut* 53(5):685-693.

Overman, E. L., Rivier, J. E. and Moeser, A. J. 2012. CRF Induces Intestinal Epithelial Barrier Injury via the Release of Mast Cell Proteases and TNF-α. *PLOS ONE* 7(6):e39935.

O'Neill, S. M., Curran, E. A., Dalman, C., Kenny, L. C., Kearney, P. M., Clarke, G., Cryan, J.G., Dinan, T.G., and Khashan, A. S. 2016. Birth by Caesarean Section and the Risk of Adult Psychosis: A Population-Based Cohort Study. *Schizophrenia Bulletin* 42(3):633-641.

O'Sullivan, A., Farver, M. and Smilowitz, J. T. 2015. The Influence of Early Infant-Feeding Practices on the Intestinal Microbiome and Body Composition in Infants. *Nutrition and Metabolic Insights* 8(Suppl 1):1-9.

Page, M. J. and Di Cera, E. 2008. Evolution of peptidase diversity. *The Journal of Biological Chemistry* 283(44):30010-30014.

Pallen, M. J. and Wren, B. W. 1997. The HtrA family of serine proteases. *Molecular Microbiology* 26(2):209-221.

Palmer, D. J., Kelly, V. C., Smit, A. M., Kuy, S., Knight, C. G. and Cooper, G. J. 2006. Human colostrum: Identification of minor proteins in the aqueous phase by proteomics. *PROTEOMICS* 6(7):2208-2216.

Pammi, M., Cope, J., Tarr, P. I., Warner, B. B., Morrow, A. L., Mai, V., Gregory, K.E., Kroll, S.j., McMurty, V., Ferris, M.J., Engstrand, L., Lilja, H. E., Hollister, E.B., Versalovic, J., and Neu, J. 2017. Intestinal dysbiosis in preterm infants preceding necrotizing enterocolitis: a systematic review and meta-analysis. *Microbiome* 5(1):31.

Pammi, M., Lal, C. V., Wagner, B. D., Mourani, P. M., Lohmann, P., Luna, R. A., Sison, A., Shivanna, B., Hollister, E.B., Abman, S.H., Versalovic, J., Connett, G.J., Bhandri, V., and Ambalavanan, N. 2019. Airway Microbiome and Development of Bronchopulmonary Dysplasia in Preterm Infants: A Systematic Review. *The Journal of Pediatrics* 204:126-133.

Parks, D. H., Tyson, G. W., Hugenholtz, P. and Beiko, R. G. 2014. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30(21):3123-3124.

Patel, S. J., Oshodi, A., Prasad, P., Delamora, P., Larson, E., Zaoutis, T., Paul, D.A., and Saiman, L. 2009. Antibiotic use in neonatal intensive care units and adherence with Centers for Disease Control and Prevention 12 Step Campaign to Prevent Antimicrobial Resistance. *The Pediatric Infectious Disease Journal* 28(12):1047-1051.

Pender, S. L. F., Braegger, C., Gunther, U., Monteleone, G., Meuli, M., Schuppan, D. and Macdonald, T. T. 2003. Matrix metalloproteinases in necrotising enterocolitis. *Pediatric Research* 54(2):160-164.

Petersen, C. and Round, J. L. 2014. Defining dysbiosis and its influence on host immunity and disease. *Cellular Microbiology* 16(7):1024-1033.

Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L., Schloss, J. A., Bonazzi, V., McEwan, J.E., Wettertrand, K.A., Deal, C., Baker, C.C., Di Francesco, V., Howcroft, T. K., Karp, R.W., Lunsford, R.D., Wellington, C.R., Belachow, T., Wright, M., Gibling, C., David, H., Mills, M., Salomon, R., Mullins, C., Akolkar, B., Begg, L., Davis, C., Grandison, L., Humble, M., Khalsa, J., Little, R.A., Peavey, H., Pontzer, C., Ponoy, M., Sayre, M.H., Starke-Reed, P., Zakhari, S., Read, J., Watosn, B., and Guyer, M. 2009. The NIH Human Microbiome Project. *Genome Research* 19(12):2317-2323.

Petrou, S. 2005. The economic consequences of preterm birth during the first 10 years of life. *BJOG: an international journal of obstetrics and gynaecology* 112 Suppl 1:10-15.

Petrou, S., Mehta, Z., Hockley, C., Cook-Mozaffari, P., Henderson, J. and Goldacre, M. 2003. The impact of preterm birth on hospital inpatient admissions and costs during the first 5 years of life. *Pediatrics* 112(6 Pt 1):1290-1297.

Philpott, D. J., McKay, D. M., Mak, W., Perdue, M. H. and Sherman, P. M. 1998. Signal transduction pathways involved in enterohemorrhagic Escherichia coli-induced alterations in T84 epithelial permeability. *Infection and Immunity* 66(4):1680-1687.

Piantino, J. H., Schreiber, M. D., Alexander, K. and Hageman, J. 2013. Culture Negative Sepsis and Systemic Inflammatory Response Syndrome in Neonates. *NeoReviews* 14(6):e294-e305.

Piterina, A. V. and Pembroke, J. T. 2013. Use of PCR-DGGE Based Molecular Methods to Analyse Microbial Community Diversity and Stability during the Thermophilic Stages of an ATAD Wastewater Sludge Treatment Process as an Aid to Performance Monitoring. *ISRN Biotechnology* 2013:162645.

Polgár, L. S. 1989. *Mechanisms of protease action*. Boca Raton, Fla: CRC Press.

Pragman, A. A., Kim, H. B., Reilly, C. S., Wendt, C. and Isaacson, R. E. 2012. The lung microbiome in moderate and severe chronic obstructive pulmonary disease. *PloS one* 7(10):e47305-e47305.

Price, M. N., Dehal, P. S. and Arkin, A. P. 2009. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Molecular Biology and Evolution* 26(7):1641-1650.

Prince, A. L., Ma, J., Kannan, P. S., Alvarez, M., Gisslen, T., Harris, R. A., Sweeney, E.L., Knox, C.L., Lambers, D.S., Jobe, A.H., Chougnet, C.A., Kallapur, S. G., and Aagaard, K. M. 2016. The placental membrane microbiome is altered among subjects with spontaneous preterm birth with and without chorioamnionitis. *American Journal of Obstetrics and Gynecology* 214(5):621-627.

Pruteanu, M., Hyland, N. P., Clarke, D. J., Kiely, B. and Shanahan, F. 2011. Degradation of the extracellular matrix components by bacterial-derived metalloproteases: implications for inflammatory bowel diseases. *Inflammatory Bowel Diseases* 17(5):1189-1200.

Puente, X. S., Sánchez, L. M., Gutiérrez-Fernández, A., Velasco, G. and López-Otín, C. 2005. A genomic view of the complexity of mammalian proteolytic systems. *Biochemical Society Transactions* 33(Pt 2):331-334.

Pugin, J. and Chevrolet, J. 1991. The intestine-liver-lung axis in septic syndrome. *Schweizerische medizinische Wochenschrift* 121(42):1538-1544.

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., Daumer, C., Heinsen, F-A., Latorre, A., Barbas, C., Seifert, J., do Santos, V.M., Ott, S.J., Ferrer, M., and Moya, A. 2012. Gut microbiota disturbance during antibiotic therapy: a multi-omic approach. *Gut* 62(11):1591-1601.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Leverenz, F., Yamada, T., Mende, D.R., Li, J., Xu, J., Li, S., Li, D., Cai, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Betalan, M., Batto, J-M., Hansen, T., Le Passieler, D., Linneberg, A., Nielsen, B.H., Peletier, E., Remault, P., Sichreitz, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Dore, J., Guarner, F., Kristiansen, K., Pdersen, O., Parhill., J., Weissenbach, J., MetaHIT Consortium, Bork, P., Ehrlich, D.S., and Wang, J. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464(7285):59-65.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., and Glöckner, F. O. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 41(Database issue):D590-D596.

Ragnarsson, E. G. E., Schoultz, I., Gullberg, E., Carlsson, A. H., Tafazoli, F., Lerm, M., Magnusson, K-E., Soderholm, J.D., and Artursson, P. 2008. Yersinia pseudotuberculosis induces transcytosis of nanoparticles across human intestinal villus epithelium via invasin-dependent macropinocytosis. *Laboratory Investigation* 88(11):1215-1226.

Rajilić-Stojanović, M., Vos, D. and M, W. 2014. The first 1000 cultured species of the human gastrointestinal microbiota. *FEMS Microbiology Reviews* 38(5):996-1047.

Reddy, B. S., MacFie, J., Gatt, M., Macfarlane-Smith, L., Bitzopoulou, K. and Snelling, A. M. 2007. Commensal bacteria do translocate across the intestinal barrier in surgical patients. *Clinical Nutrition* 26(2):208-215.

Reed, B. D., Schibler, K. R., Deshmukh, H., Ambalavanan, N. and Morrow, A. L. 2018. The Impact of Maternal Antibiotics on Neonatal Disease. *The Journal of Pediatrics* 197:97-103.

Remacle, A. G., Shiryaev, S. A. and Strongin, A. Y. 2014. Distinct interactions with cellular E-cadherin of the two virulent metalloproteinases encoded by a Bacteroides fragilis pathogenicity island. *PloS One* 9(11):e113896.

Reyes, A., Haynes, M., Hanson, N., Angly, F. E., Heath, A. C., Rohwer, F. and Gordon, J. I. 2010. Viruses in the fecal microbiota of monozygotic twins and their mothers. *Nature* 466(7304):334-338.

Richard, J. S., Randall, M. G., Cutberto, G. and Armond, S. G. 1986. Enhanced Fecal Excretion of Selected Immune Factors in Very Low Birth Weight Infants Fed Fortified Human Milk. *Pediatric Research* 20(8):711.

Rogers, A. H., Gully, N. J., Pfennig, A. L. and Zilm, P. S. 1992. The breakdown and utilization of peptides by strains of Fusobacterium nucleatum. *Oral Microbiology and Immunology* 7(5):299-303.

Rougé, C., Butel, M.-J., Piloquet, H., Ferraris, L., Legrand, A., Vodovar, M., Voyer, M., de la Cochetiere, M-F., Darmaun, D., and Rozé, J.-C. 2010. Fecal calprotectin excretion in preterm infants during the neonatal period. *PloS one* 5(6):e11083-e11083.

Roussos, A., Koursarakos, P., Patsopoulos, D., Gerogianni, I. and Philippou, N. 2003. Increased prevalence of irritable bowel syndrome in patients with bronchial asthma. *Respiratory medicine* 97(1):75-79.

Roxas, J. L., Koutsouris, A., Bellmeyer, A., Tesfay, S., Royan, S., Falzari, K., Harris, A., Cheng, H., Rhee, K.J., and Hecht, G. 2010. Enterohemorrhagic E. coli alters murine intestinal epithelial tight junction protein expression and barrier function in a Shiga toxin independent manner. *Laboratory Investigation; a Journal of Technical Methods and Pathology* 90(8):1152-1168.

Rutayisire, E., Huang, K., Liu, Y. and Tao, F. 2016. The mode of delivery affects the diversity and colonization pattern of the gut microbiota during the first year of infants' life: a systematic review. *BMC gastroenterology* 16(1):86.

Rutten, E. P. A., Lenaerts, K., Buurman, W. A. and Wouters, E. F. M. 2014. Disturbed Intestinal Integrity in Patients With COPD: Effects of Activities of Daily Living. *CHEST* 145(2):245-252.

Róka, R., Demaude, J., Cenac, N., Ferrier, L., Salvador-Cartier, C., Garcia-Villar, Fioramonti, J., and R., Bueno, L. 2007a. Colonic luminal proteases activate colonocyte proteinase-activated receptor-2 and regulate paracellular permeability in mice. *Neurogastroenterology and Motility: The Official Journal of the European Gastrointestinal Motility Society* 19(1):57-65.

Róka, R., Rosztóczy, A., Leveque, M., Izbéki, F., Nagy, F., Molnár, T., Lonovics, J., Barcia-Villar, Fioramonti, J., Wittmann, T., and Bueno, L. 2007b. A pilot study of fecal serine-protease activity: a pathophysiologic factor in diarrhea-predominant irritable bowel syndrome. *Clinical Gastroenterology and Hepatology: The Official Clinical Practice Journal of the American Gastroenterological Association* 5(5):550-555.

Saari, A., Virta, L. J., Sankilampi, U., Dunkel, L. and Saxen, H. 2015. Antibiotic exposure in infancy and risk of being overweight in the first 24 months of life. *Pediatrics* 135(4):617-626.

Sabri, A., Alcott, S. G., Elouardighi, H., Pak, E., Derian, C., Andrade-Gordon, P., Kinnally, K., and Steinberg, S. F. 2003. Neutrophil Cathepsin G Promotes Detachment-induced Cardiomyocyte Apoptosis via a Protease-activated Receptor-independent Mechanism. *Journal of Biological Chemistry* 278(26):23944-23954.

Salhab, W. A., Perlman, J. M., Silver, L. and Sue Broyles, R. 2004. Necrotizing Enterocolitis and Neurodevelopmental Outcome in Extremely Low Birth Weight Infants <1000 g. *Journal Of Perinatology* 24(9):534-540.

Salonen, A., Nikkilä, J., Jalanka-Tuovinen, J., Immonen, O., Rajilić-Stojanović, M., Kekkonen, R. A., Palva, A., and de Vos, W. M. 2010. Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: Effective recovery of bacterial

and archaeal DNA using mechanical cell lysis. *Journal of Microbiological Methods* 81(2):127-134.

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., Turner, P., Parkhill, J., Loman, N.J., and Walker, A. W. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*.12(1):87.

Salyers, A. A., Gupta, A. and Wang, Y. 2004. Human intestinal bacteria as reservoirs for antibiotic resistance genes. *Trends in Microbiology* 12(9):412-416.

Sansonetti, P. J., Tran Van Nhieu, G. and Egile, C. 1999. Rupture of the intestinal epithelial barrier and mucosal invasion by Shigella flexneri. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 28(3):466-475.

Schanler, R. J., Lau, C., Hurst, N. M. and Smith, E. O. B. 2005. Randomized trial of donor human milk versus preterm formula as substitutes for mothers' own milk in the feeding of extremely premature infants. *Pediatrics* 116(2):400-406.

Schellman, J. A. and Schellman, C. G. 1997. Kaj Ulrik Linderstrøm-Lang (1896-1959). *Protein Science: A Publication of the Protein Society* 6(5):1092-1100.

Schuijt, T. J., Lankelma, J. M., Scicluna, B. P., de Sousa e Melo, F., Roelofs, J. J. T. H., de Boer, J. D., Hoogendijk, A.J., de Beer, R., de Vos, A., Belzer, C., de Vos, W.M., van der Poll, T., and Wiersinga, W. J. 2016. The gut microbiota plays a protective role in the host defence against pneumococcal pneumonia. *Gut* 65(4):575-583.

Schulman, J., Dimand, R. J., Lee, H. C., Duenas, G. V., Bennett, M. V. and Gould, J. B. 2015. Neonatal intensive care unit antibiotic use. *Pediatrics* 135(5):826-833.

Schulzke, S. M., Deshpande, G. C. and Patole, S. K. 2007. Neurodevelopmental Outcomes of Very Low-Birth-Weight Infants With Necrotizing Enterocolitis: A Systematic Review of Observational Studies. *Archives of Pediatrics & Adolescent Medicine* 161(6):583-590.

Scott, F. I., Horton, D. B., Mamtani, R., Haynes, K., Goldberg, D. S., Lee, D. Y. and Lewis, J. D. 2016. Administration of Antibiotics to Children Before Age 2 Years Increases Risk for Childhood Obesity. *Gastroenterology* 151(1):120-129.

Sears, C. L. 2000. Molecular physiology and pathophysiology of tight junctions V. assault of the tight junction by enteric pathogens. *American Journal of Physiology. Gastrointestinal and Liver Physiology* 279(6):G1129-1134.

Sears, C. L. 2001. The toxins of Bacteroides fragilis. *Toxicon: Official Journal of the International Society on Toxinology* 39(11):1737-1746.

Segal, L. N., Alekseyenko, A. V., Clemente, J. C., Kulkarni, R., Wu, B., Gao, Z., Weiden, M. D. 2013. Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. *Microbiome* 1(1):19-19.

Sehrawat, A., Sinha, S. and Saxena, A. 2015. Helicobacter pylori neutrophil-activating protein: a potential Treg modulator suppressing allergic asthma? *Frontiers in microbiology* 6:493-493.

Seksik, P., Rigottier-Gois, L., Gramet, G., Sutren, M., Pochart, P., Marteau, P., Jian, R., and Doré, J. 2003. Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. *Gut* 52(2):237-242.

Sender, R., Fuchs, S. and Milo, R. 2016. Revised estimates for the number of human and bacteria cells in the body.*PLoS Biology* 14(8):e1002533.

Seo, K., McGregor, J. A. and French, J. I. 1992. Preterm birth is associated with increased risk of maternal and neonatal infection. *Obstetrics and gynecology* 79(1):75-80.

Shah, D. and Sinn, J. K. H. 2012. Antibiotic regimens for the empirical treatment of newborn infants with necrotising enterocolitis. *The Cochrane Database of Systematic Reviews* (8):CD007448.

Sharma, N., Chaudhry, R. and Panigrahi, P. 2012. Quantitative and Qualitative Study of Intestinal Flora in Neonates. *Journal of Global Infectious Diseases* 4(4):188-192.

Shastri, P., McCarville, J., Kalmokoff, M., Brooks, S. P. J. and Green-Johnson, J. M. 2015. Sex differences in gut fermentation and immune parameters in rats fed an oligofructose-supplemented diet. *Biology of Sex Differences* 6:13.

Shi, Z., Cai, Z., Yu, J., Zhang, T., Zhao, S., Smeds, E., Zhang, D. 2012. Toll-like receptor 11 (TLR11) prevents Salmonella penetration into the murine Peyer patches. *The Journal of Biological Chemistry* 287(52):43417-43423.

Shimizu, H., Mitomo, K., Watanabe, T., Okamoto, S. and Yamamoto, K. 1990. Involvement of a NF-kappa B-like transcription factor in the activation of the interleukin-6 gene by inflammatory lymphokines. *Molecular and cellular biology* 10(2):561-568.

Shimoda, N., Fukazawa, N., Nonomura, K. and Fairchild, R. L. 2007. Cathepsin g is required for sustained inflammation and tissue injury after reperfusion of ischemic kidneys. *The American journal of pathology* 170(3):930-940.

Shulman, R. J., Eakin, M. N., Czyzewski, D. I., Jarrett, M. and Ou, C.-N. 200). Increased gastrointestinal permeability and gut inflammation in children with functional abdominal pain and irritable bowel syndrome. *The Journal of Pediatrics* 153(5):646-650.

Silverberg, M. S., Satsangi, J., Ahmad, T., Arnott, I. D., Bernstein, C. N., Brant, S. R., Caprilli, R., Colombel, J-F., Gasche, C., Geboes, K., Jewell, D.P., Karban, A., Loftus, E.V., Salvador, A., Ridell, R.H., Sachar, D.B., Schrieber, S., Steinhart, H., A., Targan, S.R., Vermeire, S., and Warren, B. F. 2005. Toward an Integrated Clinical, Molecular and Serological Classification of Inflammatory Bowel Disease: Report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. *Canadian Journal of Gastroenterology* 19(Suppl A):5A-36A.

Sim, K., Shaw, A. G., Randell, P., Cox, M. J., McClure, Z. E., Li, M.-S., Haddad, M., Langford, P.R., Cookson, W.O.C.M., Moffatt, M.F., and Kroll, J. S. 2015. Dysbiosis anticipating necrotizing enterocolitis in very premature infants. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 60(3):389-397.

Simonovic, I., Rosenberg, J., Koutsouris, A. and Hecht, G. 2000. Enteropathogenic Escherichia coli dephosphorylates and dissociates occludin from intestinal epithelial tight junctions. *Cellular Microbiology* 2(4):305-315.

Smith, E. A. and Macfarlane, G. T. 1997a. Dissimilatory Amino Acid Metabolism in Human Colonic Bacteria. *Anaerobe* 3(5):327-337.

Smith, E. A. and Macfarlane, G. T. 1997b. Formation of Phenolic and Indolic Compounds by Anaerobic Bacteria in the Human Large Intestine. *Microbial Ecology* 33(3):180-188.

Sommer, M. O. A., Dantas, G. and Church, G. M. 2009. Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* 325(5944):1128-1131.

Sonnenburg, E. D., Smits, S. A., Tikhonov, M., Higginbottom, S. K., Wingreen, N. S. and Sonnenburg, J. L. 2016. Diet-induced extinctions in the gut microbiota compound over generations. *Nature* 529(7585):212-215.

Spiess, C., Beil, A. and Ehrmann, M. 1999. A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. *Cell* 97(3):339-347.

Srinivasan, B., Kolli, A. R., Esch, M. B., Abaci, H. E., Shuler, M. L. and Hickman, J. J. 2015. TEER measurement techniques for in vitro barrier model systems. *Journal of laboratory automation* 20(2):107-126.

Stallmach, A., Chan, C. C., Ecker, K. W., Feifel, G., Herbst, H., Schuppan, D. and Zeitz, M. 2000. Comparable expression of matrix metalloproteinases 1 and 2 in pouchitis and ulcerative colitis. *Gut* 47(3):415-422.

Stams, A. and Hansen, T. 1984. Fermentation of glutamate and other compounds by Acidaminobacter hydrogenoformans gen. nov. sp. nov., an obligate anaerobe isolated from black mud. Studies with pure cultures and mixed cultures with sulfate-reducing and methanogenic bacteria. *Archives of Microbiology* 137(4):329-337.

Stange, E. F., Travis, S. P. L., Vermeire, S., Beglinger, C., Kupcinkas, L., Geboes, K., Barakauskiene, A., Villanacci, V., Von Herbay, A., Warren, B.F., Gasche, C., Tilg, H. Schreiber, S.W., Scholmerich, J., and Reinisch, W. 2006. European evidence based consensus on the diagnosis and management of Crohn's disease: definitions and diagnosis. *Gut* 55(suppl 1):i1-i15.

Stange, E. F., Travis, S. P. L., Vermeire, S., Reinisch, W., Geboes, K., Barakauskiene, Feakins, R., Flejou, J.F., Herfarth, H., Hommes, D.W., Kupcinskas, L., Lakatox, P.L., Mantzaris, G., J., Schrieber, S., Villanacci, V., Warren, B.F., European, Crohn's and Colitis Organization. 2008. European evidence-based Consensus on the diagnosis and management of ulcerative colitis: Definitions and diagnosis. *Journal of Crohn's & Colitis* 2(1):1-23.

Stearns, J. C., Simioni, J., Gunn, E., McDonald, H., Holloway, A. C., Thabane, L., Mousseau, A., Schertzer, J.D., Ratcliffe, E.M., Rossi, L., Surette, M.G., Morrison, K.M., and Hutton, E. K. 2017. Intrapartum antibiotics for GBS prophylaxis alter colonization patterns in the early infant gut microbiome of low risk infants. *Scientific Reports* 7(1):16527.

Steck, N., Mueller, K., Schemann, M. and Haller, D. 2013. Republished: bacterial proteases in IBD and IBS. *Postgraduate Medical Journal* 89(1047):25-33.

Steinhoff, M., Buddenkotte, J., Shpacovitch, V., Rattenholl, A., Moormann, C., Vergnolle, N., Luger, T.A., and Hollenberg, M. D. 2005. Proteinase-activated receptors: transducers of proteinase-mediated signaling in inflammation and immune response. *Endocrine Reviews* 26(1):1-43.

Stewart, C. J., Embleton, N. D., Clements, E., Luna, P. N., Smith, D. P., Fofanova, T. Y., Nelson, A., Taylor, G., Orr, C.H., Petrosino, J.F., Berrington, J.E., and Cummings, S. P. 2017. Cesarean or Vaginal Birth Does Not Impact the Longitudinal Development of the Gut Microbiome in a Cohort of Exclusively Preterm Infants. *Frontiers in Microbiology* 8:1008.

Stewart, C. J., Marrs, E. C. L., Magorrian, S., Nelson, A., Lanyon, C., Perry, J. D., Embleton, N.D., Cummings, S.P., and Berrington, J. E. 2012. The preterm gut microbiota: changes associated with necrotizing enterocolitis and infection. *Acta Paediatrica* 101(11):1121-1127.

Stoll, B. J., Hansen, N., Fanaroff, A. A., Wright, L. L., Carlo, W. A., Ehrenkranz, R. A., Lemons, J.A., Donovan, E.F., Stark, A.R., Tyson, J.E., Oh, W., Bauer, C.R., Korones, S.B., Shankaran, S., Laptook, A.R., Stevenson, D.K., Papile, L-A., and Poole, W. K. 2002. Late-onset sepsis in very low birth weight neonates: the experience of the NICHD Neonatal Research Network. *Pediatrics* 110(2 Pt 1):285-291.

Stoll, B. J., Hansen, N. I., Adams-Chapman, I., Fanaroff, A. A., Hintz, S. R., Vohr, B., Higgins, R.D., National Institute of Child Health and Human Development Neonatal Research Network. 2004. Neurodevelopmental and Growth Impairment Among Extremely Low-Birth-Weight Infants With Neonatal Infection. *JAMA* 292(19):2357-2365.

Stoll, B. J., Hansen, N. I., Sánchez, P. J., Faix, R. G., Poindexter, B. B., Van Meurs, K. P., Bizzarro, M.J., Goldberg, R.N., Frantz, I.D., Hale, E.C., Shankaran, S., Kennedy, K., Carlo, W.A., Watterberg, K.L., Bell, E.F., Walsh, M.C., Schibler, K., Laptook, A.R., Shane, A.L., Schrag, S.J., Das, A., Higgins, R.D., Eunice Kennedy Shriver National Institute of Child Health., and Human Development Neonatal Research Network. 2011. Early onset neonatal sepsis: the burden of group B Streptococcal and E. coli disease continues. *Pediatrics* 127(5):817-826.

Stout, M. J., Conlon, B., Landeau, M., Lee, I., Bower, C., Zhao, Q., Roehl, K.A., Nelson, M.D., Macones, G.A., and Mysorekar, I. U. 2013. Identification of intracellular bacteria in the basal plate of the human placenta in term and preterm gestations. *American Journal of Obstetrics and Gynecology* 208(3):221-227.

Sumitomo, T., Nakata, M., Higashino, M., Terao, Y. and Kawabata, S. 2013. Group A streptococcal cysteine protease cleaves epithelial junctions and contributes to bacterial translocation. *The Journal of Biological Chemistry* 288(19):13317-13324.

Sze, M. A., Dimitriu, P. A., Hayashi, S., Elliott, W. M., McDonough, J. E., Gosselink, J. V., Cooper, J., Sin, D.D., Mohn, W.W., and Hogg, J. C. 2012. The lung tissue microbiome in chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine* 185(10):1073-1080.

Tanaka, S., Kobayashi, T., Songjinda, P., Tateyama, A., Tsubouchi, M., Kiyohara, C., Shirakawa, T., Sonomoto, K., and Nakayama, J. 2009. Influence of antibiotic exposure in the early postnatal period on the development of intestinal microbiota. *FEMS immunology and medical microbiology* 56(1):80-87.

Thomas, N. 2016. Nutritional care of preterm infants: Scientific basis and practical guidelines. *The Indian Journal of Medical Research* 143(4):531-532.

Timmerman, H. M., Rutten, N. B. M. M., Boekhorst, J., Saulnier, D. M., Kortman, G. A. M., Contractor, N., Kullen, M., Floris, E., Haarmsen, H.J.M., Vileger, A.M., Kleerebezem, M., and Rijkers, G. T. 2017. Intestinal colonisation patterns in breastfed and formula-fed infants during the first 12 weeks of life reveal sequential microbiota signatures. *Scientific Reports* 7(1):8327.

Ting, J. Y., Synnes, A., Roberts, A., Deshpandey, A., Dow, K., Yoon, E. W., Lee, K-S., Dobson, S., Lee, S.K., Shah, P.S., and Canadian Neonatal Network of Investigators. 2016. Association Between Antibiotic Use and Neonatal Mortality and Morbidities in Very Low-Birth-Weight Infants Without Culture-Proven Sepsis or Necrotizing Enterocolitis. *JAMA pediatrics* 170(12):1181-1187.

Tolia, V. N., Desai, S., Qin, H., Rayburn, P. D., Poon, G., Murthy, K., Ellsbury, D.L., and Chiruvolu, A. 2017. Implementation of an Automatic Stop Order and Initial Antibiotic Exposure in Very Low Birth Weight Infants. *American Journal of Perinatology* 34(2):105-110.

Tor, L., Kjell, O. and Björn, W. 1982. Protease Inhibitors and their Relation to Protease Activity in Human Milk. *Pediatric Research* 16(6):479-483.

Torrazza, R. M., Ukhanova, M., Wang, X., Sharma, R., Hudak, M. L., Neu, J. and Mai, V. 2013. Intestinal microbial ecology and environmental factors affecting necrotizing enterocolitis. *PloS One* 8(12):e83304.

Toshifumi, O., Nobuhiro, S., Tatuo, O., Koji, M., Masayuki, O. and Isao, O. 2002. Fusobacterium varium localized in the colonic mucosa of patients with ulcerative colitis stimulates species-specific antibody. *Journal of Gastroenterology and Hepatology* 17(8):849-853.

Tremaroli, V. and Bäckhed, F. 2012. Functional interactions between the gut microbiota and host metabolism. *Nature* 489(7415):242-249.

Tripathi, N., Cotten, C. M. and Smith, P. B. 2012. Antibiotic use and misuse in the neonatal intensive care unit. *Clinics in Perinatology* 39(1):61-68.

Trojanowska, D., Zwolinska-Wcislo, M., Tokarczyk, M., Kosowski, K., Mach, T. and Budak, A. 2010. The role of Candida in inflammatory bowel disease. Estimation of transmission of C. albicans fungi in gastrointestinal tract based on genetic affinity between strains. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research* 16(10):CR451-457.

Trompette, A., Gollwitzer, E. S., Yadava, K., Sichelstiel, A. K., Sprenger, N., Ngom-Bru, C., Blachard, C., Junt, T., Nicod, L.P., Harris, N.L., and Marsland, B. J. 2014. Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. *Nature Medicine* 20(2):159-166.

Truninger, K., W Ammann, R., E Blum, H. and Witt, H. 2001. Genetic aspects of chronic pancreatitis: Insights into aetiopathogenesis and clinical implications. *Swiss medical weekly* 131(39-40):565-574.

Turk, B. 2006. Targeting proteases: successes, failures and future prospects. *Nature Reviews. Drug Discovery* 5(9):785-799.

Turnbaugh, P. J. and Gordon, J. I. 2009. The core gut microbiome, energy balance and obesity. *The Journal of Physiology* 587(17):4153-4158.

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., Sogin, M.L., Jones, W.J., Roe, B.A., Affourtit, J.P., Egholm, M., Henrissat, B., Heath, A.C., Knight, R., and Gordon, J. I. 2009) A core gut microbiome in obese and lean twins. *Nature* 457(7228):480-484.

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

Turnbaugh, P. J., Quince, C., Faith, J. J., McHardy, A. C., Yatsunenko, T., Niazi, F., Affourtit, J., Egholm, M., Henrissat, B., Knight, R., and Gordon, J. I. 2010. Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. *Proceedings of the National Academy of Sciences of the United States of America* 107(16):7503-7508.

Turner, J. R. 2009. Intestinal mucosal barrier function in health and disease. *Nature Reviews Immunology* 9(11):799-809.

Turner, J. R., Rill, B. K., Carlson, S. L., Carnes, D., Kerner, R., Mrsny, R. J. and Madara, J. L. 1997. Physiological regulation of epithelial tight junctions is associated with myosin light-chain phosphorylation. *The American Journal of Physiology* 273(4 Pt 1):C1378-1385.

Tóthová, V., Bártlová, S., Dolák, F., Kaas, J., Kimmer, D., Maňhalová, J., Martinek, L., and Olišarová, V. 2014. Quality of life in patients with chronic diseases. *Neuro Endocrinology Letters* 35 Suppl 1:11-18.

Usher, R. H., Allen, A. C. and McLean, F. H. 1971. Risk of respiratory distress syndrome related to gestational age, route of delivery, and maternal diabetes. *American Journal of Obstetrics and Gynecology* 111(6):826-832.

Utech, M., Ivanov, A. I., Samarin, S. N., Bruewer, M., Turner, J. R., Mrsny, R. J., Parkos, C.A., and Nusrat, A. 2005. Mechanism of IFN-gamma-induced endocytosis of tight junction proteins: myosin II-dependent vacuolarization of the apical plasma membrane. *Molecular biology of the cell* 16(10):5040-5052.

Vael, C., Nelen, V., Verhulst, S. L., Goossens, H. and Desager, K. N. 2008. Early intestinal Bacteroides fragilis colonisation and development of asthma. *BMC pulmonary medicine* 8:19-19.

van Dijk, E. L., Auger, H., Jaszczyszyn, Y. and Thermes, C. 2014. Ten years of next-generation sequencing technology. *Trends in Genetics* 30(9):418-426.

Vandeputte-Rutten, L., Kramer, R. A., Kroon, J., Dekker, N., Egmond, M. R. and Gros, P. 2001. Crystal structure of the outer membrane protease OmpT from Escherichia coli suggests a novel catalytic site. *The EMBO journal* 20(18):5033-5039.

Vendrell, J., Querol, E. and Avilés, F. X. 2000. Metallocarboxypeptidases and their protein inhibitors: Structure, function and biomedical properties. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* 1477(1):284-298.

Vergnolle, N. 2000. Review article: proteinase-activated receptors - novel signals for gastrointestinal pathophysiology. *Alimentary Pharmacology & Therapeutics* 14(3):257-266.

Vergnolle, N. 2005. Clinical relevance of proteinase activated receptors (pars) in the gut. *Gut* 54(6):867-874.

Vetvicka, V., Vagner, J., Baudys, M., Tang, J., Foundling, S. and Fusek, M. 1993. Human breast milk contains procathepsin D - Detection by specific antibodies. *Biochemistry and Molecular Biology International* 30(5):921-928.

Vieira, A. T., Rocha, V. M., Tavares, L., Garcia, C. C., Teixeira, M. M., Oliveira, S. C., Cassali, G. D., Gamba, C., Martins, F.S., and Nicoli, J. R. 2016. Control of Klebsiella pneumoniae pulmonary infection and immunomodulation by oral treatment with the commensal probiotic Bifidobacterium longum 51A. *Microbes and Infection* 18(3):180-189.

Vieira, W. A. and Pretorius, E. 2010. The impact of asthma on the gastrointestinal tract (GIT). *Journal of asthma and allergy* 3:123-130.

Visse, R. and Nagase, H. 2003. Matrix metalloproteinases and tissue inhibitors of metalloproteinases: structure, function, and biochemistry. *Circulation Research* 92(8):827-839.

Wadhawan, R., Oh, W., Hintz, S. R., Blakely, M. L., Das, A., Bell, E. F., Saha, S., Laptook, A.R., Shankaran, S, Stoll, B.J., Walsh, M.C., and Higgins, R. D. 2013. Neurodevelopmental outcomes of extremely low birth weight infants with spontaneous intestinal perforation or surgical necrotizing enterocolitis. *Journal Of Perinatology* 34(1):64-70.

Wampach, L., Heintz-Buschart, A., Hogan, A., Muller, E. E. L., Narayanasamy, S., Laczny, C. C., Hugerth, L.W., Bindl, L., Bottu, J., Andersson, A.F., de Beaufort, C., and Wilmes, P. 2017. Colonization and Succession within the Human Gut Microbiome by Archaea, Bacteria, and Microeukaryotes during the First Year of Life. *Frontiers in Microbiology* 8:738.

Wang, M. L., Dorer, D. J., Fleming, M. P. and Catlin, E. A. 2004. Clinical Outcomes of Near-Term Infants. *Pediatrics* 114(2):372-376.

Wang, W., Uzzau, S., Goldblum, S. E. and Fasano, A. 2000. Human zonulin, a potential modulator of intestinal tight junctions. *Journal of Cell Science* 113 Pt 24:4435-4440.

Ward, D. M., Weller, R. and Bateson, M. M. 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* 345(6270):63-65.

Ward, D. V., Scholz, M., Zolfo, M., Taft, D. H., Schibler, K. R., Tett, A., Segata, N., and Morrow, A. L. 2016. Metagenomic Sequencing with Strain-Level Resolution Implicates Uropathogenic E. coli in Necrotizing Enterocolitis and Mortality in Preterm Infants. *Cell Reports* 14(12):2912-2924.

Weeks, C. S., Tanabe, H., Cummings, J. E., Crampton, S. P., Sheynis, T., Jelinek, R., Vanderlick, K.T., Cocco, M.J., and Ouellette, A. J. 200). Matrix Metalloproteinase-7 Activation of Mouse Paneth Cell Pro-α-defensins: ser43↓ile44 proteolysis enables membrane-disruptive activity. *Journal of Biological Chemistry* 281(39):28932-28942.

Weinstock, G. M. 2012. Genomic approaches to studying the human microbiota. *Nature* 489(7415):250-256.

Wesolowska-Andersen, A., Bahl, M. I., Carvalho, V., Kristiansen, K., Sicheritz-Pontén, T., Gupta, R. and Licht, T. R. 2014. Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis. *Microbiome* 2(1):19.

Whiteley, H. R. 1957. Fermentation of amino acids by micrococcus aerogenes. *The Journal of Bacteriology* 74(3):324.

Whittaker, R. H. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs* 30(3):279-338.

Wilkins, D., Leung, M. H. Y. and Lee, P. K. H. 2017. Microbiota fingerprints lose individually identifying features over time. *Microbiome* 5(:1.

Willing, B. P., Dicksved, J., Halfvarson, J., Andersson, A. F., Lucio, M., Zheng, Z., Jamerot, G., Tysk, C., Jansson, J.K., and Engstrand, L. 2010. A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* 139(6):1844-1854.

Wilson, C. L., Ouellette, A. J., Satchell, D. P., Ayabe, T., López-Boado, Y. S., Stratman, J. L., Hultgren, S.J., Matrisian, L.M., and Parks, W. C. 1999. Regulation of intestinal alpha-defensin activation by the metalloproteinase matrilysin in innate host defense. *Science* 286(5437):113-117.

Windey, K., De Preter, V. and Verbeke, K. 2012. Relevance of protein fermentation to gut health. *Molecular nutrition and food research* 56(1):184-196.

Wine, E., Chan, V. L. and Sherman, P. M. 2008. Campylobacter jejuni mediated disruption of polarized epithelial monolayers is cell-type specific, time dependent, and correlates with bacterial invasion. *Pediatric Research* 64(6):599-604.

Winglee, K., Howard, A. G., Sha, W., Gharaibeh, R. Z., Liu, J., Jin, D., Fodor, A.A., and Gordon-Larsen, P. 2017. Recent urbanization in China is correlated with a Westernized microbiome encoding increased virulence and antibiotic resistance genes. *Microbiome* 5(1):121.

Wirtschafter, D. D., Padilla, G., Suh, O., Wan, K., Trupp, D. and Simon Fayard, E. E. 2011. Antibiotic use for presumed neonatally acquired infections far exceeds that for central line-associated blood stream infections: an exploratory critique. *Journal Of Perinatology* 31(8):514-518.

Woese, C. R. and Fox, G. E. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America* 74(11):5088-5090.

Woodmansey, E. J., McMurdo, M. E. T., Macfarlane, G. T. and Macfarlane, S. 2004. Comparison of compositions and metabolic activities of fecal microbiotas in young adults and in antibiotic-treated and non-antibiotic-treated elderly subjects. *Applied and environmental microbiology* 70(10):6113-6122.

Wu, S., Jiang, Z.-Y., Sun, Y.-F., Yu, B., Chen, J., Dai, C.-Q., Wu, X.L., Tang, X.L., and Chen, X.-Y. 2013. Microbiota Regulates the TLR7 Signaling Pathway Against Respiratory Tract Influenza A Virus Infection. *Current Microbiology* 67(4):414-422.

Wu, S., Lim, K. C., Huang, J., Saidi, R. F. and Sears, C. L. 1998. Bacteroides fragilis enterotoxin cleaves the zonula adherens protein, E-cadherin. *Proceedings of the National Academy of Sciences of the United States of America* 95(25):14979-14984.

Wójkowska-Mach, J., Różańska, A., Borszewska-Kornacka, M., Domańska, J., Gadzinowski, J., Gulczyńska, E., Helwich, E., Kordek, A., Pawlik, D., Szczapa, J., and Heczko, P. B. 2014. Necrotising Enterocolitis in Preterm Infants: Epidemiology and Antibiotic Consumption in the Polish Neonatology Network Neonatal Intensive Care Units in 2009. *PLOS ONE* 9(3):e92865.

Xu, Z., Wang, Z., Yuan, C., Liu, X., Yang, F., Wang, T., Wang, J., Manabe, K., Qin, Q., Wang, X., Zhang, Y., and Zhang, M. 2016. Dandruff is associated with the conjoined interactions between host and microorganisms. *Scientific Reports* 6:24877.

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K.-H., Whitman, W.B., Euzeby, J., Amann, R., and Rosselló-Móra, R. 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews. Microbiology* 12(9):635-645.

Yassour, M., Vatanen, T., Siljander, H., Hämäläinen, A.-M., Härkönen, T., Ryhänen, S. J., Franzosa, E.A., Vlamakis, H., Huttenhower, C., Gevers, D., Lander, E.S., Knip, M., Diabimmune Study Group and Xavier, R. J. 2016. Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. *Science Translational Medicine* 8(343):343-381.

Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., Heath, A.C., Warner, B., Reader, J., Kuczynski, J., Caporaso, G.J., Lozupone, C.A., Lauber, C., Clemente, J.C., Knights, D., Knight R., and Gordon, J. I. 2012. Human gut microbiome viewed across age and geography. *Nature* 486(7402):222.

Yazar, A., Atis, S., Konca, K., Pata, C., Akbay, E., Calikoglu, M. and Hafta, A. 2001. Respiratory Symptoms and Pulmonary Functional Changes in Patients With Irritable Bowel Syndrome. *American Journal of Gastroenterology* 96(5):1511-1516.

Yeruva, L., Spencer, N. E., Saraf, M. K., Hennings, L., Bowlin, A. K., Cleves, M. A., Marcer, K., Chintapalli, S.V., Shankar, K., Rank, R.G., Badger, T.M., and Ronis, M. J. J. 2016. Formula diet alters small intestine morphology, microbial abundance and reduces VE-cadherin and IL-10 expression in neonatal porcine model. *BMC gastroenterology* 16(1):40.

Yuan, S., Cohen, D. B., Ravel, J., Abdo, Z. and Forney, L. J. 2012. Evaluation of Methods for the Extraction and Purification of DNA from the Human Microbiome. *PLOS ONE* 7(3):e33865.

Yurkovetskiy, L., Burrows, M., Khan, Aly a., Graham, L., Volchkov, P., Becker, L., Antonopoulos, D., Umesaki, Y., and Chervonsky, A.V. 2013. Gender Bias in Autoimmunity Is Influenced by Microbiota. *Immunity* 39(2):400-412.

Zeitz, J., Biedermann, L., Fried, M., Frei, P., Rogler, G., Scharl, M., Vavricka, S.R., Brulisauer, K., Schuppler, M., and Loessner, M. J. 2014. Smoking Cessation Alters Intestinal Microbiota: Insights from Quantitative Investigations on Human Fecal Samples Using FISH. *Inflammatory Bowel Diseases* 20(9):1496-1501.

Zhou, Y., Shan, G., Sodergren, E., Weinstock, G., Walker, W. A. and Gregory, K. E. 2015. Longitudinal analysis of the premature infant intestinal microbiome prior to necrotizing enterocolitis: a case-control study. *PloS One* 10(3):e0118632.

Zindel, U., Freudenberg, W., Rieth, M., Andreesen, J., Schnell, J. and Widdel, F. 1988. Eubacterium acidaminophilum sp. nov., a versatile amino acid-degrading anaerobe producing or utilizing H 2 or formate. *Archives of Microbiology* 150(3):254-266.

Zou, Z.-H., Liu, D., Li, H.-D., Zhu, D.-P., He, Y., Hou, T. and Yu, J.-L. 2018. Prenatal and postnatal antibiotic exposure influences the gut microbiota of preterm infants in neonatal intensive care units. *Annals of Clinical Microbiology and Antimicrobials* 17(1):9.

Zwittink, R. D., Renes, I. B., Lingen, R. A. v., Zoeren-Grobben, D. v., Konstanti, P., Norbruis, O. F., Martin, R., Jebbink, L.J.M.G., Knol, J., and Belzer, C. 2018. Association between duration of intravenous antibiotic administration and early-life microbiota development in late-preterm infants. *European Journal of Clinical Microbiology & Infectious Diseases* 37(3):475-483.

Appendices

Appendix 1. Parental Consent and Information for Microbiome Study.

Parental Information Sheet (Bristol)

Information sheet for Parents/Guardians – Babies admitted to the neonatal unit

Principal Investigator: Dr R Wach, Consultant Neonatologist.

1. Study Title
The microbiome of the neonatal lung and its effect on chronic lung disease of prematurity

2. Invitation
You are being invited to take part in a research study being led by Cardiff University. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Please take time to decide whether or not you wish to take part.

Thank you for reading this leaflet.

3. What is the purpose of the study?
Chronic Lung Disease of prematurity (CLD), which is also often called BPD (for bronchopulmonary dysplasia), is a common disease of premature babies. Babies with CLD require oxygen for prolonged periods and are sometimes discharged home on oxygen. Our research is trying to understand why some babies develop CLD.

Until the past few years, it was thought that the lungs of healthy babies (and adults) were sterile, i.e. contained no bacteria. New techniques, however, have shown that there are low levels of bacteria in all people's lungs (known as the microbiome of the lung). It is not known when babies acquire these bacteria, or what type of bacteria are present in healthy baby's lungs. Our research seeks learn more about the acquisition of this bacteria.

It is becoming clear that these bacteria may play a part in many lung diseases. Our research will also look into the role these bacteria may have in the development of chronic lung disease

of prematurity.

4. Why have I been chosen?
We would like to invite 3 groups of babies to join the study:

Study group: Babies who have been born prematurely (At or before 32 weeks gestation) and require the assistance of a breathing machine to support their breathing

Control group of ventilated babies: Babies who have been born at term (i.e. at or after 37 weeks gestation) and who need help with a breathing machine to support breathing for non-respiratory reasons) commonly if they undergo surgery)

*We would like to invite you to join the study because your baby was born at 32 weeks gestation or less and needs a breathing machine to help his/her breathing **OR** your baby was born at or after 37 weeks gestation and requires a breathing machine to help his/her breathing*

5. Does my baby have to take part?
It is up to you to decide whether or not to consent for your baby to take part. If you do allow your baby to take part in the study, you will be given this information sheet to keep and be asked to sign a consent form. Even if you do decide to consent to your baby taking part, you are still free to withdraw your baby at any time without giving any reasons. A decision to withdraw at any time, or a decision not to take part, will not affect the care that your baby will receive.

6. What will happen to my baby if we take part?

Airway Fluid Samples
We would like to obtain airway fluid from your baby. All babies who are on a breathing machine regularly have the breathing tube sucked out to clear the secretions to prevent the tube from blocking. The secretions sucked out are usually thrown away. We would use these secretions for our study.

Samples will be taken only when the nurse or doctor caring for your baby feels the baby needs suctioning. There are no extra risks associated with collecting the secretions.

Nasal Fluid Samples
Babies on the neonatal unit often have the secretions from their nose suctioned out. When the nursing staff do this suctioning we would like to use the sample for our study.

We would like to collect nose fluid samples daily during the first week of life and then twice weekly until 4 weeks of age, while the baby remains on the neonatal unit, but only when it is being performed by the nursing staff as needed by the baby. Once the breathing tube is removed we would continue to samples from the nose.

Stool Samples
There is some evidence that the bacteria in the gut can influence the lung bacteria or microbiome. We would therefore like to analyse stool samples from your baby. We would aim to collect the first stool sample and one sample per week for the first month, or until your baby is discharged from the neonatal unit, whichever is sooner.

7. Will this affect my baby's treatment?
The medical care of your baby will not be affected by this study. The information from this study will not be used to diagnose or treat your baby.

8. Is there any benefit to taking part in this study?
There will be no benefit to you or your baby from taking part in this study. The information from this study will not be used to diagnose or treat your baby. The study should improve our understanding the role of germs play in the development of lung disease in premature babies.

9. What will happen to the samples collected?
The samples will be transferred to our laboratories at Cardiff University. We shall analyse the fluid from the airways in a number of ways:

1) We will use molecular biological techniques to identify any bacterial DNA present in the fluid to identify what species of bacteria are present
2) We will measure markers of inflammation in the fluid to help us identify if the bacteria are causing infection.
3) We will look for molecules that many germs may produce (metabolites) in the lung fluid and blood to see if blood tests can be used to identify any bacteria.

We shall analyse the samples in our laboratories at Cardiff University but some analyses may be conducted by commercial companies or other university laboratories which have expertise to analyse the samples. The samples will have a code and will not have any information about the baby.

The baby's DNA will be extracted alongside the bacterial DNA but it will not be used further in this study. With your permission any remaining samples, including DNA (baby and bacterial), may be stored for future research into chronic lung disease of prematurity. The samples will be anonymised before use in future studies and may be accessed by researchers in the UK and abroad, the research may include genetic (e.g. DNA), commercial and animal research. You may withdraw your consent for the storage and future use of your baby's samples at any point. If you do withdraw your consent your baby's samples will not be used in any subsequent studies and will be destroyed according to locally approved practices. Any samples already distributed for use in research prior to the withdrawal of consent will continue to be used in that study and any samples remaining at the end of the study will be destroyed.

10. What are the risks of taking part?
Babies who receive mechanical ventilation are monitored closely for their heart rate and oxygen levels. The risks are the same as those of routine suctioning that the baby may have. Sometimes the babies may need extra oxygen, typically 5 – 10%, for 5 – 10 minutes and sometimes especially when the suction tube is placed the heart rate may drop for a few seconds (usually less than 30 seconds). We would monitor the baby throughout the procedure and stop it if the baby becomes unwell in any way.

11. What if something goes wrong?
If your baby is harmed by taking part in this research project, there are no special compensation arrangements. If your baby is harmed due to someone's negligence, then you may have grounds for a legal action but you may have to pay for it. Regardless of this, if you wish to complain, or have any concerns about any aspect of the way you or your baby have been approached or treated during the course of this study, the normal National Health Service complaints mechanisms should be available to you.

12. Will my baby's taking part be kept confidential?
All information which is collected about your baby during the course of the research will be kept strictly confidential. We will assign a number to each baby and use this to label the samples

obtained for the study.

13. What will happen to the results of the study?
We will publish the results in reputable medical journals and present the data at scientific and medical meetings. Your baby's name and details will NOT be revealed at any stage. Please let us know if you would like a copy of the report.

14. Who is paying for the study?
The study is being funded by departmental funds.

15. Who had reviewed the study?
This study has been reviewed by the Wales Research Ethics Committee 2 and also by the North Bristol NHS Trust Research and Development committee.

16. Who can I contact for further information?
You may contact Dr Richard Wach by asking one of the staff on the neonatal unit, by telephone 0117 414 6800 or by email Richard.wach@nbt.nhs.uk . Alternatively, please contact Dr David Gallacher by email gallacherdj@cardiff.ac.uk or telephone 029 20 74 3375.

Thank you for taking time to read this information leaflet at this time. Please do not hesitate to ask Dr David Gallacher or Dr Richard Wach if you would like to discuss anything further.

Dr David Gallacher	Dr Richard Wach	Professor Sailesh Kotecha
Clinical Research Fellow	**Consultant Neonatologist**	**Consultant in Neonatal Medicine**

Patient Identification Number for this study:

Sticker

Patient

PARENT/GUARDIAN CONSENT FORM

Project Title: The Microbiome of the Neonatal Lung and its Effect on Chronic Lung Disease of Prematurity

Principle Investigators: **Dr R Wach, Consultant Neonatologist**

Contact Details: Neonatal Unit, 0117 414 6800

This form should be read in conjunction with the Patient Information Leaflet (Neonatal Unit Patients), version no 4(Bristol) dated 4th March 2015

Please initial relevant

1. I confirm that I have read and understand the information sheet dated 4th March 2015 (version 4- babies admitted to the neonatal unit) for the above study and have had the opportunity to ask questions. ☐

2. I understand that my and my baby's participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my or my baby's medical care or legal rights being affected. ☐

3. I understand that sections of any of my baby's medical notes and data collected during the study, may be looked at by individuals from Cardiff University, from regulatory authorities or from the NHS Trust, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records. ☐

4. I agree for my baby to take part in the above study. ☐

5. I understand that samples of my baby's airway fluid will be collected for this study. I understand that my baby's DNA will be extracted alongside the bacterial DNA but will not be used in this study. ☐

6. I give permission for any remaining samples (including my baby's DNA) to be used in future for chronic lung disease of prematurity research in the UK and abroad, which may include genetic (e.g. DNA), commercial or animal research. I understand I am free to withdraw my consent to future research at any point and that all samples will be destroyed as detailed in the information sheet.

Yes ☐ No ☐

_____ _____

_____

Name of Parent/Guardian	Date	Signature
_____	_____	_____
Name of Person taking consent (if different from researcher)	Date	Signature
_____	_____	_____
Researcher	Date	Signature

1 for patient; 1 for researcher; 1 to be kept with hospital notes

Parental Information Sheet (Cardiff)

Cardiff and Vale University Health Board
Bwrdd Iechyd Prifysgol Caerdydd a'r Fro

University Hospital of Wales
Ysbyty Athrofaol Cymru

Heath Park,
Cardiff,
CF14 4XW
Phone (029) 2074 7747
Fax (029) 2074 3838

Parc Y Mynydd Bychan,
Caerdydd,
CF14 4XW
Ffôn (029) 2074 7747
Ffacs (029) 2074 3838

Information sheet for Parents/Guardians – Babies admitted to the neonatal unit

Principal Investigator: Prof Sailesh Kotecha, Consultant Neonatologist.

1. Study Title

The microbiome of the neonatal lung and its effect on chronic lung disease of prematurity

2. Invitation

You are being invited to take part in a research study led by Cardiff University. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Please take time to decide whether or not you wish to take part.

Thank you for reading this leaflet.

3. What is the purpose of the study?

Chronic Lung Disease of prematurity (CLD), which is also often called BPD (for bronchopulmonary dysplasia), is a common disease of premature babies. Babies with CLD require oxygen for prolonged periods and are sometimes discharged home on oxygen. Our research is trying to understand why some babies develop CLD.

Until the past few years, it was thought that the lungs of healthy babies (and adults) were sterile, i.e. contained no bacteria. New techniques, however, have shown that there are low levels of bacteria in all people's lungs (known as the microbiome of the lung). It is not known when babies acquire these bacteria, or what type of bacteria are present in healthy baby's lungs. Our research seeks learn more about the acquisition of this bacteria.

It is becoming clear that these bacteria may play a part in many lung diseases. Our research will also look into the role these bacteria may have in the development of chronic lung disease of prematurity.

4. Why have I been chosen?

We would like to invite 3 groups of babies to join the study:

Study group: Babies who have been born prematurely (At or before 32 weeks gestation) and require the assistance of a breathing machine to support their breathing

Control group of ventilated babies: Babies who have been born at term (i.e. at or after 37 weeks gestation) and who need help with a breathing machine to support breathing for non-respiratory reasons, commonly if they undergo surgery.

Control group of babies on the postnatal wards: Babies born at term without any complications. Only the nasal passage samples will be collected.

We would like to invite you to join the study because your baby was born at 32 weeks gestation or less and needs a breathing machine to help his/her breathing ***OR*** *your baby was born at or after 37 weeks gestation and requires a breathing machine to help his/her breathing*

5. Does my baby have to take part?

It is up to you to decide whether or not to consent for your baby to take part. If you do allow your baby to take part in the study, you will be given this information sheet to keep and be asked to sign a consent form. Even if you do decide to consent to your baby taking part, you are still free to withdraw your baby at any time without giving any reasons. A decision to withdraw at any time, or a decision not to take part, will not affect the care that your baby will receive.

6. What will happen to my baby if we take part?

Lung Fluid Samples

We would like to obtain lung fluid from your baby for the first 28 days of age, daily for the first week then twice weekly. Whilst the baby needs help with a breathing machine, the breathing tube is often sucked out by the nurses to prevent it from blocking. We would perform our suctioning when the nursing staff caring for the baby would be performing the routine suctioning. It will replace the need for the routine suctioning, so it does not need to be performed twice.

In order to compare the results with other baby's results, we have standardised this method of suctioning: we place the baby on his/her back and turn the head to the left side to encourage the suction tube to go down the right lung. We will then gently place a suction tube through the breathing tube into the lungs and through the tube insert saline (salt water). The amount of saline is based on the baby's weight using 1 ml for each kilogram of the baby's weight (one teaspoon is 5 ml). After instilling the saline we will suck up as much fluid as possible and repeat the procedure once more. The returned fluid will consist of the saline and will also have the baby's lung fluid which we can use for our research.

The suctioning performed to obtain the research samples involves inserting the suction tube further into the lungs than the routine suctioning and is done primarily for sample collection.

Babies may be uncomfortable during the procedure. Occasionally the baby's heart rate

may drop briefly as the suction tube is inserted, and the baby sometimes may need a little more oxygen (usually 5-10%) for a short period of time. We will monitor the baby's heart rate and oxygen saturation during and after the procedure, and stop the procedure if the baby becomes unwell in any way.

Babies on breathing machines regularly have their breathing tube cleared of the secretions by the nursing staff. The secretions are usually thrown away but we would like to use these secretions to compare results with the standardised collection of fluid. We would like to use the secretions collected by the nursing staff for our study. These Samples will only be taken when the nurse or doctor caring for your baby thinks the baby needs suctioning. These samples would also be collected once daily in the first week of life and then twice weekly until 28 days of age.

To compare the bacteria in the lungs with that in the upper airways we would like to obtain samples of secretions from your baby's nose around the same time the lung fluid is collected. This would involve inserting a thin suction tube into both nostrils and suctioning any secretions that are present. These secretions will also be analysed for any bacteria. The nursing staff regularly suction babies' noses to clear secretions. We would replace this suctioning wherever possible so it does not need to be performed twice. Once your baby has had the breathing tube removed, we would like to continue to collect the samples from the nose daily during the first week of life and the twice weekly until 4 weeks of age, while the baby remains on the neonatal unit.

Stool Samples

There is some evidence that the bacteria in the gut can influence the lung bacteria or microbiome. We would therefore like to analyse stool samples from your baby. We would aim to collect the first stool sample and one sample per week for the first month, or until your baby is discharged from the neonatal unit, whichever is sooner.

Blood Samples

We would also like to take 0.5 to 1 ml of your baby's blood once a week for the first 4 weeks. Babies on the neonatal unit have regular blood tests. We would ask the doctors performing these tests to take an extra 0.5 to 1ml of blood (1 tea spoon is 5mls) once a week when they are taking blood routinely. Samples will only be collected if the team looking after your baby are doing a blood test on your baby.

7. Will this affect my baby's treatment?

The medical care of your baby will not be affected by this study. The information from this study will not be used to diagnose or treat your baby.

8. Is there any benefit to taking part in this study?

There will be no benefit to you or your baby from taking part in this study. The information from this study will not be used to diagnose or treat your baby. The study should improve our understanding the role of germs play in the development of lung disease in premature babies.

9. What will happen to the samples collected?

The samples will be transferred to our laboratories at Cardiff University. We shall analyse the fluid from the nasal passages in a number of ways:

4) We will use molecular biological techniques to identify any bacterial DNA present in the fluid to identify what species of bacteria are present
5) We will measure markers of inflammation in the fluid to help us identify if the bacteria are causing infection.
6) We will look for molecules that many germs may produce (metabolites) in the lung fluid and blood to see if blood tests can be used to identify any bacteria.

We shall analyse the samples in our laboratories at Cardiff University but some analyses may be conducted by commercial companies or other university laboratories which have expertise to analyse the samples. The samples will have a code and will not have any information about the baby.

The baby's DNA will be extracted alongside the bacterial DNA but it will not be used further in this study. With your permission any remaining samples, including lung fluid, stool, blood and DNA (baby and bacterial), may be stored for future research into chronic lung disease of prematurity. The samples will be anonymised before use in future studies and may be accessed by researchers in the UK and abroad, the research may include genetic (e.g. DNA), commercial and animal research. You may withdraw your consent for the storage and future use of your baby's samples at any point. If you do withdraw your consent your baby's samples will not be used in any subsequent studies and will be destroyed according to locally approved practices. Any samples already distributed for use in research prior to the withdrawal of consent will continue to be used in that study and any samples remaining at the end of the study will be destroyed.

10. What are the risks of taking part?

Babies who receive mechanical ventilation are monitored closely for their heart rate and oxygen levels. The risks are similar to those of routine suctioning that the baby may have. Sometimes the babies may need extra oxygen, typically 5 – 10%, for 5 – 10 minutes and sometimes especially when the suction tube is placed the heart rate may drop for a few seconds (usually less than 30 seconds). We would monitor the baby throughout the procedure and stop it if the baby becomes unwell in any way.

Taking blood from your baby will only be done when your baby is having a necessary blood test by the doctors treating him/her. Preterm babies often need blood transfusions. Taking an extra 0.5 – 1 ml should not affect this.

11. What if something goes wrong?

If your baby is harmed by taking part in this research project, there are no special compensation arrangements. If your baby is harmed due to someone's negligence, then you may have grounds for a legal action but you may have to pay for it. Regardless of this, if you wish to complain, or have any concerns about any aspect of the way you or your baby have been approached or treated during the course of this study, the normal National Health Service complaints mechanisms should be available to you.

12. Will my baby's taking part be kept confidential?

All information which is collected about your baby during the course of the research will be kept strictly confidential. We will assign a number to each baby and use this to label the samples obtained for the study.

13. What will happen to the results of the study?

We will publish the results in reputable medical journals and present the data at scientific and medical meetings. Your baby's name and details will NOT be revealed at any stage. Please let us know if you would like a copy of the report.

14. Who is paying for the study?

The study is being funded by departmental funds.

15. Who had reviewed the study?

This study has been reviewed by theWales Research Ethics Committee 2and also by the Cardiff and Vale University Health Board.

16. Who can I contact for further information?

You may contact Dr David Gallacher or Professor Sailesh Kotecha by asking one of the staff on the neonatal unit or by telephoning 029 20 74 3375 or by mail to: gallacherdj@cardiff.ac.uk .

Professor Sailesh Kotecha, Neonatal Unit, Heath Hospital, Heath Park, Cardiff CF14 4XN.

Thank you for taking time to read this information leaflet at this time. Please do not hesitate to ask Dr David Gallacher or Professor Sailesh Kotecha if you would like to discuss anything further.

Dr David Gallacher

Clinical Research Fellow

Professor Sailesh Kotecha

Consultant in Neonatal Medicine

Patient Identification Number for this study:

Sticker

Patient

PARENT/GUARDIAN CONSENT FORM

Project Title: The Microbiome of the Neonatal Lung and its Effect on Neonatal Lung Disease
Principle Investigators: Professor Sailesh Kotecha, Professor of Child Health
Contact Details: Neonatal Unit, 029 20 74 3375

This form should be read in conjunction with the Patient Information Leaflet (Neonatal Unit Patients), version no 4 (Cardiff) dated 4th March 2015.

Please initial relevant

1.I confirm that I have read and understand the information sheet dated 4th March 2015 (version 4-neonatal unit patients) for the above study and have had the opportunity to ask questions. ☐

2. I understand that my and my baby's participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my or my baby's medical care or legal rights being affected. ☐

3. I understand that sections of any of my baby's medical notes and data collected during the study, may be looked at by individuals from Cardiff University, from regulatory authorities or from the NHS Trust, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records. ☐

4. I agree for my baby to take part in the above study. ☐

5. I understand that samples of my baby's lung fluid, blood, stool and nasal fluid will be collected for this study. I understand that my baby's DNA will be extracted alongside the bacterial DNA but will not be used in this study. ☐

6. I give permission for any remaining samples (including my baby's DNA) to be used in future for chronic lung disease of prematurity research in the UK and abroad, which may include genetic (e.g. DNA), commercial or animal research. I understand I am free to withdraw my consent to future research at any point and that all samples will be destroyed as detailed in the information sheet.

Yes ☐ No ☐

_____ _____

_____

Name of Parent/Guardian	Date	Signature
_____	_____	_____
Name of Person taking consent (if different from researcher)	Date	Signature
_____	_____	_____
Researcher	Date	Signature

1 for patient; 1 for researcher; 1 to be kept with hospital notes

Parental Information Sheet – Post Natal Ward (Cardiff)

Cardiff and Vale University Health Board
Bwrdd Iechyd Prifysgol Caerdydd a'r Fro

University Hospital of Wales
Ysbyty Athrofaol Cymru

Heath Park,
Cardiff,
CF14 4XW
Phone (029) 2074 7747
Fax (029) 2074 3838

Parc Y Mynydd Bychan,
Caerdydd,
CF14 4XW
Ffôn (029) 2074 7747
Ffacs (029) 2074 3838

Information sheet for Parents/Guardians – Postnatal Ward Babies

Principal Investigator: Prof Sailesh Kotecha, Consultant Neonatologist.

1. Study Title

The microbiome of the neonatal lung and its effect on chronic lung disease of prematurity

2. Invitation

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Please take time to decide whether or not you wish to take part.

Thank you for reading this leaflet.

3. What is the purpose of the study?

Chronic Lung Disease of prematurity (CLD), which is also often called BPD (for bronchopulmonary dysplasia), is a common disease of premature babies. Babies with CLD require oxygen for prolonged periods and are sometimes discharged home on oxygen. Our research is trying to understand why some babies develop CLD.

Until the past few years, it was thought that the lungs of healthy babies (and adults) were sterile, i.e. contained no bacteria. New techniques, however, have shown that there are low levels of bacteria in all people's lungs (known as the microbiome of the lung). It is not known when babies acquire these bacteria, or what type of bacteria are present in healthy baby's lungs. Our research seeks learn more about the acquisition of this bacteria.

It is becoming clear that these bacteria may play a part in many lung diseases. Our research will also look into the role these bacteria may have in the development of chronic lung disease of prematurity.

4. Why have I been chosen?

We would like to invite 3 groups of babies to join the study:

Study group: Babies who have been born prematurely (At or before 32 weeks gestation) and require the assistance of a breathing machine to support their breathing

Control group of ventilated babies: Babies who have been born at term (i.e. at or after 37 weeks gestation) and who need help with a breathing machine to support breathing for non-respiratory reasons) commonly if they undergo surgery)

Control group of babies on the postnatal wards: Babies born at term without any complications. Only the nasal passage samples will be collected.

We would like to invite you to join the study because your baby has been delivered at or after 37 weeks gestation and is well on the postnatal wards.

5. Does my baby have to take part?

It is up to you to decide whether or not to consent for your baby to take part. If you do allow your baby to take part in the study, you will be given this information sheet to keep and be asked to sign a consent form. Even if you do decide to consent to your baby taking part, you are still free to withdraw your baby at any time without giving any reasons. A decision to withdraw at any time, or a decision not to take part, will not affect the care that your baby will receive.

6. What will happen to my baby if we take part?

For babies who are well and on the postnatal ward, we would like to collect samples of the fluid and mucous in the nasal passages, to act as a marker of what may be happening in the lungs. This would involve inserting a thin suction tube into both nostrils and suctioning any secretions that are present. These secretions will be analysed for any bacteria. We would like to suction the nasal passages daily during the first week of life, or for as long as you and your baby remain on the postnatal ward, whichever is sooner.

We would also like to collect daily stool samples from your baby's dirty nappy while he/she is in hospital.

7. Will this affect my baby's treatment?

The medical care of your baby will not be affected by this study. The information from this study will not be used to diagnose or treat your baby.

8. What will happen to the samples collected?

The samples will be transferred to our laboratories at Cardiff University. We shall analyse the fluid from the nasal passages and the stool in a number of ways:

1) We will use molecular biological techniques to identify any bacterial DNA present

in the samples to identify what species of bacteria are present

2) We will measure markers of inflammation in the samples to help us identify if the bacteria are causing infection.
3) We will look for molecules that many germs may produce (metabolites) in the samples to see if blood tests can be used to identify any bacteria.

We shall analyse the samples in our laboratories at Cardiff University but some analyses may be conducted by commercial companies or other university laboratories which have expertise to analyse the samples. The samples will have a code and will not have any information about the baby.

The baby's DNA will be extracted alongside the bacterial DNA but it will not be used further in this study. With your permission any remaining samples, including lung fluid, stool, blood and DNA (baby and bacterial), may be stored for future research into chronic lung disease of prematurity. The samples will be anonymised before use in future studies and may be accessed by researchers in the UK and abroad, the research may include genetic (e.g. DNA), commercial and animal research. You may withdraw your consent for the storage and future use of your baby's samples at any point. If you do withdraw your consent your baby's samples will not be used in any subsequent studies and will be destroyed according to locally approved practices. Any samples already distributed for use in research prior to the withdrawal of consent will continue to be used in that study and any samples remaining at the end of the study will be destroyed.

9. Is there any benefit to taking part in this study?

There will be no benefit to you or your baby from taking part in this study. The information from this study will not be used to diagnose or treat your baby. The study should improve our understanding the role of germs play in the development of lung disease in premature babies.

10. What are the risks of taking part?

Suctioning of the nasal passages is a procedure regularly performed on babies on the neonatal unit and babies admitted to paediatric wards. The procedure only takes 20 – 30 seconds and is only mildly uncomfortable for the babies.

11. What if something goes wrong?

If your baby is harmed by taking part in this research project, there are no special compensation arrangements. If your baby is harmed due to someone's negligence, then you may have grounds for a legal action but you may have to pay for it. Regardless of this, if you wish to complain, or have any concerns about any aspect of the way you or your baby have been approached or treated during the course of this study, the normal National Health Service complaints mechanisms should be available to you.

12. Will my baby's taking part be kept confidential?

All information which is collected about your baby during the course of the research will be kept strictly confidential. We will assign a number to each baby and use this to label the samples obtained for the study.

13. What will happen to the results of the study?

We will publish the results in reputable medical journals and present the data at scientific and medical meetings. Your baby's name and details will NOT be revealed at any stage. Please let us know if you would like a copy of the report.

14. Who is paying for the study?

The study is being funded by departmental funds.

15. Who had reviewed the study?

This study has been reviewed by the Wales Research Ethics Committee 2 and also by the Cardiff and Vale University Health Board.

16. Who can I contact for further information?

You may contact Dr David Gallacher or Professor Sailesh Kotecha by asking one of the staff on the neonatal unit or by telephoning 029 20743375 or by mail to: gallacherdj@cardiff.ac.uk .

Professor Sailesh Kotecha, Neonatal Unit, Heath Hospital, Heath Park, Cardiff CF14 4XN.

Thank you for taking time to read this information leaflet at this time. Please do not hesitate to ask Dr David Gallacher or Professor Sailesh Kotecha if you would like to discuss anything further.

Dr David Gallacher — Clinical Research Fellow

Professor Sailesh Kotecha — Consultant in Neonatal Medicine

Patient Identification Number for this study:

Patient Sticker

PARENT/GUARDIAN CONSENT FORM

Project Title: The Microbiome of the Neonatal Lung and its Effect on Chronic Lung Disease of Prematurity

Principle Investigators: Professor Sailesh Kotecha, Consultant Neonatologist

Contact Details: Neonatal Unit, 029 20 74 3375

This form should be read in conjunction with the Patient Information Leaflet (postnatal ward babies), version 4 dated 28th July 2016.

Please initial relevant

1. I confirm that I have read and understand the information sheet dated 28th July 2016 (version 4 – post natal ward babies) for the above study and have had the opportunity to ask questions. ☐

2. I understand that my and my baby's participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my or my baby's medical care or legal rights being affected. ☐

3. I understand that sections of any of my baby's medical notes may be looked at by the research individuals I give permission for these individuals to have access to my baby's records. ☐

4. I agree for my baby to take part in the above study. ☐

5. I understand that samples of my baby's nasal fluid and stool will be collected for this study. I understand that my baby's DNA will be extracted alongside the bacterial DNA but will not be used in this study. ☐

6. I give permission for any remaining samples (including my baby's DNA) to be used in future for chronic lung disease of prematurity research in the UK and abroad, which may include genetic (e.g. DNA), commercial or animal research. I understand I am free to withdraw my consent to future research at any point and that all samples will be destroyed as detailed in the information sheet.

Yes ☐ No ☐

_____	_____	
	_____	
Name of Parent/Guardian	Date	Signature
_____	_____	
	_____	
Name of Person taking consent (if different from researcher)	Date	Signature
_____	_____	
	_____	

Researcher	Date	Signature

1 for patient; 1 for researcher; 1 to be kept with hospital notes

Appendix 2. Primer Sequences

Oligo Name	Sequence (5'-3')
SA501	AATGATACGGCGACCACCGAGATCTACACATCGTACGTATGGTAATTGGCCTACGGGNGGCWGCAG
SA502	AATGATACGGCGACCACCGAGATCTACACACTATCTGTATGGTAATTGGCCTACGGGNGGCWGCAG
SA503	AATGATACGGCGACCACCGAGATCTACACTAGCGAGTTATGGTAATTGGCCTACGGGNGGCWGCAG
SA504	AATGATACGGCGACCACCGAGATCTACACCTGCGTGTTATGGTAATTGGCCTACGGGNGGCWGCAG
SA505	AATGATACGGCGACCACCGAGATCTACACTCATCGAGTATGGTAATTGGCCTACGGGNGGCWGCAG
SA506	AATGATACGGCGACCACCGAGATCTACACCGTGAGTGTATGGTAATTGGCCTACGGGNGGCWGCAG
SA507	AATGATACGGCGACCACCGAGATCTACACGGATATCTTATGGTAATTGGCCTACGGGNGGCWGCAG
SA508	AATGATACGGCGACCACCGAGATCTACACGACACCGTTATGGTAATTGGCCTACGGGNGGCWGCAG
SB501	AATGATACGGCGACCACCGAGATCTACACCTACTATATATGGTAATTGGCCTACGGGAGGCWGCAG
SB502	AATGATACGGCGACCACCGAGATCTACACCGTTACTATATGGTAATTGGCCTACGGGAGGCWGCAG
SB503	AATGATACGGCGACCACCGAGATCTACACAGAGTCACTATGGTAATTGGCCTACGGGAGGCWGCAG
SB504	AATGATACGGCGACCACCGAGATCTACACTACGAGACTATGGTAATTGGCCTACGGGAGGCWGCAG
SB505	AATGATACGGCGACCACCGAGATCTACACACGTCTCGTATGGTAATTGGCCTACGGGAGGCWGCAG
SB506	AATGATACGGCGACCACCGAGATCTACACTCGACGAGTATGGTAATTGGCCTACGGGAGGCWGCAG
SB507	AATGATACGGCGACCACCGAGATCTACACGATCGTGTTATGGTAATTGGCCTACGGGAGGCWGCAG
SB508	AATGATACGGCGACCACCGAGATCTACACGTCAGATATATGGTAATTGGCCTACGGGAGGCWGCAG
SA701	CAAGCAGAAGACGGCATACGAGATAACTCTCGAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
SA702	CAAGCAGAAGACGGCATACGAGATACTATGTCAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
SA703	CAAGCAGAAGACGGCATACGAGATAGTAGCGTAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
SA704	CAAGCAGAAGACGGCATACGAGATCAGTGAGTAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
SA705	CAAGCAGAAGACGGCATACGAGATCGTACTCAAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
SA706	CAAGCAGAAGACGGCATACGAGATCTACGCAGAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
SA707	CAAGCAGAAGACGGCATACGAGATGGAGACTAAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
SA708	CAAGCAGAAGACGGCATACGAGATGTCGCTCGAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
SA709	CAAGCAGAAGACGGCATACGAGATGTCGTAGTAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
SA710	CAAGCAGAAGACGGCATACGAGATTAGCAGACAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
SA711	CAAGCAGAAGACGGCATACGAGATTCATAGACAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
SA712	CAAGCAGAAGACGGCATACGAGATTCGCTATAAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT

Appendix 3. Mothur Pipeline

module load mothur/1.39.5

mother

make.contigs(file=XXXX.txt, processors=10)

summary.seqs(fasta=XXXX.trim.contigs.fasta, processors=10)

screen.seqs(fasta=XXXX.trim.contigs.fasta, group=XXXX.contigs.groups, summary=XXXXX.trim.contigs.summary, maxambig=12, maxlength=429, processors=10)

count.groups(group=XXXX.contigs.good.groups)

unique.seqs(fasta=XXXX.trim.contigs.good.fasta)

count.seqs(name=XXXX.trim.contigs.good.names, group=XXXX.contigs.good.groups)

align.seqs(fasta=XXXX.trim.contigs.good.unique.fasta, reference=silva.bacteria.fasta, processors=10)

summary.seqs(fasta=XXXX.trim.contigs.good.unique.align, count=XXXX.trim.contigs.good.count_table, processors=10)

screen.seqs(fasta=XXXX.trim.contigs.good.unique.align, count=dest2.trim.contigs.good.count_table, summary=XXXX.trim.contigs.good.unique.summary, start=6428, end=23444, maxhomop=6)

count.groups(count=XXXX.trim.contigs.good.good.count_table)

filter.seqs(fasta=XXXX.trim.contigs.good.unique.good.align, vertical=T, trump=.)

unique.seqs(fasta=XXXX.trim.contigs.good.unique.good.filter.fasta, count=dest2.trim.contigs.good.good.count_table)

pre.cluster(fasta=XXXX.trim.contigs.good.unique.good.filter.unique.fasta, count=XXXX.trim.contigs.good.unique.good.filter.count_table, diffs=2, processors=10)

chimera.vsearch(fasta=XXXX.trim.contigs.good.unique.good.filter.unique.precluster.fasta, count=XXXX.trim.contigs.good.unique.good.filter.unique.precluster.count_table, dereplicate=t)

remove.seqs(fasta=XXXX.trim.contigs.good.unique.good.filter.unique.precluster.fasta, accnos=dest2.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vsearch.accnos)

split.abund(fasta=XXXX.trim.contigs.good.unique.good.filter.unique.precluster.pick.fasta, count=XXXX.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vsearch.pick.count_table, cutoff=1)

count.groups(count=XXX.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vsearch.pick.abund.count_table)

classify.seqs(fasta=XXX.trim.contigs.good.unique.good.filter.unique.precluster.pick.abund.fasta, count=XXX.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vsearch.pick.abund.count_table, reference=trainset16_022016.rdp.fasta, taxonomy=trainset16_022016.rdp.tax, cutoff=80, method=wang, processors=10)

remove.lineage(fasta=XXX.trim.contigs.good.unique.good.filter.unique.precluster.pick.abund.fasta, count=XXX.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vsearch.pick.abund.count_table, taxonomy=XXX.trim.contigs.good.unique.good.filter.unique.precluster.pick.abund.rdp.wang.taxonomy, taxon=Chloroplast-Mitochondria-unknown-Archaea-Eukaryota)

cluster.split(fasta=XXX.trim.contigs.good.unique.good.filter.unique.precluster.pick.abund.pick.fasta, count=XXX.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vsearch.pick.abund.pick.count_table,

taxonomy=XXX.trim.contigs.good.unique.good.filter.unique.precluster.pick.abund.rdp.wang.pick.taxonomy, splitmethod=classify, taxlevel=4, cutoff=0.03, processors=10)

make.shared(list=XXX.trim.contigs.good.unique.good.filter.unique.precluster.pick.abund.pick.opti_mcc.unique_list.list,
count=XXX.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vsearch.pick.abund.pick.count_table, label=0.03)

classify.otu(list=XXX.trim.contigs.good.unique.good.filter.unique.precluster.pick.abund.pick.opti_mcc.unique_list.list,
count=XXX.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vsearch.pick.abund.pick.count_table,
taxonomy=XXX.trim.contigs.good.unique.good.filter.unique.precluster.pick.abund.rdp.wang.pick.taxonomy, label=0.03)

count.groups(count=XXX.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vsearch.pick.abund.pick.count_table)

sub.sample(shared=XXX.trim.contigs.good.unique.good.filter.unique.precluster.pick.abund.pick.opti_mcc.unique_list.shared, size=ZZZZ)

dist.seqs(fasta=XXX.final.fasta, output=lt, processors=10)

collect.single(shared=XXX.final.subsample.shared, calc=chao-invsimpson-shannon-npshannon, freq=100)

summary.single(calc=nseqs-sobs-chao-ace-invsimpson-npshannon-coverage-shannon)

quit

fasttree -gtr -nt XXXX.final.fasta > XXXX.final.tre

unifrac.weighted(tree=XXXX.final.tre, name=dest2.final.names,
group=XXXX.final.groups, distance=square, processors=10, random=F,
subsample=ZZZZ)

get.oturep(phylip=XXXX.final.phylip.dist, list=XXXX.final.list,
XXXX=Dave1.final.fasta, label=0.03)

alternatively use below commands if the above crashes:

get.oturep(fasta=xx.final.fasta,
count=xx.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vsearch.pick.abund.pick.count_table, list=xx.final.list, label=0.03, method=abundance)

Use fasta file made by get.oturep command to generate species identity using usearch.

1. Exit Mothur
2. Transfer fasta file from HIVE to local computer
3. Edit file to remove tabs and -. (Select all text(Ctrl A), go to replace tool, \t for tab replace with _, replace all, use extended option)
4. Rename file using .edited in file name
5. Transfer file to HIVE directory
6. Ensure rdp_download_453650seqs_newnames.fa reference file is in working directory
7. Use command line: usearch -usearch_global je.final.reps_edited --db rdp_download_453650seqs_newnames.fa --uc je.usearch_dd_97%_3.txt --id 0.97 --maxaccepts 3 --maxrejects 0 --strand plus -threads 10
   (Note: yellow = input file name, blue = output file to be created, these need to be specified, grey = reference database)

Appendix 4. R Script for Visualisation of Sequencing Data

```
setwd("/Volumes/UNI DRIVE/External Hard Drive/Stool Microbiome Analysis")

library("phyloseq")

library(vegan)

library(ggplot2)

otu=as.matrix(read.table(file.choose(), header=TRUE, check.names = FALSE))

tax=as.matrix(read.table(file.choose(), header=TRUE))

meta=as.data.frame(read.table(file.choose(), header=TRUE))

PCoA <- phyloseq(otu_table(otu, taxa_are_rows = TRUE), tax_table(tax),
sample_data(meta1))

PCoA.rel <- transform_sample_counts(PCoA, function(x) 100 * x/sum(x))

PCoA.rel.ord <- ordinate(PCoA.rel, "PCoA", "bray")

PCoA.rel.ord.figure = plot_ordination(PCoA.rel, PCoA.rel.ord, type="samples")

PCoA.rel.ord.figure.baby
```

Note: To create an NMDS plot the PCoA commands were replaced with NMDS.

```
library(beeswarm)

setwd("C:/Users/c1563648/OneDrive - Cardiff University/PhD/3rd Year PhD/Stool
Microbiome Analysis/Files for Analysis")

BSwarm.Diversity = read.table(file.choose(), header=TRUE)

boxplot(Number_of_Sequences ~ Sample, data = BSwarm3, outline = FALSE, main =
'boxplot + beeswarm')

beeswarm(Number_of_Sequences ~ Sample, data = BSwarm3, col = 4, pch = 16, add =
TRUE)
```

```
BSwarm.Plot3 = boxplot(Number_of_Sequences ~ Sample, data = BSwarm3, outline =
FALSE, main = 'Number of Sequences Produced by Illumina MiSeq 16S Sequenicng',
xlab="Samples and Controls", ylab="Number of Reads")

BSwarm.Plot3 = beeswarm(Number_of_Sequences ~ Sample, data = BSwarm3, col =
4, pch = 16, add = TRUE)

BSwarm.Plot3 = beeswarm(Number_of_Sequences ~ Sample, data = BSwarm4, pwcol
= as.factor(Sample), pch = 16, add = TRUE)

BSwarm.Plot3 = beeswarm(Number_of_Sequences ~ Sample, data = BSwarm4, col =
3, pch = 16, add = TRUE)
```