The Gut Mucosal Microbiome of HIV- exposed Uninfected Infants in Africa

Phumudzo Khomunala
(KHMPHU003)

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN

In fulfilment of the requirements for the degree:

MSc (Med) Chemical Biology

Supervisor: Prof. Jonathan Blackburn

Co-Supervisor: Dr S Fortuin

Institute of Infectious Disease and Molecular Medicine (IDM)

Division of Chemical and Systems Biology

Department of Integrative Biomedical Sciences

Faculty of Health Sciences

University of Cape Town

The financial assistance of The National Research Foundation towards this research is hereby acknowledged. The opinions and conclusions are those of the author and are not to be attributed to any of the above-mentioned parties

Declaration

I, Phumudzo Khomunala , hereby declare that the work on which this thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university. I authorise the University to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signed by candidate

Signature: . Date: ……07/02/2020…

Abstract

Background: South Africa has a large HIV disease burden, with the highest rate of infection occurring in young women of the childbearing age. This gave impetus to the Prevention-of-Mother-To-Child-Transmission (PMTCT) program, that has been successfully implemented. Due to the success of the PMTCT program, HIV-exposed-uninfected (HEU) infants represent a growing population in South Africa. However, these infants have been found to have increased morbidity and mortality rates compared to their HIV Unexposed Uninfected (HUU) peers, as well as altered immune and vaccine responses. The reasons for this remain unclear, but one hypothesis is that altered gut microbiomes in HEU adversely affect the developing infant immune system. The microbiome (a collection of an array of microorganisms, their genes, genomes, proteomes, and metabolites) is an area of emerging research interest; dysbiosis of the gut microbiome has recently been associated with disease outcomes and progression in several disease areas. The microbial colonisation of the infant begins in utero and continues after birth. It is affected by several factors: birth mode, age of gestation, feeding mode, maternal health status as well as environmental factors.

Aim: To elucidate the microbiomes of HEU infants in Africa, compared to HUU controls

Design: Ultra-high-performance liquid chromatography mass spectrometry was used to analyse and characterize a subset of existing stool samples stored from the InFANT cohort study. The infant gut metaproteome of 34 HEU versus 29 HUU infants, from the South African arm of the study was analysed. Cross-sectional samples were collected and analysed at two-time points, namely at birth and within the first week of life (between 4 to 7 days after birth).

Results: Comparative analysis of the HEU and HUU reveal differences in the microbial composition between the two groups at birth and day 4-7, with the most apparent difference occurring at birth. In our comparison we found that the relative abundances of Bacteroidetes and Firmicutes were different between the HEU and HUU at both birth and day 4-7. There was a dramatic shift in the microbial composition within the first week of life.

Conclusion: It is evident from our analysis that the HEU infant has a different gut microbiome to that of the HUU infant at birth. The HEU microbiome is characterised by a high microbial diversity at birth. This could be associated with more severe outcomes from childhood ailments. The human breast milk (HBM) microbiome greatly influences and mitigates the differences upon subsequent breastfeeding, but differences in the measured microbiomes of HEU and HUU nonetheless remain.

Recommendations: A longitudinal study should be carried out to better monitor the long-term effects of the microbiome on infant immune priming. A study of the HBM microbiome should also be investigated to better understand the role of HBM in mediating and priming the infant's immune system. Further, a study of the metabolome of the infant gut and the matching HBM of the mother may identify potential metabolites that could be used as biomarkers for vaccine responses.

Acknowledgements

I would like to thank my parents for all their unconditional support and encouragement. Thank you for the prayers and words of encouragement. I would like to specially thank my wombmate, Avhasei for letting me chew off her ear when times were tough.

A special thank you to my friends, Samantha for the cheerleading and prayers and encouragement, thank you to Thando for always reminding me to eat and investing time energy and caffeine into this pursuit. Thank you to my partner Samu for the encouragement and check-ins

Thank you to my supervisors Prof. Jonathan Blackburn for the opportunity and freedom, thank you to my co-supervisor Dr. Suereta Fortuin. I would also like to thank our collaborators: Namely Dr. Jerome Wendoh, Matthys Potgieter, Dr. Andrew Nel, Dr Heather Jaspan, and all the members of the project who made this possible.

A big thank you to Imane ,Bridget and all the members of the Blackburn Lab group, for the laughs, cake, coffee and safe space to work, I enjoyed your company and all the intellectual debate and conversations in the space. It made for an enjoyable, productive workspace.

A special thank you in memory of my Auntie Ruth, thank you for the guidance and love while you were still with us

Table of contents

Declaration 1

Abstract 2

Acknowledgements 3

Table of contents 4

List of abbreviations 6

List of tables 9

List of figures 9

Introduction 11

Chapter 1. Literature review 14

1.1 The microbiome 14

1.2 Microbiome analysis tools 20

Chapter 2. Methodology 28

2.1 Selection Criteria and ethical considerations 28

2.2 Stool homogenization and Protein Extraction 31

2.3 Protein quantification and Digestion 32

2.4 Desalting and purification 32

2.5 Liquid Chromatography-Mass Spectrometer analysis 32

2.6 Data Analysis 33

2.7 Taxonomic analysis 34

2.8 Functional analysis 35

2.9 Protein analysis 35

Chapter 3. Results 37

3.1. MS data quality 37

3.2. Taxonomic identification 41

3.3 Functional analysis of metaproteomic data 63

3.4 Protein analysis 66

Chapter 4. Discussion 79

4.1. Taxonomic analysis of metaproteomic data 79

4.2. Functional analysis of Metaproteomic data **83**

4.2.1 Peptide analysis **83**

4.2.2 Protein analysis of metaproteomic data **84**

4.3. Limitations **86**

Chapter 5. Conclusion and recommendations **88**

References **89**

Appendix **104**

List of abbreviations

A:	Solvent A
ABC:	Ammonium Bicarbonate
ACN:	Acetonitrile
AIDS:	Acquired immune deficiency syndrome
B:	Solvent B
BCA:	Bicinchoninic acid
BSA:	Bovine serum albumin
buffer A:	0.1% FA/dH2O
buffer B:	0.1%FA/ACN
C 18:	Octadecyl bonded silica
CSV:	Comma separated values
CV:	Coefficient of variation
Da:	Dalton
DDA:	Data dependent acquisition
dH2O:	Deionised water
DTT:	Dithiothreitol
ESI:	Electrospray ionisation
FA:	Formic acid
FDR:	False discovery rate
FF:	Exclusive formula feeding
HAART:	Highly Active Antiretroviral therapy
HBM:	Human breast milk
HCD:	High energy collision dissociation
HEU:	HIV-exposed uninfected
HIV:	Human immunodeficiency virus
HPLC:	High performance liquid chromatography
HUU:	HIV-Unexposed uninfected
IAA:	Iodoacetamide
IDM:	Institute of Infectious Diseases and Molecular Medicine
IEF:	Isoelectric focusing
Ig:	Immunoglobulin
IgG:	Immunoglobulin G
IgG1:	Immunoglobulin G subclass 1
IgG3:	Immunoglobulin G subclass 3

IgG4:	Immunoglobulin G subclass 4
IT:	Injection time
kDa:	Kilodalton
LC:	Liquid chromatography
LC-MS/MS:	Liquid chromatography-tandem mass spectrometry
LFQ:	Label-free quantitation
m/z:	Mass-to-charge ratio
MALDI:	Matrix-assisted laser desorption ionisation
Max IT:	Maximum injection time
MS:	Mass spectrometry
MS1:	Mass spectrometry at peptide ion level
MS1-TIC:	Total MS1 ion chromatogram
MS2:	Tandem mass spectrometry (at fragment ion level)
MS2-TIC:	Total MS2 ion chromatogram
mV:	Millivolt
MWCO:	Molecular weight cut-off
Na^+:	Sodium ion
NCE:	Normalised collision energy
ESI:	Electrospray ionisation
LC:	liquid chromatography
NGS:	Next generation sequencing
PBS:	phosphate buffered saline
PEP:	Posterior error probability
PI:	Principal investigator
ppm:	Parts per million
PSM:	Peptide spectral match
PTM:	Post translational modification
PMTCT:	Prevention of mother to child transmission
Q1:	Quadrupole 1
Q2:	Quadrupole 2
Q3:	Quadrupole 3
QIT:	Quadrupole ion trap
QTOF:	Quadrupole time-of-flight
RPLC:	Reversed-phase liquid chromatography
SA:	South Africa

SCX:	Strong cation exchange
SDS:	sodium dodecyl sulphate
SDS-PAGE:	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEM:	Standard error of the mean
SRM:	Single reaction monitoring
TB:	Tuberculosis
TFA:	Trifluoroacetate
TIC:	Total ion chromatogram
TOF:	Time-of-flight
TSQ:	Triple stage quadrupole
TSV:	Tab separated values
uHPLC:	Ultra-high-pressure liquid chromatography
UV:	Ultraviolet
WHO:	World Health Organisation

List of tables

Table 1: Table of different methods applied to microbiome studies. **26**
Table 2: Eligibility criteria for sample recruitment in InFANT study **29**
Table 3: Clinical information of InFant mother diads recruited for Study. **30**
Table 4: BSA concentration of samples **37**
Table 5: Numbers of peptides identified for each class **44**
Table 6: Number of peptides identified that are associated with Mycobacterium Tuberculosis **59**
Table 7*:* Parasitic phyla identified **59**
Table 8: Top 10 Biological process for Chordata peptides identified on Unipept online in HEU birth. **63**
Table 9: Top 10 Biological processes for Chordata peptides identified on Unipept online HUU birth. **64**
Table 10: Top 10 Biological processes for Chordata peptides identified on Unipept online HEU day 4-7 **64**
Table 11: Top 10 Biological processes for Chordata peptides identified on Unipept online HUU day 4-7. **65**
Table 12: Top 10 protein classes identified for Birth HEU infant group **67**
Table 13: Top 10 protein classes identified for Birth HUU infant group **68**
Table 14: Top10 Protein Classes identified Day4-7_HEU **70**
Table 15: Top 10 protein class identified Day4-7_HUU **72**
Table 16: Top 3Go enrichment Biological process exclusive Birth HEU **77**
Table 17: Top 10 Go enrichment Biological process exclusive to Birth HUU **77**
Table 18: Top 10 Biological process exclusive to day4-7HEU **77**
Table 19: Top 7 Biological process exclusive to day4-7HUU **78**
Table 20: Top 10 Biological processes common in all four groups. **78**

List of figures

Figure 1: Physiologic adaptations of the body during pregnancy that may favour an increased bacterial translocation **19**
Figure 2: Factors influencing the colonization of the infant gut. **20**
Figure 3: Schematic diagram of the Q-exactive Mass spectrometer **23**
Figure 4: Flowchart of the workflow followed for sample processing and analysis **31**
Figure 5: Bar graph of missed cleavages generated from summary script in R, displays the quality of the digestion of the trypsin. **40**
Figure 6: Chromatograms generated from samples run on the Q-Exactive mass spectrometer **41**
Figure 7: Venn diagram of the Peptides that are shared amongst the different groups at both timepoints. **42**
Figure 8: bacterial metaproteome based on Phyla in the: HEU vs HUU at Birth and Day4-7. **42**
Figure 9: Stacked bar graph indicating the relative abundance of the metaproteomic classification based on class of bacteria in HEU and HUU infants at both time points **43**
Figure 10: Bacterial class abundance **45**
Figure 11: Sunburst chart generated on Unipept providing an overview of the organism identified HUU at birth. **47**
Figure 12: Expanded sunburst chart of the HUU microbiota birth, detailing the bacteria identified. **48**
Figure13: Treemap generated on Unipept of the organisms identified in HUU birth infants **49**

Figure 14: Sunburst chart generated on the Unipept platform providing an overview of the organism identified HEU birth. **50**
Figure 15: Expanded sunburst chart of the HEU birth **51**
Figure 16: Tree map generated of the organisms present in the Birth HEU infants **52**
Figure 17: Sunburst chart providing an overview of the organisms identified in D4-7 HUU infants **53**
Figure 18.: Expanded sunburst chart of the D4-7 HUU infants, details on the bacterial composition **54**
Figure 19: Treemap of the organisms identified in D4-7 HUU. **55**
Figure 20: Sunburst chart providing an overview of the organism identified in the HEU. D4-7 **56**
Figure 21: Expanded sunburst chart of the HEUD4-7 microbiota 57
Figure 22: Treemap of the organisms present in the D4-7 HEU infants. **58**
Figure 23: PCA plot of the samples clustering according to the phyla present within both groups at each timepoint. **60**
Figure 24: Hierarchical cluster chart of the all expressed classes of organisms in the HEU and HUU at birth **61**
Figure 25: Hierarchical cluster chart of all the expressed classes of organism in HEU and HUU at Day 4-7 **62**
Figure 26: All protein classes identified in the Birth HEU infant group **66**
Figure 27: All protein classes identified for Birth HUU infant group **68**
Figure 28: All protein classes identified for Day 4-7 HEU infant group **70**
Figure 29: All protein classes identified for Day 4-7 HUU infant group **71**
Figure 30: Volcano plots generated on Perseus (a)Birth HEU vs HUU (b) day 4-7 HEU vs HUU **73**
Figure 31: Venn diagram of the protein groups that overlap at the different timepoints (a) Birth timepoint (b) Day 4-7 timepoint**.** **75**
Figure 32: Venn diagram of all the Human proteins shared in HEU and HUU at both the Birth and Day 4-7 timepoint **76**

Introduction

HIV in South Africa

The trajectory of human immunodeficiency virus (HIV) infection in South Africa is different to other parts of the world, due to the socio-economic history in the country (Loewenson *et al.*, 1997; Hargreaves *et al.*, 2007; Weiser *et al.*, 2007; CDC, 2016). In other parts of the world the majority of HIV infections occurred between men who have sex with men (MSM) or through the use of injectable drugs (El-Sadr, Mayer and Hodder, 2010; Aids, 2018), whereas the onset of the epidemic within the South African context saw a rapid rise in the rate of HIV infection and transmission among women of reproductive age between 25-35 years old (Williams *et al.*, 2000). This further contributed to an increase in the number of HIV-infected and –exposed (HEU) infants born in South Africa. Hesseling and colleagues observed that these HEU infants were immunocompromised at birth and suffered adverse side effects from routine vaccinations such as BCG (Hesseling *et al.*, 2007, 2009). These findings highlighted that we clearly do not understand the impact of HIV exposure and of Highly Active Antiretroviral Therapy (HAART) on immune development in HEU infants in utero (Garcia-Knight *et al.*, 2015).

In 2002, the South African government implemented a Prevention-of-Mother-to-child-transmission (PMTCT) program, which has grown into one of the world's largest PMTCT programs (Goga *et al.*, 2017). Since its initiation the program has undergone several changes and the policy has been amended (Goga *et al.*, 2017). In a bid to reduce the number of HIV positive babies born, the government changed the healthcare policies and mandated HIV testing for all pregnant women receiving prenatal care from a government facility (Doherty *et al.*, 2010, 2011). Since 2013, if the mother tested HIV positive, she is placed on lifelong HAART treatment regardless of her CD4 count. The successful implementation of this intervention reduced the number of vertical transmission from mother-to-child by 78% (Burton, Giddy and Stinson, 2015; Sherman *et al.*, 2017). As a result

of this, however, the number of HIV-exposed babies born in South Africa increased exponentially over the last few years (Anon, 2016).

Despite the success of the program in preventing vertical transmission of the HIV from mother-to-child, HEU infants are still experiencing adverse outcomes i.e. low birthweight, increased morbidity and mortality due to opportunistic infections, compared to their HIV-unexposed-uninfected (HUU) peers (Ramokolo *et al.*, 2014; Hofer *et al.*, 2016; le Roux *et al.*, 2016; Smith *et al.*, 2017). Some of these poor health outcomes can be attributed to recommendations that were initially introduced by the World Health Organisation (WHO) and the South African Department of Health and the National AIDS Council to exclusively formula feed (FF) HEU infants ,amongst a number of factors (*HIV and infant feeding Guidelines on*, 2010; Sherman *et al.*, 2017). In South Africa, most HIV positive mothers live in resource poor conditions and thus formula feeding is expensive, and the WHO recommendations did not consider the socio-economic status of these mothers. Therefore a large proportion of the poor health outcomes experienced by these HEU infants could be attributed to bad feeding practices, resulting in increased diarrheal diseases and infections such as oral thrush (le Roux *et al.*, 2016). Issues related to access to clean water and sanitation are not unique to South Africa and indeed similar disease trends were also observed in other low-income to middle income countries with high HIV disease burden thus leading to increased diarrheal illness and malnourishment (Powis *et al.*, 2011; Gibb *et al.*, 2012; Arikawa *et al.*, 2016). A policy change in 2009 by the WHO encouraged exclusive breastfeeding practices for HIV exposed/positive infants, regardless of the mothers HIV status, for the first 6 months of the infants life (Natchu *et al.*, 2012). This recommendation greatly reduced the morbidity and mortality of HEU infants due to bad feeding practices (Wood *et al.*, 2017): during the first 6 months of life, protection against diarrhoea was substantially greater (odds ratio 6.1 [4.1-9.0]) than against deaths due to acute respiratory infection with exclusive breast feeding (Victora and Barros, 2000). Although the recommended

feeding practices have changed, only 50% of South African HIV positive mothers are choosing to exclusively breast feed for the first six months of the infant's life, whilst the remainder are choosing to continue mix feeding practices of a combination of breastfeeding, formula feeding and solid foods (Coutsoudis, 2005; Kuhn *et al.*, 2015).

HEU infants have been reported to have increased opportunistic infections such as lower respiratory infection, pneumonia and gastroenteritis, with less successful outcomes to intervention and treatment by IV antibiotics during hospitalisation (Slogrove *et al.*, 2012, 2016; Adler *et al.*, 2015; Slogrove, Archary and Cotton, 2016, Passmore etal., 2017). Previous studies revealed that HEU infants have an altered innate and adaptive immune response, including the natural killer and TH4 memory cell functions (Smith *et al.*, 2017). This group of HEU infants present are thus an ever-increasing and uniquely vulnerable population that may place an enormous burden on South Africa's public health facilities..

Chapter 1. Literature Review

1.1 The Microbiome

According to Lederberg's definition the microbiome is an "ecological system of commensal, symbiotic, and perhaps pathogenic microorganisms that reside in the human body"(Lederberg and McCray, 2001). Lederg popularised this term in 2001 and it has since been used frequently to describe the microbial composition of an environmental sample.

Despite the technological advances in omics over the last decade which allows for an in-depth investigation of the microbiome, understanding the biological impact of the gut microbiome on disease remains in its infancy. Here the microbiome is defined as "a collection of an array of microorganisms their genes, genomes and metabolites, that have colonised the human body" (Pflughoeft and Versalovic, 2012). A large proportion of these microorganisms form an integral part of the first line of defence but were previously studied as pathogens believed to be the cause of disease (Pallen, 2011). When the microbes are in a symbiotic relationship the body functions optimally, however a dysbiosis can result in alterations in normal bodily functions (Pflughoeft and Versalovic, 2012; Cerea *et al.*, 2015). The microbiome has been linked to several disease states and the relationship between health and microbiome composition has recently been established as not only correlative but also causative (Turnbaugh *et al.*, 2006a; Round and Mazmanian, 2009). Illnesses like Irritable Bowel Syndrome (IBS) as well as Crohns disease have all been linked to a dysbiosis in gut bacterial composition. The microbiome and its involvement in disease progression is still not well understood causing gaps in the knowledgebase.

Background

The microbial colonisation of the human has recently been shown to begin in utero (Jeurink *et al.*, 2013). Until recently, it was thought that the infant developed in a sterile environment within the amniotic sac. However, new evidence suggests that the amniotic sac is not sterile, as evidence of bacteria in the meconium of healthy new-borns has been found (Jeurink *et al.*, 2013; Aagaard *et al.*, 2014; Wassenaar and Panigrahi, 2014; Perez-Muñoz *et al.*, 2017). Currently there is no consensus on the theory of a non-sterile womb. Recently, Rodriguez and colleagues investigated the reagents that were used in the first study to

report the presence of microbes in the womb and the reagents themselves could give rise to a DNA signal thought to be due to a contaminant, termed the "kit-ome", that results in DNA signals being picked up (Leiby *et al.*, 2018; Lim, Rodriguez and Holtz, 2018; Theis *et al.*, 2019). Salter and colleagues, highlighted the possibility of DNA contamination of samples with a low biomass by various sources such as molecular biology grade water, PCR reagents and DNA extraction kits themselves (Salter *et al.*, 2014).The argument in favour of the theory of a sterile womb is the successful delivery of germ free infants in several species including humans (Kirk, 2012; Leiby *et al.*, 2018). It remains possible therefore that the DNA being recorded may just be contaminant DNA (Leon *et al.*, 2018).

During the birthing process, the mode of delivery influences the microbial colonisation and composition of the infant (Armanian *et al.*, 2016). Furthermore after birth the microbial colonisation of the infant gut continues and is influenced by mode of feeding (Armanian *et al.*, 2016). The gastrointestinal tract is a large component of the human immune system (Wittig and Zeitz, 2003) and is considered as an extra organ. Therefore, its modulation can have a severe impact on disease progression and outcome (Kinross, Darzi and Nicholson, 2011). Several studies suggest that the human gut microbiome interacts with the human immune system and in turn influences immune function thereby affecting disease progression of many illnesses ranging from infectious diseases to non-communicable diseases (Turnbaugh et al. 2006b; Jeurink et al. 2013; Magne et al. 2017; Dietert 2017).

The microbiome in healthy individuals

The human microbiome project (HMP) was launched in 2007 with the aim of characterizing the microbial composition of a healthy human microbiota and help generate a better understanding the human microbe interactions (NIH HMP Working Group *et al.*, 2009). Interestingly, the HMP has enabled the classification and investigation of the microbiomes of various sites on healthy individuals (Turnbaugh *et al.*, 2007). The data collected revealed that the microbial population was inter and intra-variable suggesting that the microbial population varied among different habitats within the same host (Grice *et al.*, 2009; Pallen, 2011). Further, the same habitats in different hosts have different microbial composition albeit each of the habitats had signature phyla present although in varying degrees (Costello *et al.*, 2009; Grice *et al.*, 2009; Caporaso *et al.*,

2011; Ursell *et al.*, 2012). Despite the taxonomical differences, signature bacterial species/phyla were associated with specific functional and metabolic pathways (Huttenhower *et al.*, 2012). This suggests that a large microbial variation is to be expected within healthy controls. On the contrary, diseased conditions had a dysbiosis with very low microbial diversity as well as the presence of pathogenic bacteria (Li *et al.*, 2016; Pammi *et al.*, 2017). Furthermore, no pathogenic species of bacteria were identified in any of the healthy individuals suggesting that the presence of a pathogenic bacteria (organism) could be associated with a disease state (Grice *et al.*, 2009; Consortium *et al.*, 2012). The inter-individual variation could be attributed to functional relevance within the host and the environment (Consortium *et al.*, 2012). The HMP has defined the microbiome of a healthy individual as reference point when studying the microbiome and its dysbiosis. Recent findings have revealed that although taxonomical variation may exist the functional metabolic pathways that are presented remain broadly the same (The Human Microbiome Consortium Project,2012). Various areas of the human body have been studied: the skin, mouth and the gut. Each of these areas possesses various microenvironments. The skin, the organ most exposed to interaction with the environment and microbes (Fredricks, 2001; Kong *et al.*, 2012), has different microenvironments with different microbiome compositions (Gao *et al.*, 2007; Costello *et al.*, 2009; Grice *et al.*, 2009; Kong *et al.*, 2012). For example, the armpits are characteristically moist and dominated by *Corynebacterium*, whilst dry areas like the knees and elbows are dominated by *β-Proteobacteria* and *Flavobacteria* (Grice *et al.*, 2009). The oral microbiome which includes the mouth also has distinct areas such that the microbiome between the gums and the teeth is different to that found on the buccal cavity (Paster *et al.*, 2006; Zaura *et al.*, 2009; Dewhirst *et al.*, 2010). Often the microbes such as *Streptococcus, Fusobacterium, Actinomyces, Veillonella,* and *Treponema* studied in the mouth have been associated with a pathogenesis (gum disease, tonsillitis and cavities) (Zaura *et al.*, 2009). It has been known that the mouth harbours microbes, but only recently have "healthy" mouth microbes been discussed (Dewhirst *et al.*, 2010). A 2009 study by Zaura *et al.* revealed the possibility of a core oral microbiome in healthy individuals. The study showed that there was a 72% overlap in the taxa of the microbes that were isolated from the three healthy Caucasian male adults (Zaura *et al.*, 2009).

Whilst each area has a distinct microenvironment a pattern exists within each area that defines what a "healthy" microbiota is. For example, a healthy vaginal microbiome is characterised by low diversity with an abundance of *Lactobacillus* (Ling *et al.*, 2010; Jacques Ravel, Gajer, Abdo, Schneider, Koenig, Stacey L McCulle, *et al.*, 2011; Ma, Forney and Ravel, 2012). The composition of the vaginal microbiome has been profiled and arranged into five groups by abundance/dominance: group I is *Lactobacillus crispatus* dominant, with group II being *Lactobacillus gasseri* dominant, group III is *Lactobacillus iners* dominant and group V is *Lactobacillus jensenii* dominant. However group IV is a heterogeneous group of strict anaerobes (J. Ravel *et al.*, 2011). Interestingly a low diversity was not always the case for a healthy vaginal microbiota, as differences in microbial composition have been found amongst women of different ethnic groups. Women with African-American ancestry have a more diverse bacterial composition compared to women with European ancestry (Jacques Ravel, Gajer, Abdo, Schneider, Koenig, S. L. McCulle, *et al.*, 2011; Fettweis *et al.*, 2014).

To date, the gut microbiome has been the most studied due to its size and impact on health. A dysbiosis in the gut microbiome has been associated with a range of ailments. Diseases such as diabetes, a metabolic dysbiosis, have been associated with the relative abundance of Firmicutes been found to be lower, whilst the proportion of Bacteroidetes and Proteobacteria was higher in diabetic persons compared to non-diabetics (Larsen *et al.*, 2010). The impact of the gut microbiome on inflammatory bowel diseases such as Crohn's disease and ulcerative colitis has also been investigated. Cohn's disease is characterised by chronic inflammation of the lining of the digestive tract which can lead to abdominal pain (Gevers *et al.*, 2014). Recent studies have revealed a link between the disease states, host genetics, the gut microbiome and alterations in specific metabolomics pathways (Fung *et al.*, 2012; Gevers *et al.*, 2014). Studies of the intestinal gut microbiota imply that an unbalanced microbial community composition is associated with a dysregulated immune response, suggesting the role of the microbiome in pathogenesis (Fung *et al.*, 2012).

Studies based on Crohn,s disease phenotype showed a strong positive correlation with clinical disease severity and negative correlation with species richness. This suggests that a severe disease state manifests with severely reduced species diversity in favour of a dysbiosis (Gevers *et al.*, 2014). The dominant species

in Crohn's disease were *Escherichia coli, Eikenella corrodens,Haemophilus parainfluenzae (Pasteurellaceae), , Fusobacterium nucleatum Veillonella parvula, (Neisseriaceae) and Gemella moribillum* (Huttenhower *et al.*, 2012; Morgan *et al.*, 2012b; Sheehan and Shanahan, 2017). Whereas the species that decreased in Crohn's disease *were Bacteroides vulgatus, Bacteroides caccae, Bifidobacterium bifidum, Bifidobacterium longum, Bifidobacterium adolescentis, Bifidobacterium dentum, Blautia hansenii, Ruminococcus gnavus, Clostridium nexile, Faecalibacterium prausnitzii, Ruminoccus torques, Clostridium bolteae, Eubacterium rectale, Roseburia intestinalis,* and *Coprococcus* (Morgan *et al.*, 2012a; Gevers *et al.*, 2014).

Interestingly, the microbial population showed less variety in the stool samples than in the mucosal tissue samples (Gevers *et al.*, 2014; Slingerland *et al.*, 2017). This highlights a disadvantage with using faecal samples for microbiome studies, however this sample type is the most easily, non-invasive retrievable sample type.

The Microbiome in pregnant women

Many changes occur in a woman's physiology during pregnancy and lactation, which affect virtually all systems, including the cardiovascular, respiratory, genitourinary and digestive systems. Interestingly, such adaptations may favour an increased bacterial translocation during late pregnancy and early lactation (Figure 1). The main effects of gestation on the digestive tract are associated with the displacement of the abdominal organs by the growing uterus, decreased motility and a large increase in the size and complexity of the maternal intestine (Jimenez et al ,2008). Overall, the digestive tract is characterized by weakened barriers against bacterial growth, increased permeability, and reduced peristalsis, these three factors that are closely associated with bacterial translocation (Perez *et al.*, 2007; Urbaniak *et al.*, 2016).

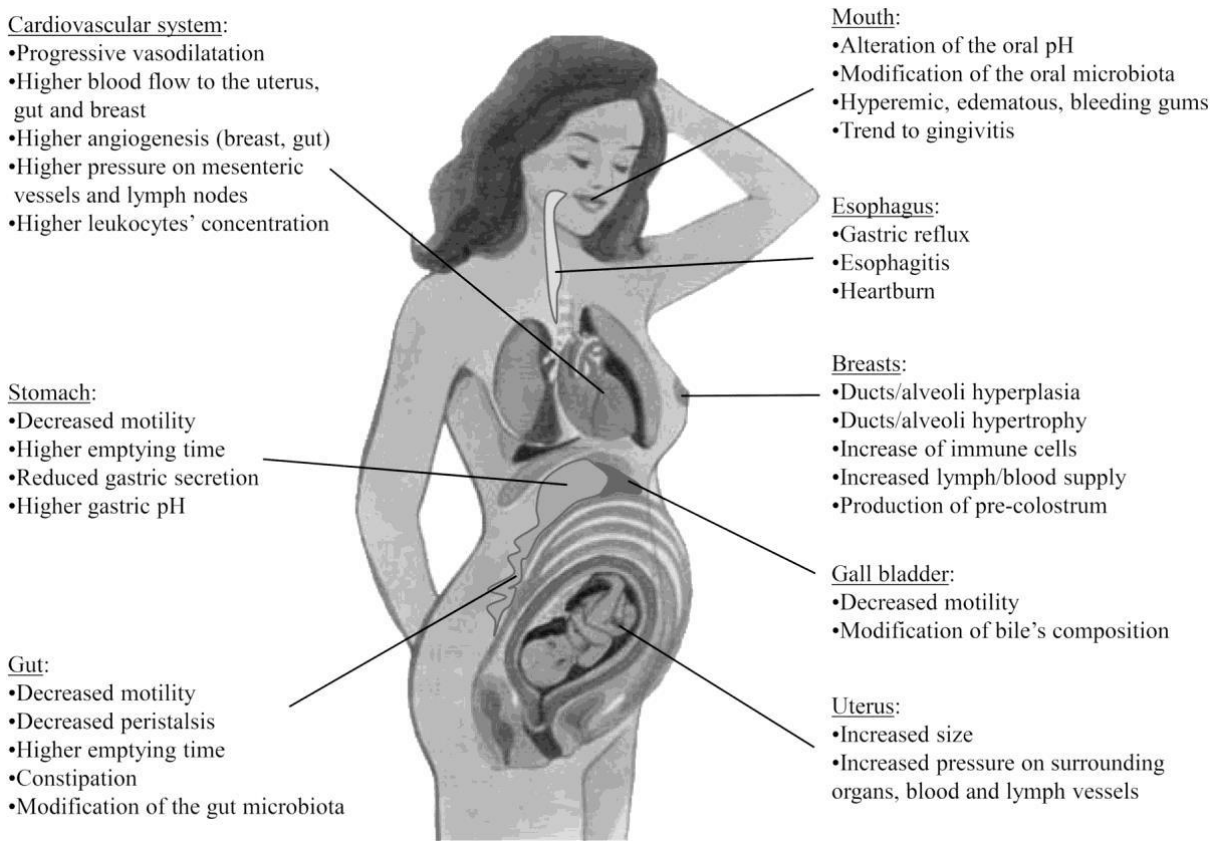


Figure 1. Physiologic adaptations of the body during pregnancy that may favour an increased bacterial translocation (Rodríguez, 2014).

There are several factors that influence the gut microbiome of the infant (Figure 2). Colonization of the infant gut begins in-utero and continues during birth. Babies delivered vaginally will have a gut microbiome similar to that of their mothers' vaginal microbiomes including *Lactobacillus* and *Bifidobacterium* species (Kumar *et al.*, 2016; Urbaniak *et al.*, 2016), while those babies delivered by a caesarean will have a gut microbiome similar to the skin of their mothers, with lower numbers of *Bifidobacteria* and *Bacteroides*, and more often colonized with *Clostridium difficile (C.difficile)* compared with vaginally born infants (Dominguez-Bello *et al.*, 2010, 2013; Song, Dominguez-Bello and Knight, 2013). Infants born by elective caesarean delivery were shown to have a particularly low bacterial richness and diversity (Dominguez-Bello *et al.*, 2013) .

After birth, the mode of feeding further affects the infant gut microbiome composition (Le Huërou-Luron, Blat and Boudry, 2010; Fan *et al.*, 2013; Mueller *et al.*, 2015). Formula-fed infants have been shown to have a significantly different gut microbiome mainly dominated by *Bifidobacterium fragilis (B. fragilis)* compared

to breast-fed infants dominated by *Bifidobacterium infantis* (*B.infantis)* (Le Huërou-Luron, Blat and Boudry, 2010; Jost *et al.*, 2013; Victora *et al.*, 2016) .

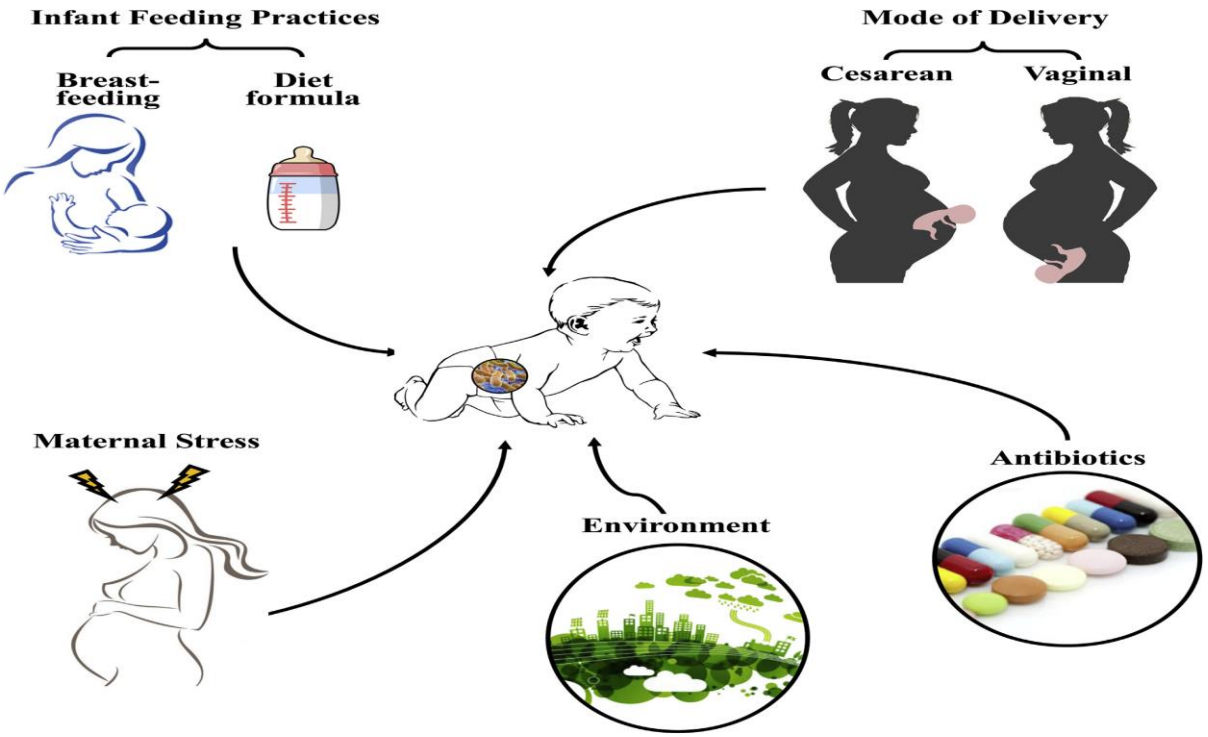


Figure. 2. Factors influencing the colonization of the infant gut. Widely performed perinatal interventions such as mode of delivery, infant feeding practices, antibiotic usage and environment affect the colonization of the infant gut, Other factors such as maternal stress during pregnancy, gestational age, genetics, and infections (pre- and/or postnatal) also influence the microbial composition of the infant gut microbiota(Heijtz, 2016)

1.2 Microbiome analysis tools

Until recently microbiome studies focused on the microbial composition, with an emphasis on the bacterial component as methods have already been established to isolate and culture certain bacteria. The methods to study the microbiome have since expanded to include non-culture dependant methods.

Culture dependent techniques

The microbiota has been traditionally studied using culture-dependent techniques. This method is cheap, easy, quick and offered high taxonomical discrimination (Vaz-Moreira *et al.*, 2011). It also gives information

on viability of the strains which culture-independent microbiological techniques do not. However, culture dependent techniques are limited as a large proportion of the organisms found in environmental samples are fastidious and do not thrive in culture (Jost *et al.*, 2013). This led to the need to adopt different methods to analyse the microbiota in samples.

Culture Independent Tools

Metagenomics

The recent advancement and development of genomic technologies has given rise to the ability to study the microbial profile in samples without the use of culture. Massively parallel sequencing, also called Next-generation sequencing (NGS), targets the bacterial 16S rRNA gene allowing for an in-depth insight into the bacterial community. The 16S rRNA region in bacteria is ideal for this purpose, as it is a well conserved region of the bacterial genome that is able to provide species specific identification of bacteria due to its variable regions (De Filippo *et al.*, 2010; Ghanbari, Kneifel and Domig, 2015). The use of NGS allows for the identification of fastidious microbes that do not thrive in culture. One of the objectives of the MetaHIT (Metagenomics of the Human Intestinal Tract) project was to compile a human gut microbial catalogue using metagenomic sequencing (Qin *et al.*, 2010). This study revealed that the highest abundance of phyla observed were in *Bacteroidetes* and F*irmicutes. The* study also revealed clusters of bacteria at the genus /family levels. *Bacteroidetes and Dorea/Eubacterium/Ruminococcus* groups well as *BIifidobacteria/Proteobacteria* and *streptococci/lactobacilli* groups. These findings revealed that similar groups of bacteria may be shared amongst individuals (Qin *et al.*, 2010; Morgan, Segata and Huttenhower, 2013).

However, this technique is not without its drawbacks, the most notorious being PCR bias (Xuan *et al.*, 2013). The overall bacterial community representation may be skewed depending on the bias introduced using the species-specific primers. In addition, information on the viability of the bacterium is not provided (Xuan *et al.*, 2013). Another shortcoming of metagenomic sequencing is depth bias, which is the inability to identify bacteria that are present in lower concentrations and taking into account the complexity of stool samples a

portion of the population will remain unidentified (Lagier *et al.*, 2015). Whilst the method was able to reveal a more complex microbial community it also highlighted that many of the microbiota found were unidentified and uncharacterised. This gave rise to the need for better culture techniques resulting in a technique called culturomics

Culturomics

Culturomics is the growing of different bacteria under different culture conditions, that mimic the environmental conditions from which the sample was collected (Greub, 2012b). The rise of culturomics was facilitated by the number of unknown bacteria that were identified using NGS. The use of microbial culturomics was first described by Lagier and colleagues in 2012 (Lagier *et al.*, 2012, 2015). Culturomics made use of the advances made in mass spectrometry methods to rapidly identify bacterial colonies that were grown under different culture techniques. This method allowed for the identification of previously unidentified bacteria (Lagier et al. 2015; Greub 2012; Fournier et al. 2015).

Whilst this technique does have its advantages, the biggest problem with culturomics is the time-consuming process and how costly it is to setup each culture condition. However, the technique does allow for the verification and culturing of clinically relevant viable species identified.

Importantly, the microbiome consists of not only microorganisms, but it also made up of proteins and metabolites that affect how the microorganisms within the environment interact with each other and those must be considered if we are to truly understand how the microbiome affects health. Today, mass spectrometry techniques allow identification of multiple components of the microbiome.

Mass spectrometry-based methods for microbiome /omics

A mass spectrometer is an instrument that is used to measure the mass to charge ratio of a molecule. Mass spectrometry is a technique where the compounds are ionized into charged molecules, these ions are fragmented and the ratio of their mass to charge (m/z) is measured (Hoffmann and Stroobant, 2007; Singhal *et al.*, 2015). It consists of three parts:

(i)an ion source that functions to charge the molecules into a gas phase

(ii) a mass analyser that functions to evaluate the mass-to-charge(m/z) ratio of the charged ions passing through the mass spectrometer and

(iii) a mass detector that measures the number of ions at each m/z value and generates a spectrum as an output.

A variety of ionisation techniques are used for mass spectrometry, the most preferred method for proteomics being electrospray ionisation (ESI) (Griffiths and Wang, 2009). ESI relies on the formation of gas phase ions from analytes in a liquid solution. The solution is propelled through a capillary by an electrical field, evaporated and charge is transferred to the analyte, thereby creating gaseous ions (Di Falco, 2018). The ionisation of compounds is not completely understood and two theories have arisen to try and explain the ESI process: the ion evaporation model (IEM) and the charge residue model (CRM) (Hoffmann and Stroobant, 2007; Mandal *et al.*, 2013).

In a Time-of-flight (TOF) analyser the ions are accelerated by an electric field into the analyser and the time taken for the ions to drift through a field-free chamber to reach the detector is then measured. The time taken is used to infer the ions m/z ratio (Peng *et al.*, 2003; Di Falco, 2018).

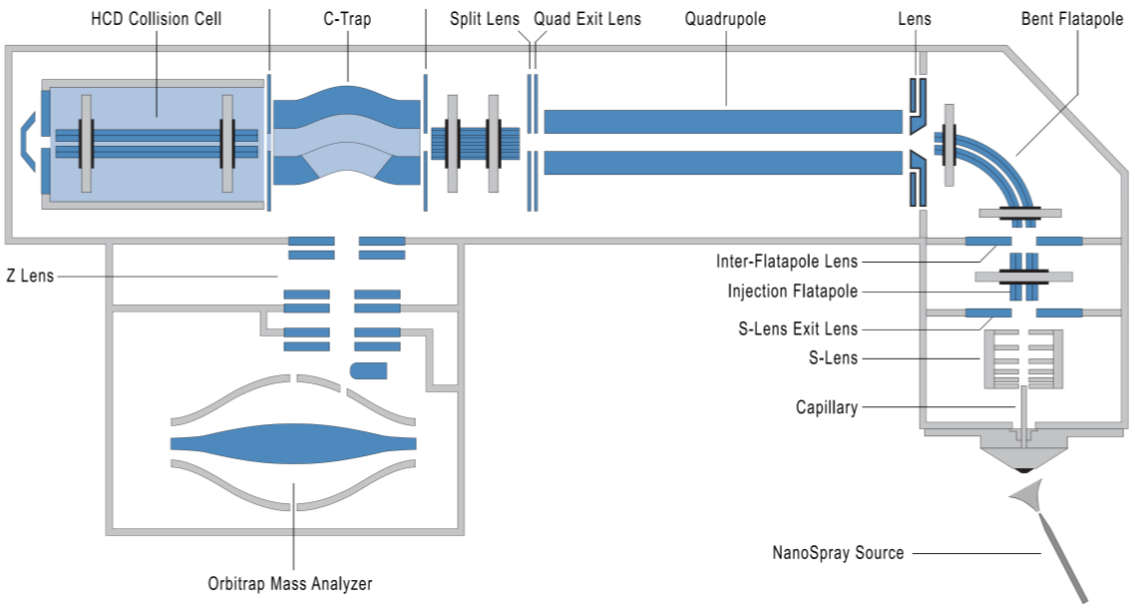


Figure 3: Schematic diagram of the Q-exactive Mass spectrometer

Most "-omic" experiments are conducted on tandem mass spectrometers (Griffiths and Wang, 2009). Tandem mass spectrometers are MS/MS analysers. Initially a peptide ion is detected and selected for fragmentation within the mass spectrometer (MS) (Di Falco, 2018). This generates an m/z ratio of the parent ion as well as an m/z ratio of its fragment ions. The most common type of fragmentation is collision-induced dissociation (CID) (Gundry *et al.*, 2010).

Mass spectrometry coupled technologies can be applied to a wide range of "omics "studies: proteomics, lipidomics and metabolomics. The ability for a single instrument to be applied to various sample types has mass spectrometry at the forefront of microbiome studies. This has already been applied to a field like culturomics. The use of mass spectrometry allowed for the identification of previously unidentifiable bacteria (Lagier et al. 2015; Greub 2012; Fournier et al. 2015)

Proteomics-based mass spectrometry

This then brings us to the application of mass spectrometry for proteomic analysis. Proteomics is defined by Patterson and Aebersold as "the systematic study of the many and diverse properties of proteins in a parallel manner with the aim of providing detailed descriptions of the structure, function and control of biological systems in health and disease" (Patterson and Aebersold, 2003). MS methods for protein identification have evolved as better techniques and technologies have arisen. At the onset of MS coupled proteomics, various workflows were used to separate proteins. Initially, use of MS techniques in proteomics relied on the use of gel separations such as sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE) extractions. SDS PAGE was extensively used at the onset of proteomic studies, however it was prone to sample loss and it suffered from low resolution (Thelen and Peck, 2007). This necessitated the use of alternative technique for protein separation giving rise to the use of the mass spectrometer coupled to chromatographic separation techniques for proteomics. The method allows for better identification of proteins in complex protein mixtures. This resulted in improved separation and identification of proteins.

The most popular coupled mass spectrometry technique for proteomics is Liquid Chromatography (LC) MS/MS, also known as shotgun proteomics (Di Falco, 2018). Typically, samples are prepared in a bottom up

or top down manner. Top down proteomics aims to characterise proteins by ionizing and dissociating an intact protein (McLafferty *et al.*, 2007), whereas bottom up proteomics relies on the digestion of a complex protein mixture by a protease – usually trypsin - to form peptides (Aguilar, 2004; Gundry *et al.*, 2009). The peptide mixture is then separated by online chromatography coupled to an ESI and analysed by MS/MS (Patterson and Aebersold, 2003; Aguilar, 2004; Gundry *et al.*, 2010). The generated spectra are searched against a database for the identification of proteins from the peptide sequences. The need for of a well curated database for spectral matching as well as the presence of a unique peptide as an identifier for positive protein identification is a challenge facing proteomics in the microbiome field.

MALDI-TOF identification of bacterial isolates grown in culture

Once cultured, bacterial isolates can be identified using MALDI-TOF (Matrix Assisted Laser Desorption Ionization-Time of Flight) mass-spectrometry (Clark *et al.*, 2013; Singhal *et al.*, 2015). The compound to be analysed is dissolved in a solvent, called the matrix (Hoffmann and Stroobant, 2007). The analyte is embedded in the matrix prior to irradiation by a laser. The irradiation results in the ionisation of the analyte and these ions are accelerated by an electrostatic field towards the analyser (Dingle and Butler-Wu, 2013). The detector measures the m/z ratio and generates spectra. The Bruker 'Biotyper' system relies on the unique ribosomal protein signatures generated by different species of bacteria. An m/z spectrum is generated, and the spectra are matched to those in the Bruker database. A score is assigned as to how accurately the microbe was identified (TeKippe *et al.*, 2013). MALDI-TOF is affordable, quick and accurate, but is limited by the availability of a reference spectra for the successful identification of the bacterial isolate. It also relies on culture techniques and some bacteria do not thrive in culture (Singhal *et al.*, 2015).

Metaproteomics

Whilst metagenomic approaches have allowed for the identification of bacterial species that do not thrive in culture, it does not reflect the functional proteins expressed within the microbiome. Metaproteomics is a field of study that is concerned with identifying proteins within an environmental sample. This sample type is more complex than those investigated by regular proteomics. As previously described the microbiome consists not only of bacterial components (although these have been the most well defined and studied) it

also includes metabolites and proteins from non -bacterial sources. Metaproteomics in the microbiome space allows for the identification of bacterial contributions as well as the identification of other microbial components such as viruses, fungi and even human/host effectors. The biggest challenge facing the use of MS for microbiome studies is the lack of an agreed upon protocol. As this type of an experiment would generate a large amount of data, further downstream bioinformatic processes are also required.

Table 1. Table of different methods applied to microbiome studies.

Technique	**Pros**	**Cons**
Culture	Affordable Infers viability	Fastidious organisms do not thrive Time consuming Expensive
Metagenomics	Quick Easy Data analysis tools for downstream analysis readily available.	PCR primer bias Does not infer viability Skewed towards only bacterial components
Metaproteomics	Quick Easy Identify all proteins: greater coverage	Preparation sample loss Identification relies on well curated database Most abundant proteins mask the less abundant proteins Data analysis tools for downstream analysis not readily available. Standardised sample preparation protocol doesn't exist.
Culturomics	Fastidious organisms grown Ability to identify new organisms	Expensive Time consuming Relies on well curated database

Research Hypothesis

The profile of the gut microbiome in HIV-exposed uninfected infants in Africa will be different from that of HIV -Unexposed infants and this difference may contribute to altered immune responses

Aims and Objectives:

1.Elucidate the gut microbiota in HIV- exposed uninfected infants using liquid chromatography mass spectrometry (LC-MS)

2. Use LC-MS to identify bacterial effectors (proteins) which potentially mediate the interaction between the gut microbiome and the host immune system.

Chapter 2. Methodology

The South African infant mother cohort was recruited from Khayelitsha, an urban township with a high prevalence of unemployment and informal housing that is located 35km from Cape Town. Khayelitsha has a high prevalence of HIV positive mothers who are on ARV treatment**.** The Mother and infant pairs were enrolled at the Maternal Obstetric Unit (MOU) at Site B, the largest of two public sector delivery units that serve the estimated 400,000 living in Khayelitsha.An estimated 30-32% of the deliveries are to HIV-infected mothers. Unemployment and informal housing are highly prevalent in the area . (Appendix A.1). Stool was collected from the infant's nappy at birth and day 4-7, placed in stool specimen containers. The samples were stored at -80 ^{0}C till sample preparation.

2.1 Selection Criteria and ethical considerations

This was a study, nested within a previously approved study: *Innate, Adaptive and Mucosal Immune Responses in HIV-1 Exposed Uninfected Infants: A Human Model to Understand Correlates of Immune Protection* with **HREC/REF:285/2012,** referred to as the "**InFANT study**" seeks to identify correlates of protection against HIV conferred by exclusive breast feeding.. The parent study has approved informed consent documents that are currently in use. In addition, details regarding the participants, inclusion and exclusion criteria, data confidentiality, ethical considerations and risks to and benefits for participants are described in the parent study **HREC/REF: 285/2012**. This sub study has ethics approval to carry out metaproteomic work on a subset of the infant's stool samples: **HREC 062/2018 (**Appendix A.2).

For this study, we analyzed existing stool samples stored from the InFANT cohort to characterize the infant gut proteome and metabolome of 34 HIV exposed uninfected (HEU) versus 29 HIV-unexposed (HU) infants, from the South African arm of the study, during the first week after birth in breastfed infants. The cross-sectional samples were collected at two-time points, namely Birth and Day 4-7, and analysed.

Table 2. Eligibility criteria for sample recruitment in InFANt Study

HIV exposed cohort	**HIV unexposed cohort**
1) Mother is HIV infected	1) Mother is HIV uninfected
2) Age of mother ≥ 18 yrs.	2) Age of mother ≥ 18 yrs.
3) Mother has self-chosen to breast feed her infant	3) Mother has self-chosen to breast feed her infant
4) Mother is able and willing to do the follow up assessments and provide informed consent	4) Mother is able and willing to do the follow up assessments and provide informed consent
Infant Factors	**Infant Factors**
1) Gestational age ≥35 weeks	1) Gestational age ≥35 weeks
2) Birth weight ≥ 2.0kg	2) Birth weight ≥ 2.0kg

Exclusion criteria

Maternal Factors: Complications during pregnancy and delivery such as chorioamnionitis

Infant factors: Hypoxic injury /seizures/sepsis/ intrauterine growth retardation

Table 3. Clinical information of infant mother diads recruited for study.

Status	**HEU**		**HUU**	
Timepoint	Birth	D4-7	Birth	D4-7
Female infants	10	8	2	11
Male infants	3	13	5	11
Average infants birth weight (g)	3184,6	3051,4	3251,4	3082,7
Average infants birth height (cm)	48,2	47,8	49,4	46,8
Average infants head circumference (cm)	33,5	32,7	34,4	33,3
Average infants temperature (celsius)	36,4	36,6	36,6	36,5
Average Infants gestational age in weeks	39	38	38	39
Infants on Arv	Yes	Yes	N/A	N/A
Average Mothers CD4 count	427	463	N/A	N/A
Average Mothers age	25	28	24	26

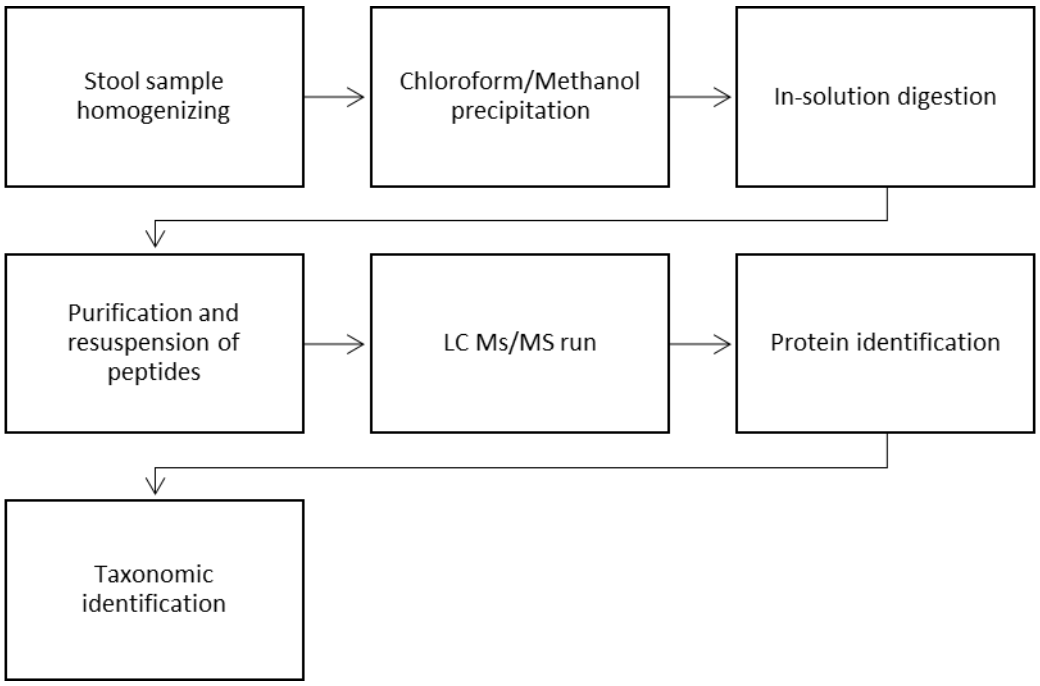


Figure 4: Flowchart of the workflow followed for sample processing and analysis

2.2 Stool homogenization and Protein Extraction

Each stool sample was handled independently and snap frozen using liquid nitrogen and homogenized mechanically to a fine powder by use of a pestle and mortar. Once homogenized the powder was placed in amber glass vials and isopropanol added. The samples were incubated over night for 18 hours in a fridge at -20^{o} C. The stool was further homogenized, and cells disrupted by sonication at maximum power for six cycles of 45 seconds with 1-minute cooling on ice between cycles (Rezwan et al ,2007). The tubes were centrifuged at 3500 rpm for 30 minutes and the supernatant was removed and placed in amber glass vials for storage at -80^{o} C. Supernatant was removed from the pellet and stored in a separate glass vials after each centrifugation steps during the sample preparation. The supernatant was stored for future metabolomics study but not analyzed as part of this study. The pellet was then resuspended in a 1:1 (v: v) chloroform-methanol solvent and vortexed for a minute prior to 18-hour incubation at -20^{o}C. After incubation, the chloroform-methanol resuspension was centrifuged at 3200rpm for 30 minutes. The pellet was then resuspended in a 100% methanol and centrifuged at 3400 rpm for 30 minutes at 4^{o}C. The supernatant was decanted off and the tubes were air dried, after which the pellet was resuspended in Denaturation Buffer (10mM Tris-HCL, 6M

urea, 2M thiourea, pH8) and stored at -20^{o}C for downstream protein digestion and mass spectrometry analysis.

2.3 Protein quantification and Digestion

A modified Bradford assay (Ramagli, 1999), using 1% HCl solution, Bradford Reagent and Bovine Serum Albumin (BSA) as a standard, was carried out to quantify the amount of protein in each sample. Total protein was calculated and 200µg of proteins were aliquoted out and reduced in 1µL of 1M DTT for 1 hour followed by incubation in 1µL of 550mM iodoacetamide (IAA) for another hour. The samples were diluted four-fold in 20mM ABC pH 8 prior to trypsination, with sequence grade modified trypsin (Promega, Madison, USA) added at a 1:50 (w/w) ratio, 2M $CaCl_2$ was added to the vials at a 1:10(v/v) ratio. The vials were wrapped in foil and incubated for 16 hours on a shaker at medium speed. Trypsination was ended by addition 1µL of 100% Formic acid.

2.4 Desalting and purification

10 µg aliquots of the resulting tryptic peptides were desalted using in-house produced C18 stage tips. Activation of the C18 discs was carried out by three rinses with 80% acetonitrile (ACN), followed by three rinses with 2% ACN. 10µg of each sample was added to the C18 discs, one sample per stage tip, and centrifuged. The sample were desalted by three rinses of 2% ACN followed by elution into glass capillary tubes by three rounds of 100µL of 60% ACN, 0.1% formic acid. The activation, equilibration, peptide wash and elution were carried out at 4000rpm for 5 minutes each rinse. The samples were dried in a vacuum concentrator Speedy Vac at 30 ^{0} C for 45 minutes and reconstituted in 40µL of 2% ACN, 0.1% formic acid to 250ng/µL solution, prior to LC-MS/MS analysis.A pooled sample of 1µL from each of the reconstituted samples was also prepared prior to LC-MS/MS analysis

2.5 Liquid Chromatography-Mass Spectrometer analysis

Samples were analysed on a Thermo Scientific Dionex Ultimate 3000 UHPLC (Thermo Fisher Scientific, Waltham, MA, USA) coupled to a Thermo Scientific Q-Exactive hybrid quadrupole orbitrap mass spectrometer. The samples were loaded onto a 2 cm Luna C18 100 µM internal diameter fused silica trap, packed in-house, and then separated on a 30 cm Aeris peptide C18 75 µM internal diameter analytical column, packed in-house and maintained at 40^{o}C. The samples were randomised and run individually, and injection volumes were optimised batch-wise. All samples were eluted on a 70-minute gradient, with a constant flow rate of 300nl/min. Solvent A (0.1% Formic acid/ HPLC grade water) and solvent B (0.1% Formic Acid/ Acetonitrile). The gradient consisted of 2% solvent B for 10 minutes then increasing to 6% B for 2 minutes, followed by increasing to 40% B at 40 minutes; followed by washing with 80% after 40 minutes and a 20-minute calibration, with a 45-minute wash between every 4 samples.

Mass spectrometric analysis was facilitated by Xcalibur software (Thermo Scientific). The acquisition parameters were optimised in-house based on the settings described by (Pirmoradian et al, 2013) with a few amendments:. A top 10, data dependent, positive ion mode was used to acquire mass spectra with automatic cycling between MS and MS/MS scans as previously described by (Domon & Aebersold 2010). Ion fragmentation was performed by higher energy collision induced dissociation (HCD). MS1 settings included a resolution of 70000, scan range of 300-1750 m/z automatic gain control (AGC) target of 3E6, and an ion injection time of 250 ms. At the MS2 level the setting were as follows: isolation window of 4.0 m/z, a normalised collision energy (NCE) 25, resolution of 17500, AGC target of 5E4 and ion injection time of 80 ms. Data dependent settings included an underfill ratio of 1% (which equates to an intensity threshold of 1.7x104), peptide match set to “preferred, isotopic exclusion, and a dynamic exclusion of 30 s. Charge exclusion was set to all unassigned charges, as well as all charges other than 2 or 3.

2.6 Data Analysis

The raw mass spectra were converted to Mascot generic files (mgf) format on MS convert. The mgf files were processed using an inhouse developed metaproteomic analysis pipeline, MetaNovo (Potgieter *et al.*, 2019)**.**

Briefly, *de novo* sequence tags were derived from the MS/MS spectra using DeNovoGUI version 1.15.11 (Muth *et al.*, 2014), with DirecTag (Tabb *et al.*, 2008). The resultant sequence tags were mapped to the entire UniProt database (ca. 120m sequences) using a high performance and parallelized computing pipeline, generating compact databases for downstream target-decoy analysis and False Discovery Rate (FDR) controlled protein identification. The output protein IDs from the initial MetaNovo search were then re-searched with the original raw MS/MS dataset under stringent conditions using MaxQuant (v.1.5.5.0).

The MetaNovo pipeline was used to generate a database with Fasta files from the raw files. MaxQuant version (1.5.3.12) software package was used to search the mgf file, using the Andromeda search engine, against the MetaNovo generated database. The default Maxquant settings were used with the following modifications: Trypsin was selected as an enzyme with a maximum of 2 missed cleavages, LFQ was selected for Label free quantitation, the default MaxQuant false discovery rate (FDR) cut-off of 0.01 (1%) was used at the peptide-spectrum level with use of a decoy database which included a nonsense reverse entry for every predicted protein. Protein identification was set at a minimum of two unique peptides required for protein identification. Match between runs was selected to increase identification of protein groups. Quantitation was *via* the MS1-based LFQ method embedded within MaxQuant All searches were performed on a virtual Windows server running on UCT High Performance Computer cluster (HPC). Statistical analysis was carried out in R to identify metaproteins that are differentially expressed between groups.

MS data quality and identification summary statistics were assessed with the aid of a custom script in the R environment.

2.7 Taxonomic analysis

The taxonomical analysis was performed using the UniPept online software tool. Peptide sequences from the *peptide.txt* files for Birth and Day 4-7 were uploaded separately to the UniPept-web based platform with the following parameters: a) Equate I and L; b) Filter duplicate peptides and c) Advanced missed cleavage handling. The files were uploaded as separated by timepoint and exposed status.

2.8 Functional analysis

Functional analysis of the taxa identified was carried out on the UniPept QuickGo plugin which generates Go Terms associated with proteins at the different Hierarchical level. The LCA level chosen was Chordata and Biological process was analysed.

2.9 Protein analysis

Proteingroups.txt file from the MetaNovo combined folder for each timepoint was uploaded onto Microsoft Excel, total number of peptides in each group were assessed and 2 or more peptides for positive protein identification was used. Each timepoint was analysed separately. Proteins were filtered for human only proteins and functional analysis of the proteins was executed on the Pantherdb.org web-based tool. In Panther Gene List Analysis, the following parameters were selected: 1. 'ID list' selected as list type, 2. species set to Homo sapiens, and 3. Functional classification viewed on graphic charts. The Panther dB analysis was used to identify the protein classes identified in each group at both time points.

Analysis of the protein groups was carried on Perseus 1.6.8.0. The *proteinGroups.txt* folder for Birth and Day4-7 were uploaded separately onto the Perseus platform, LFQ values were selected under the category of main. Rows were filtered based on categorical column, (Site, Reverse, and Contaminant). The rows were annotated categorically then checked for normal distribution, the rows were further filtered based on numerical /column, based on unique peptides with, >1 as the threshold. The data was then log transformed and rows filtered on valid values with a minimum of 2 in each group. A two-sample t test was carried out, with Benjamin-Hochberg FDR set for truncation and a volcano plot generated. Whilst there were no significantly upregulated proteins identified (p-value 0.05, FDR0.05) at either timepoint, some proteins were differentially expressed.

A presence-absence analysis of the human proteins was carried out *Proteingroups.txt* file from the MetaNovo combined folder for each timepoint was uploaded onto Excel, total number of peptides in each group were assessed and 3 or more peptides for positive protein identification was used. Proteins were

filtered out for human only proteins. the human proteins were then uploaded onto Venny2.1.0 online platform to generate Venn diagrams and protein list exclusive to each group at each timepoint.

From the presence-absence analysis proteins that were exclusively found in each sample set were uploaded onto the STRING online tool and an enrichment analysis with background statistics set to whole genome was run.

Chapter 3: Results

3.1 MS data quality

During the sample preparation the concentration of the samples was measured using a modified Bradford assay. Once the concentration for each sample was determined the volume required for 200µL of protein was calculated and aliquoted into clean vials for Trypsin digestion.

Table 4: BSA concentration of samples

Sample ID	STATUS	Concentration (ug\uL)	Timepoint
4	HEU	3,75	Birth
8	HEU	2,391	Birth
9	HEU	4,103	D4-7
10	HEU	(-)	D4-7
11	HEU	(-)	D4-7
13	HEU	(-)	D4-7
16	HEU	3,56	D4-7
17	HEU	8,61	D4-7
18	HEU	5,452	D4-7
20	HEU	0,183	Birth
26	HEU	3,257	Birth
38	HEU	(-)	D4-7
54	HEU	(-)	Birth
55	HEU	(-)	Birth
61	HEU	(-)	D4-7
62	HEU	3,524	Birth
63	HEU	1,111	Birth
66	HEU	0,113	
71	HEU	0,939	Birth
78	HEU	(-)	D4-7
81	HEU	(-)	Birth
88	HEU	3,98	Birth
92	HEU	1,562	D4-7
100	HEU	5,77	Birth
121	HEU	1,699	D4-7
127	HEU	1,836	D4-7
138	HEU	(-)	D4-7
140	HEU	5,61	D4-7
142	HEU	6,34	D4-7
143	HEU	0,34	D4-7
156	HEU	6,589	D4-7
157	HEU	(-)	D4-7

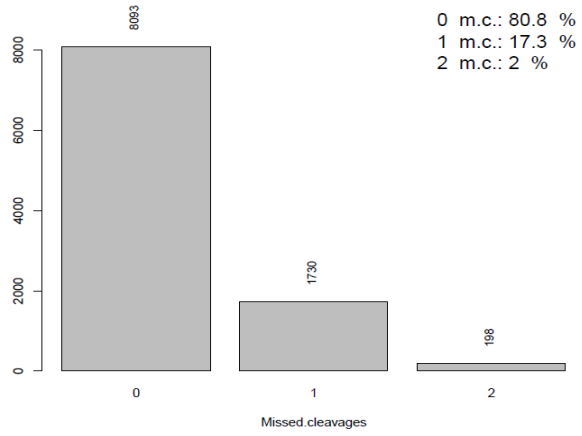
159	HEU	4,69	D4-7
161	HEU	2,15	D4-7
302	HU	5,866	Birth
305	HU	(-)	D4-7
307	HU	3,404	Birth
308	HU	5,188	D4-7
309	HU	1,754	D4-7
312	HU	1,73	D4-7
314	HU	1,64	D4-7
315	HU	7,83	D4-7
316	HU	(-)	D4-7
318	HU	(-)	D4-7
319	HU	5,15	D4-7
320	HU	1,83	D4-7
325	HU	0,267	D4-7
332	HU	3,027	D4-7
337	HU	1,892	D4-7
347	HU	1,304	D4-7
345	HU	didn't process	D4-7
350	HU	(-)	D4-7
354	HU	2,268	Birth
357	HU	3,8	Birth
359	HU	8,99	Birth
360	HU	6,88	D4-7
365	HU	8,48	Birth
374	HU	(-)	Birth
377	HU	3,98	D4-7
381	HU	6,54	D4-7
383	HU	7,86	D4-7
387	HU	0,836	D4-7
389	HU	2,91	D4-7

The data analysis for the Birth and Day4-7 was carried out separately. MS1 and MS2 data were acquired for all samples on the LC-MS.

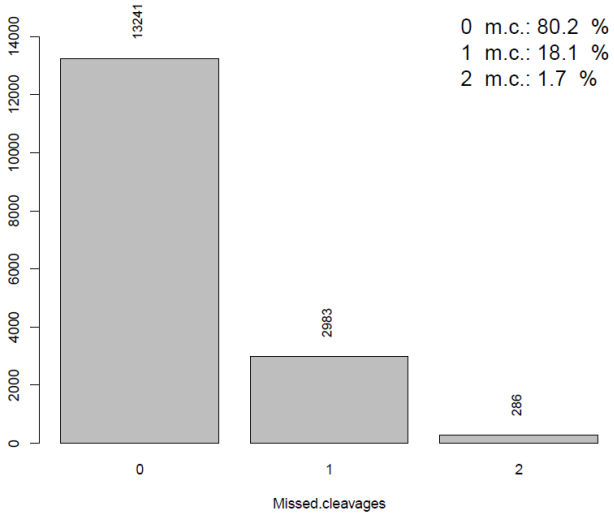
For the Birth timepoint 13 HEU vs 7 HUU stool samples 1026286 spectra were submitted to MaxQuant software for analysis, 78320 (7.6%) were mapped to known tryptic peptide sequences in the MetaNovo generated protein database corresponding to an identification of 2816 protein groups and 10021 non-redundant peptides in total across the birth samples. Whilst the data had a good tryptic digest, with few missed cleavages, the data did have 4.9% contaminants present.

In the day 4-7 samples 21 HEU vs 22 HUU stool samples were processed, 828997 spectra were submitted to MaxQuant software and 68994 (8.3%) of the spectra were identified, corresponding with an identification of 4712 protein groups and 16510 non redundant peptides were identified. The data had 2.5% contaminants, but the tryptic digestions were of good quality.

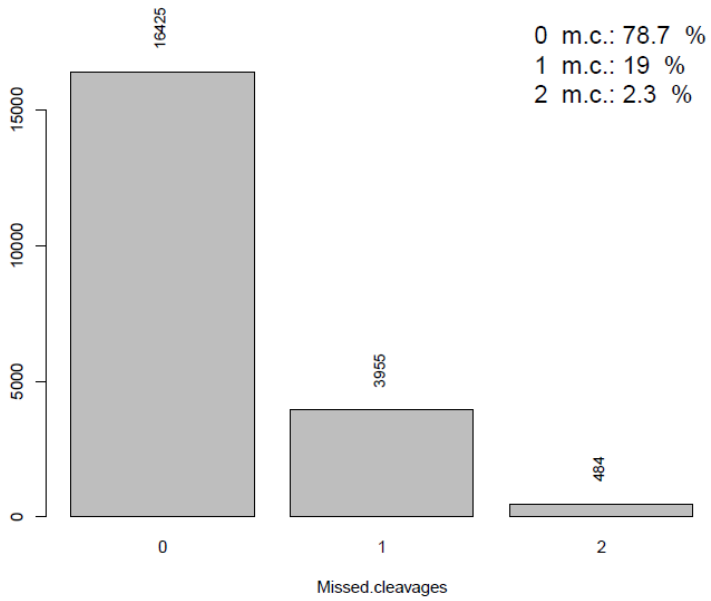
An analysis of all the samples from both timepoints was also carried out. 2681796 spectra were submitted to MaxQuant software for analysis, 220488 (8.2%) of the spectra were mapped to known tryptic peptides in the MetaNovo generated database. This correspond to 5597 protein groups and 20864 identified peptides. The data generated was of good quality, with 0.8% contaminants identified in the protein groups. The number of identified peptide sequences varied amongst the samples and no batch effect was detected. Thus, all the samples could be analysed concurrently.



a.



b.

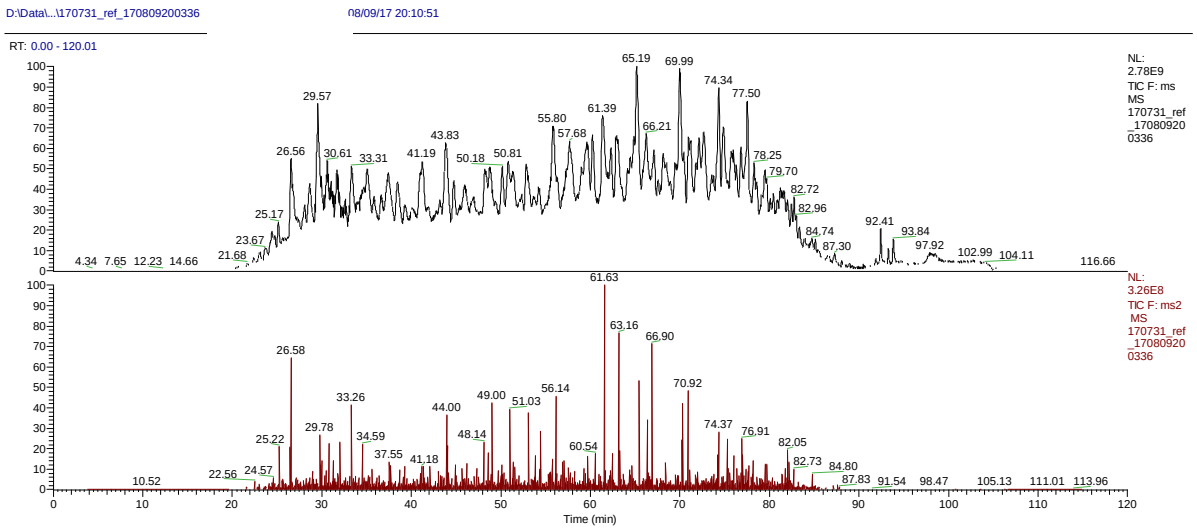


c

Figure 5: Summary statistics generated on R a) Bar graph of missed cleavages in the Birth samples generated from summary script in R displays the quality of the digestion of the trypsin, and indicates that the digestion of the protein into peptides was successful with 80.5% of the peptides identified revealing no missed cleavage .b) Bar graph of the missed cleavages in the Day4-7 batch of samples , the digestion of the protein into peptides was successful with only 80.2% missed cleavages .c) Missed cleavages for all samples from both time points .

The chromatograms generated had a good gradient and distribution over the duration of the run (refer to figure 6a**).** The reference sample and the pooled sample both show good charge and distribution (Appendix B.1) The pooled sample functions as quality control to aid in determining the optimum injection volume as well as ensuring no contaminants are in the samples.

a.



b.

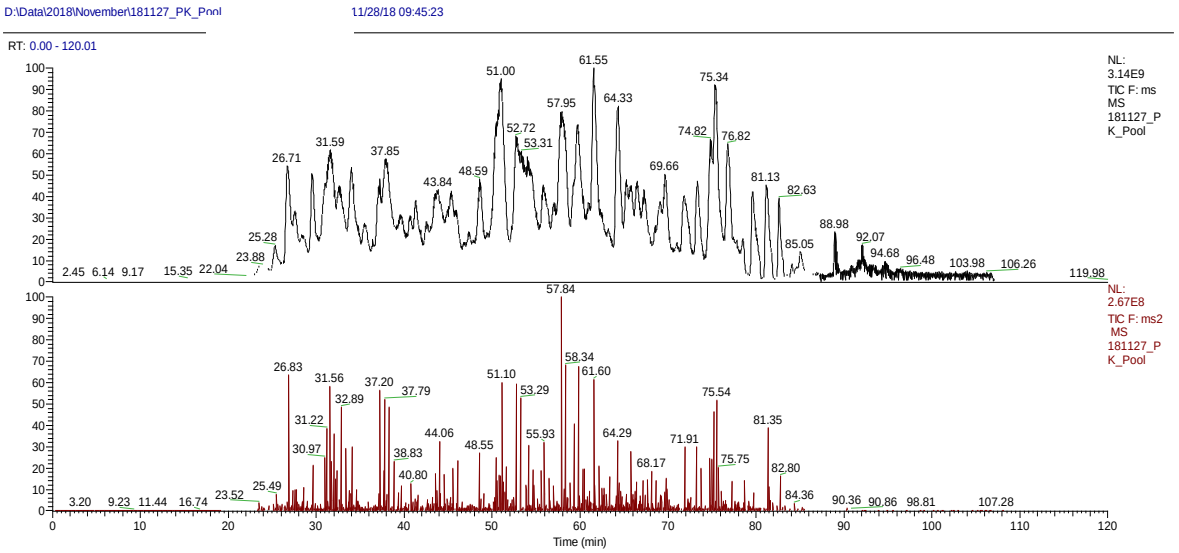


Figure 6: Chromatograms generated from samples run on the Q-Exactive mass spectrometer a) reference sample SH-SY5Y, used to confirm the machine is running well

b) pooled sample of stool sample used as an indicator of the optimum injection volume as well as assess if there is any contamination The injection volume was set to 2uL.

3.2 Taxonomic identification

Metaproteome analysis

The Venn diagram reveals that there is some overlap in peptides that are identified within each group at each time point, however some peptides that are specific to each group have also been identified. Most of

the unique peptides were found in HEU at day 4-7 (20.6%), followed by HUU at day 4-7 (19.5%), HEU at birth (8.2%) and finally HUU at birth (7.3%).

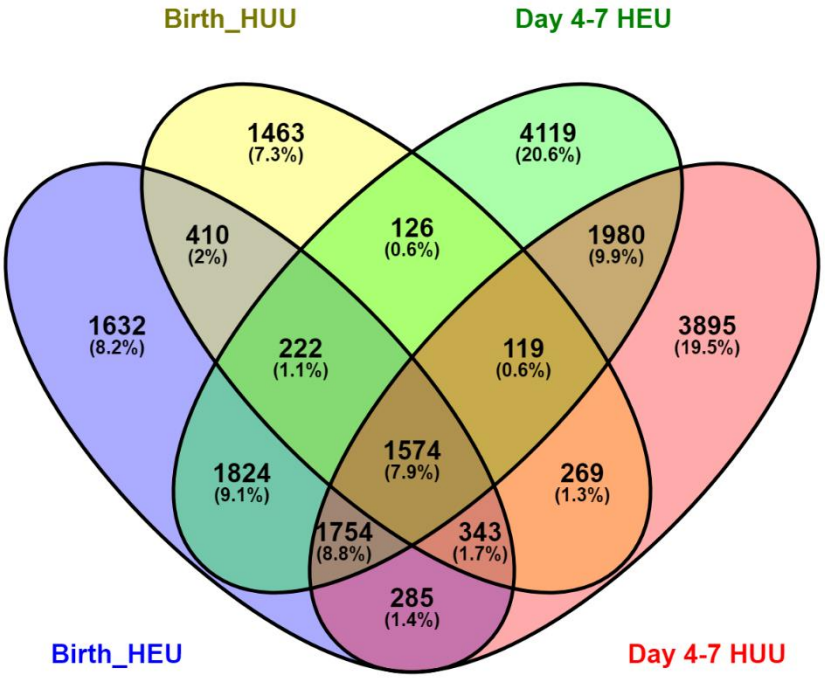


Figure:7: Venn diagram of the Peptides that are shared amongst the different groups at both timepoints.

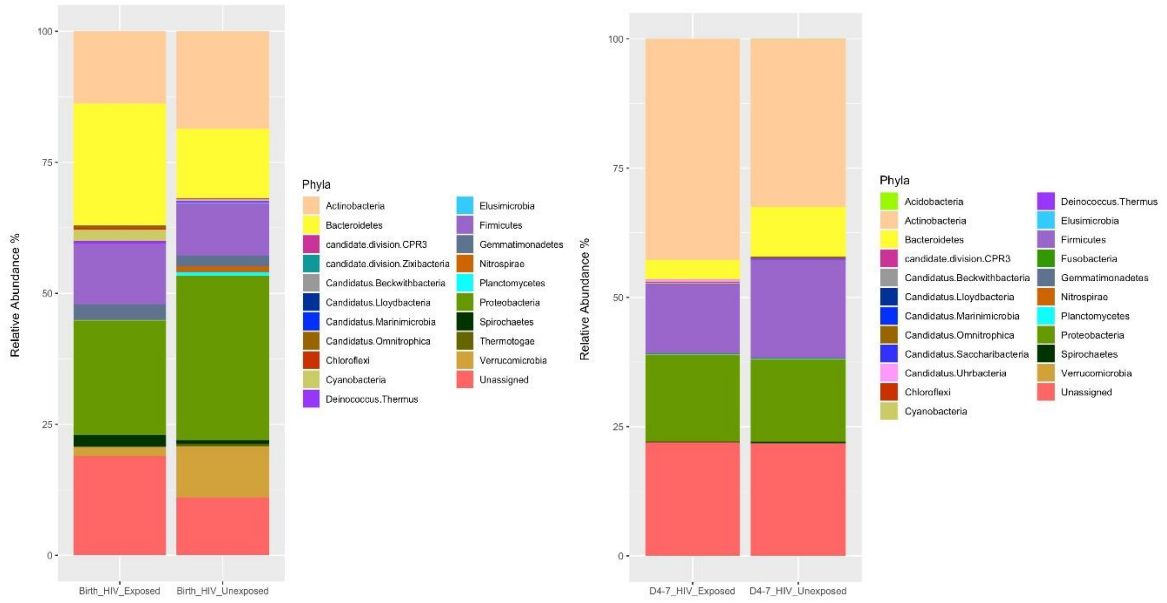


Figure 8: Bacterial metaproteome based on Phyla in the HEU vs HUU at Birth and during Day 4-7.

The difference in the relative abundance of the phyla at birth and at timepoint day 4-7 is depicted in Figure 8. At Birth the microbiota of the HEU infant has more Bacteroidetes than HUU whereas at day 4-7 the HEU relative abundance of Bacteroidetes decrease and the relative abundance of Actinobacteria increases. There is a shift to a microbiome in the HEU that is dominated by Actinobacteria. Interestingly, the HUU birth microbiota has a greater abundance of Actinobacteria, however at day 4-7 the HUU is still dominated by Actinobacteria but the relative abundance is less than that observed in HEU infants.

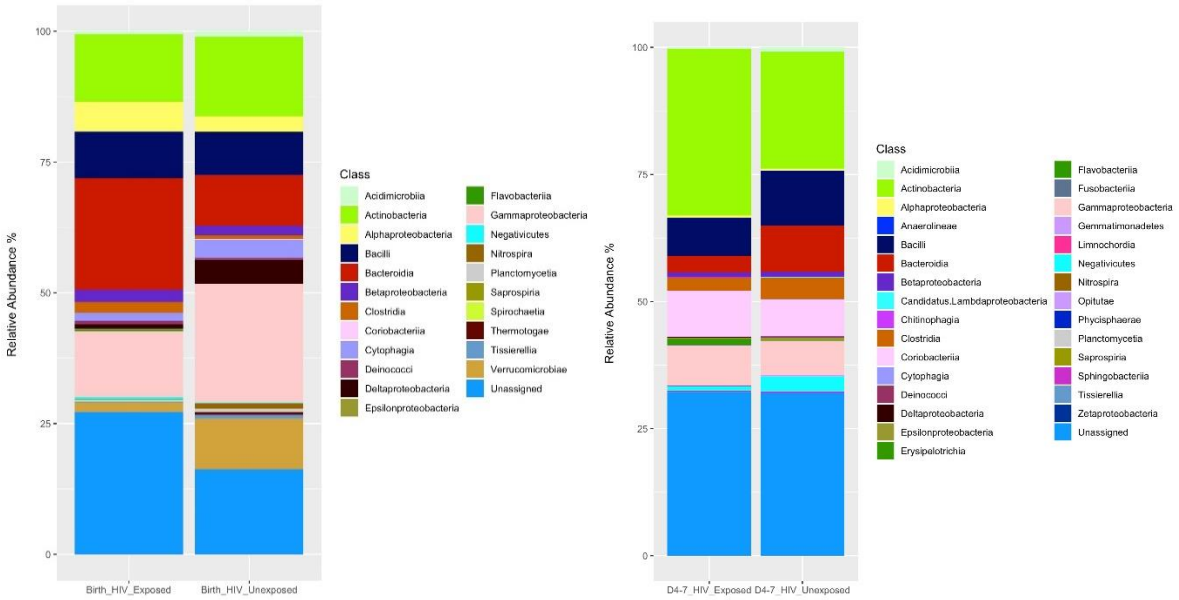


Figure 9: Stacked bar graph indicating the relative abundance of the metaproteomic classification based on Intensity of Class of bacteria in HEU and HUU infants at both time points

The Class distribution of the microbiota belonging to the kingdom Bacteria

At birth the HEU is and HUU are comprised of the same class of bacteria, however the relative abundance of each class in each group varies. In the HEU the majority is *Bacteroidia*, *Actinobacteria, Gammaproteobacteria* and *Bacilli*. Whereas in the HUU infant at birth has a microbiota comprised of mostly *Gammaproteobacteria, Actinobacteria, Bacteroida* and *Bacilli* respectively. The microbial composition of both the HEU and the HUU shift to an *Actinobacteria* dominated microbiota at day 4-7, with an increase in *Coriobacteria* in both infant groups (Table 5) .

Table 5: Numbers of peptides identified for each class

Bacterial Class	Number of peptides			
	D4-7HEU	D4-7HUU	Birth HEU	Birth HUU
Acidimicrobiia	1	1	0	1
Acidobacteriia	1	1	0	0
Actinobacteria	2565	1548	1812	429
Alphaproteobacteria	18	17	11	7
Anaerolineae	1	1	0	0
Ardenticatenia	0	0	1	0
Bacilli	247	142	27	59
Bacteroidia	450	389	309	29
Betaproteobacteria	25	15	17	4
Candidatus Lambdaproteobacteria	1	1	0	0
Chitinophagia	1	1	1	0
Chlamydiia	0	1	0	0
Clostridia	443	551	224	11
Coriobacteriia	1198	707	2	9
Cytophagia	5	4	2	2
Deinococci	1	1	1	1
Deltaproteobacteria	5	3	3	2
Epsilonproteobacteria	2	3	1	0
Erysipelotrichia	103	0	1	0
Flavobacteriia	5	2	7	0
Fusobacteriia	8	0	0	0
Gammaproteobacteria	642	276	219	13
Gemmatimonadetes	1	1	1	1
Limnochordia	1	1	0	0
Negativicutes	339	213	32	1
Nitrospira	3	3	1	1
Opitutae	0	1	0	0
Phycisphaerae	3	3	0	0
Planctomycetia	2	1	1	1
Saprospiria	2	0	1	1
Sphingobacteriia	4	1	1	0
Spirochaetia	1	1	1	0
Theionarchaea	0	0	1	0
Thermotogae	1	1	1	1
Tissierellia	1	0	1	1
unassigned	2160	1401	837	134
Verrucomicrobiae	0	0	1	1
Zetaproteobacteria	1	0	0	0

From the data generated it is observed that there is a difference in the composition of the class of bacteria in the two groups at each time point.

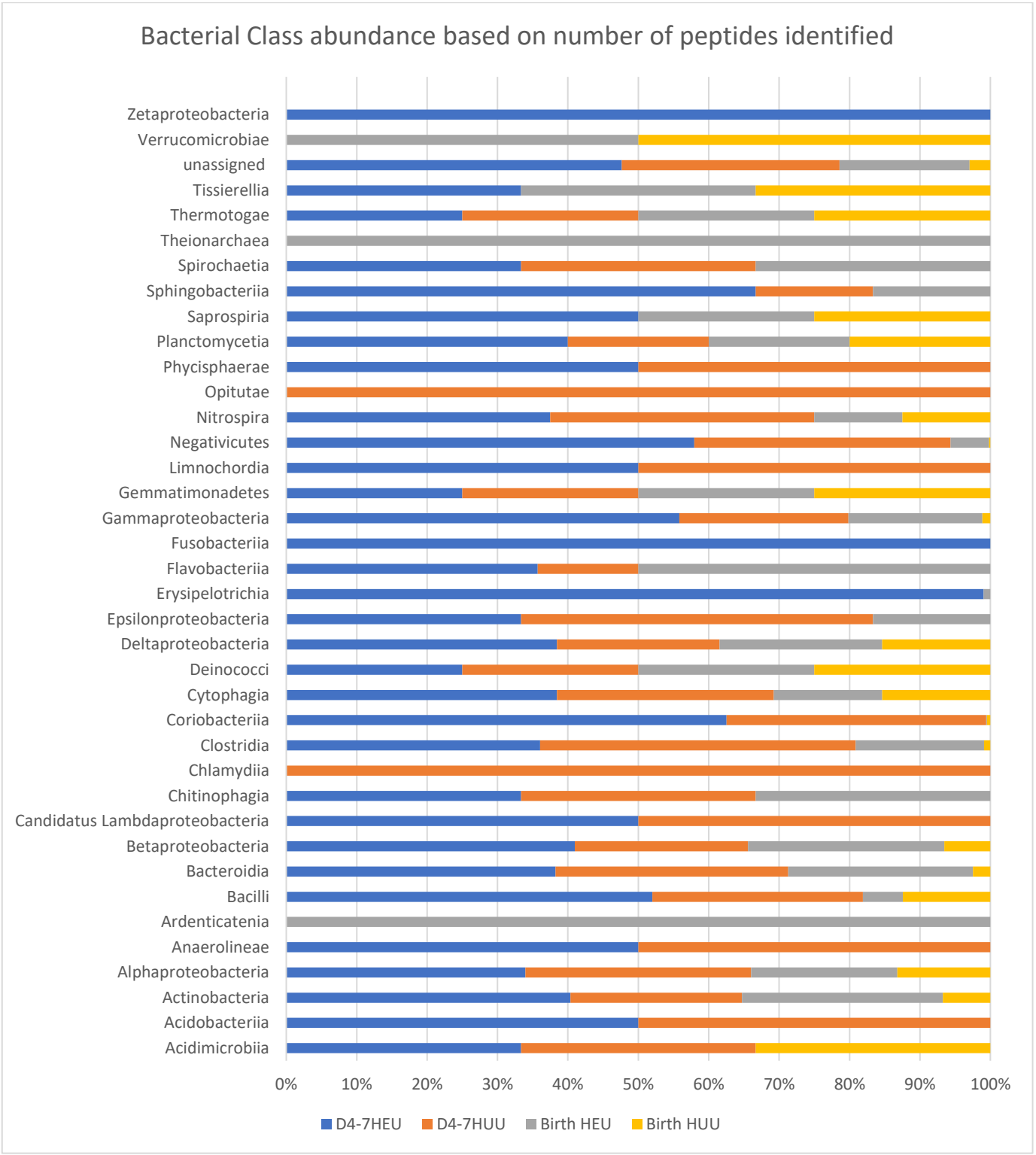


Figure 10: Bacterial class abundance.

The Contribution of each timepoint to each bacterial class is illustrated in Figure 10. The timepoint that has the greatest contribution to the bacterial classes is the D4-7HEU group. Some of the classes identified were

attributed to one group only, Chlamydia and Opitutae were exclusively present in the D4-7HUU group whilst Theinoarchae and Ardenticatenia were only present in the Birth HEU group. Interestingly Erysipelotrichia was exclusive to the HEU groups at both time points.

Birth HUU

In the stool samples of the HUU infant at birth 4394 out of 4526 of the peptides uploaded onto UniPept were identified. Many of the peptide's sequences, 74% in total, that were uploaded were mapped to Eukaryotes. Whereas only 16% of the peptides were mapped to the domain Bacteria.

The bacterial composition of the HUU infant at birth is primarily composed of bacteria belonging to order Corynebacteriales, Bifidobacteriales and Lactobacteriales. All three of these orders have been previously identified in the faeces of humans. Of the Actinobacteria identified, the majority are mapped to peptides belonging to the family Bifidobacteriacea. The two species with the genus Bifidobacterium that were identified were *B. bifidum, B. breve and B. longum..*

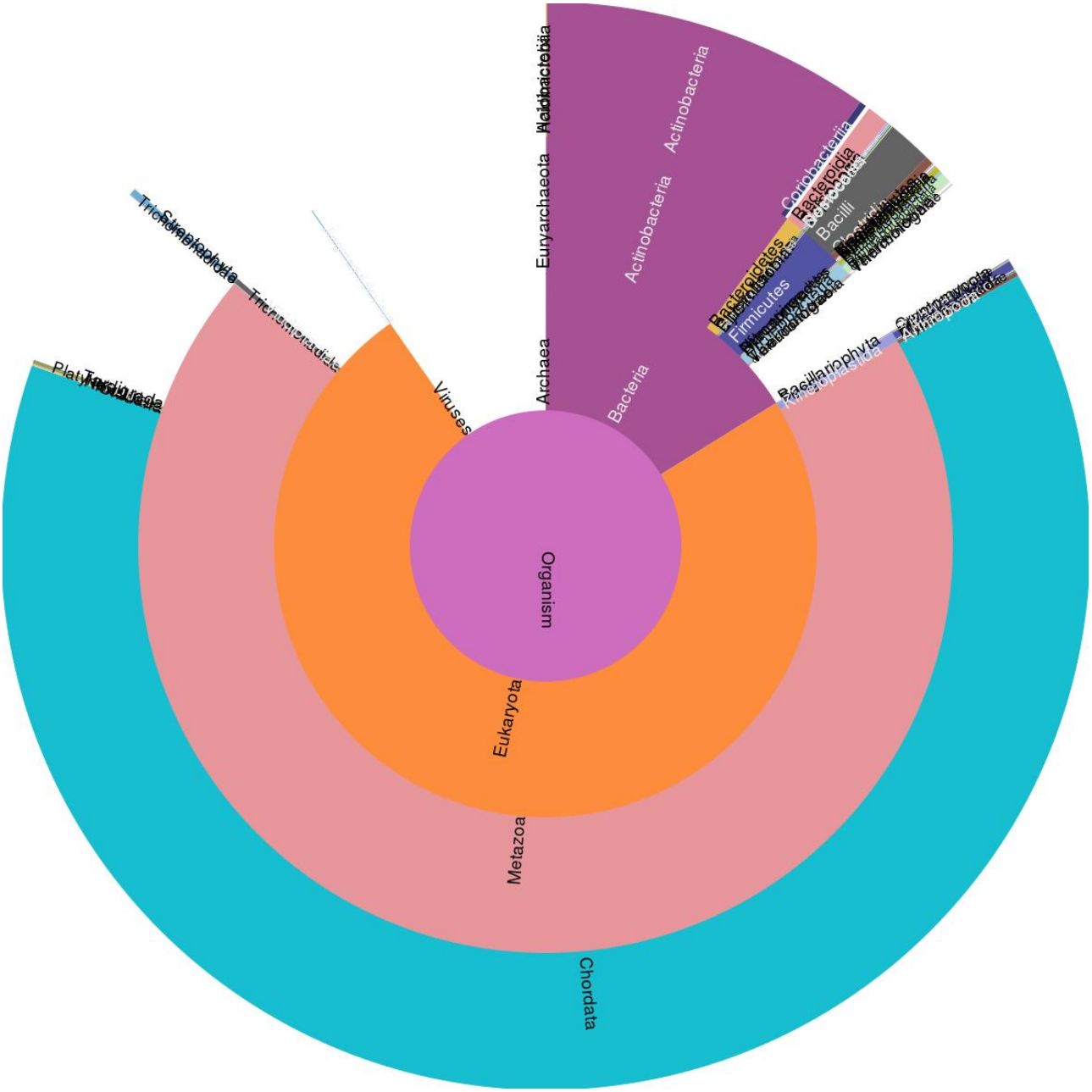


Figure 11: Sunburst chart generated on UniPept providing an overview of the organism identified HUU. 74% of the peptides identified belong to Eukaryotes Mesuere et al. (2015) Proteomics doi:10.1002/pmic.201400361

The Bacteria in HUU at birth are not as abundant as in the HEU infant. Most of the peptides identified in the HUU infant belong to the Eukaryotes namely Chordata, assumed to be human peptides, as these are human samples.

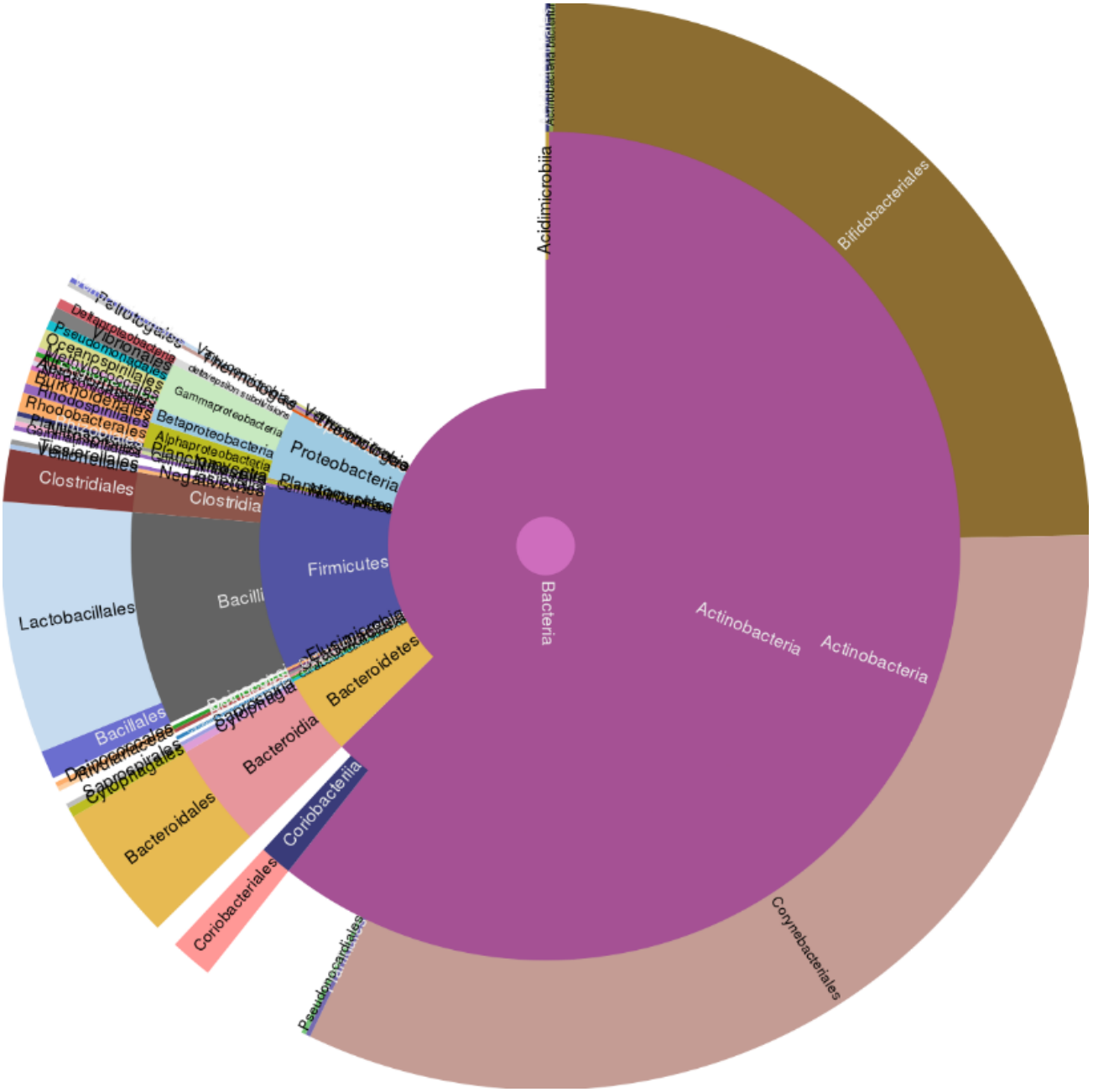


Figure 12: Expanded sunburst chart of the HUU microbiota at birth, detailing the bacteria identified, with 16% of the peptides identified belonging to the bacteria superkingdom Mesuere et al. (2015) Proteomics doi:10.1002/pmic.201400361

Interestingly, Mycobacterial *tuberculosis complex* was also identified in the HUU birth stool samples(Table 6).

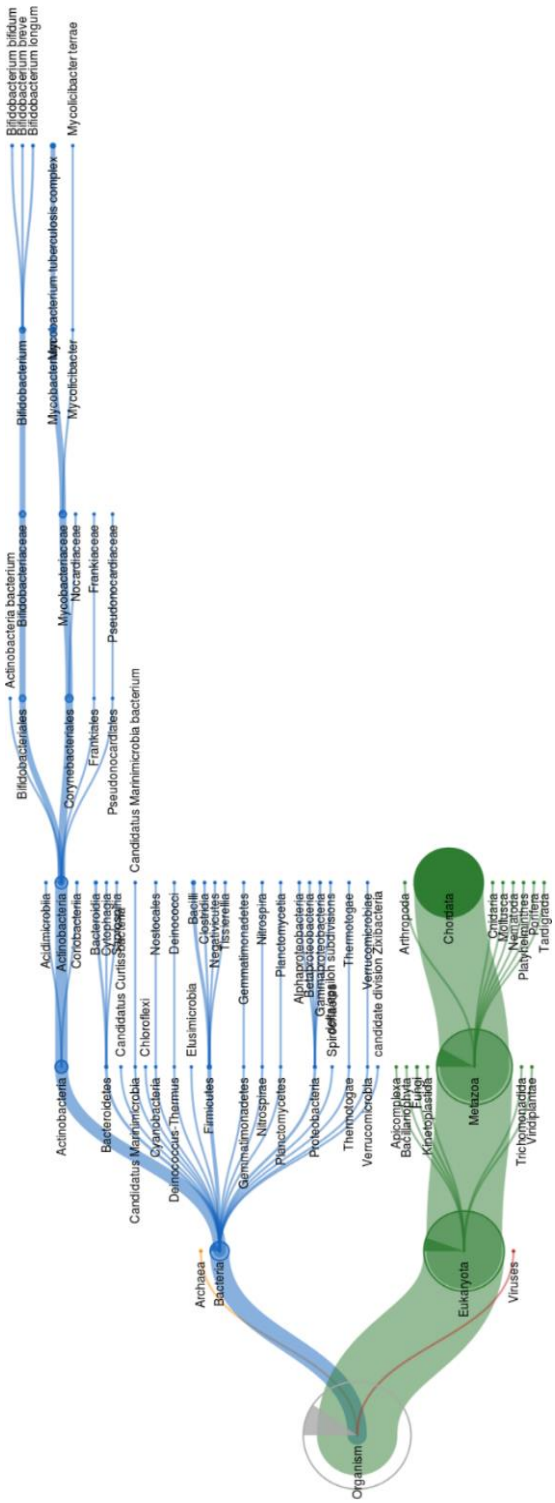


Figure 13: Treemap generated on UniPept of the organisms identified in HUU birth infants
Mesuere et al. (2015) Proteomics doi:10.1002/pmic.201400361

Birth HEU

In the HEU birth infant's UniPept analysis 7778 out of 8044 peptides were identified on the platform. In HEU infants only 44% of the peptides identified were matched to Eukaryotes, which suggests that the HEU

microbiome is more diverse than the HUU birth microbiome. With 45% of the peptides identified belonging to the Kingdom bacteria, more than in the HUU birth samples

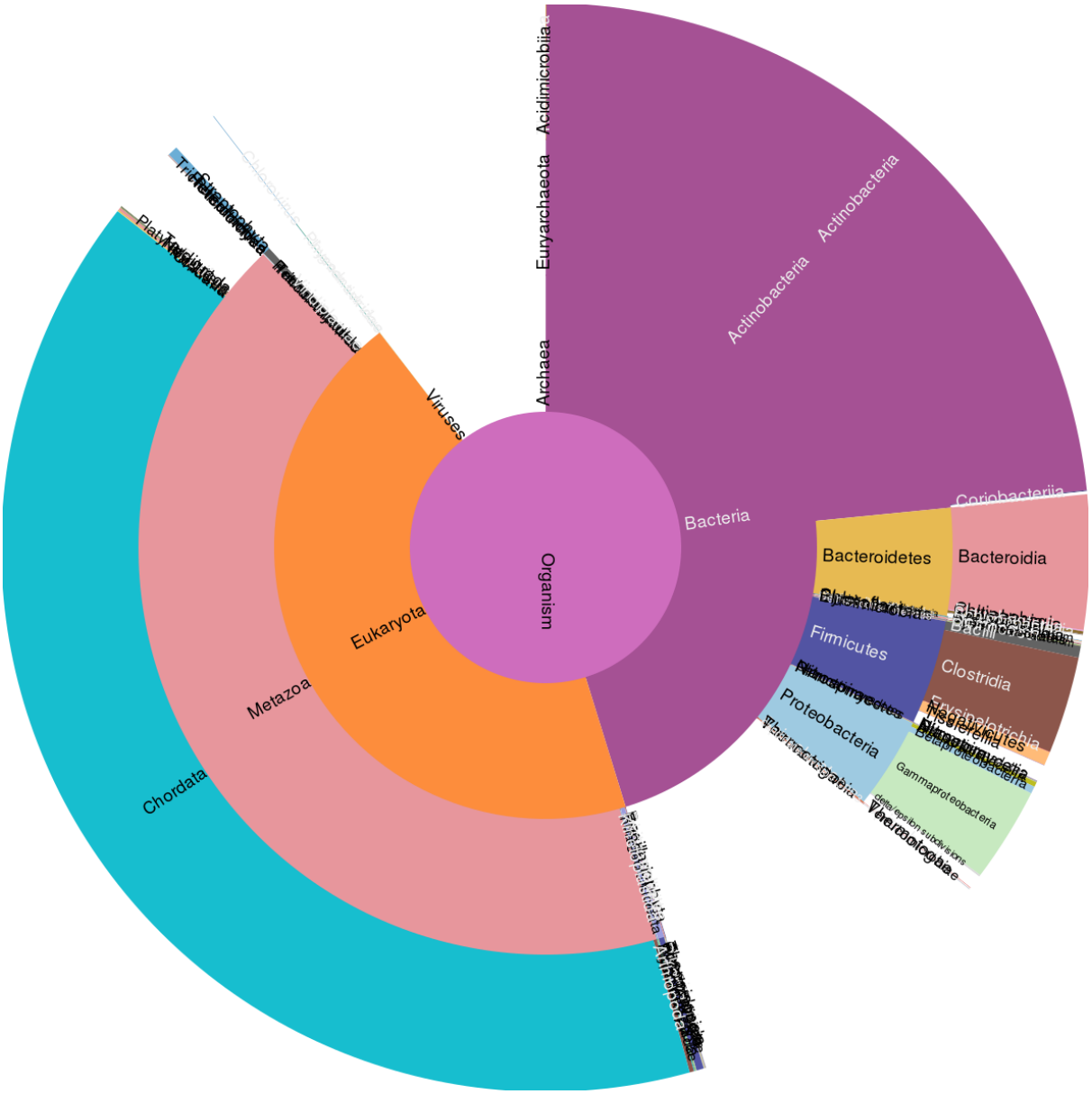


Figure 14:Sunburst chart generated on the UniPept platform providing an overview of the organism identified HEU birth, with 44%of the peptides identified belonging to Eukaryotes Mesuere et al. (2015) Proteomics doi:10.1002/pmic.201400361

Within the bacteria superkingdom, many the peptides belonged to the order Bifidobacteriales, Bacteriodales (phylum Bacteroidetes) and Clostidiales (Phylum Firmicutes) Fewer Lactobacillales peptides were identified than in the HUU peers at birth. The species identified belonging to the genus Bifidobacterium were *B. asteroides, B. bifidum, B. commune, B. psychraerophilum, B. huminatium* and *B. longum.*

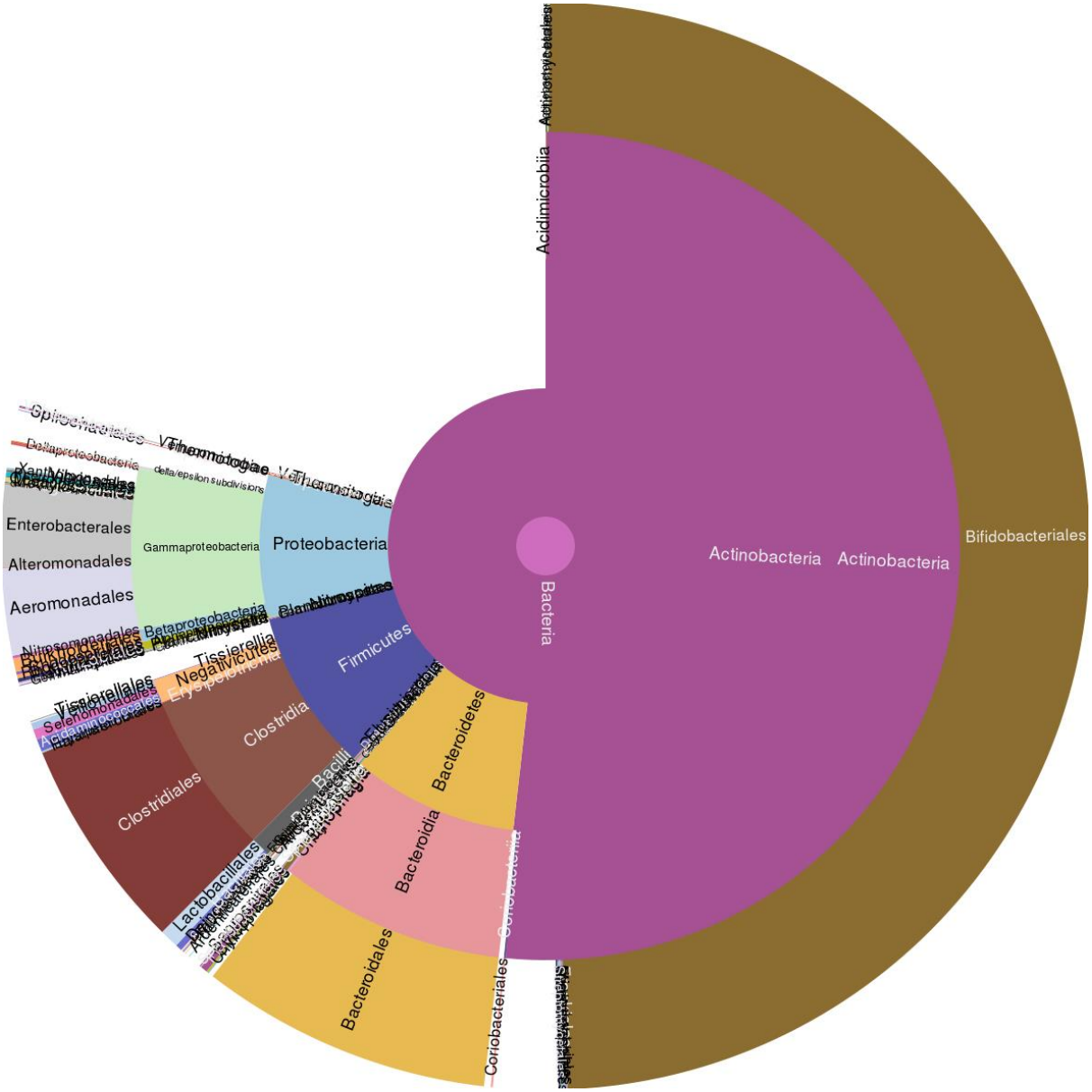


Figure 15:Expanded sunburst chart of the HEU birth microbiota , detailed on the bacteria ,45% of the peptides identified belong to the bacteria superkingdom Mesuere et al. (2015) Proteomics doi:10.1002/pmic.201400361

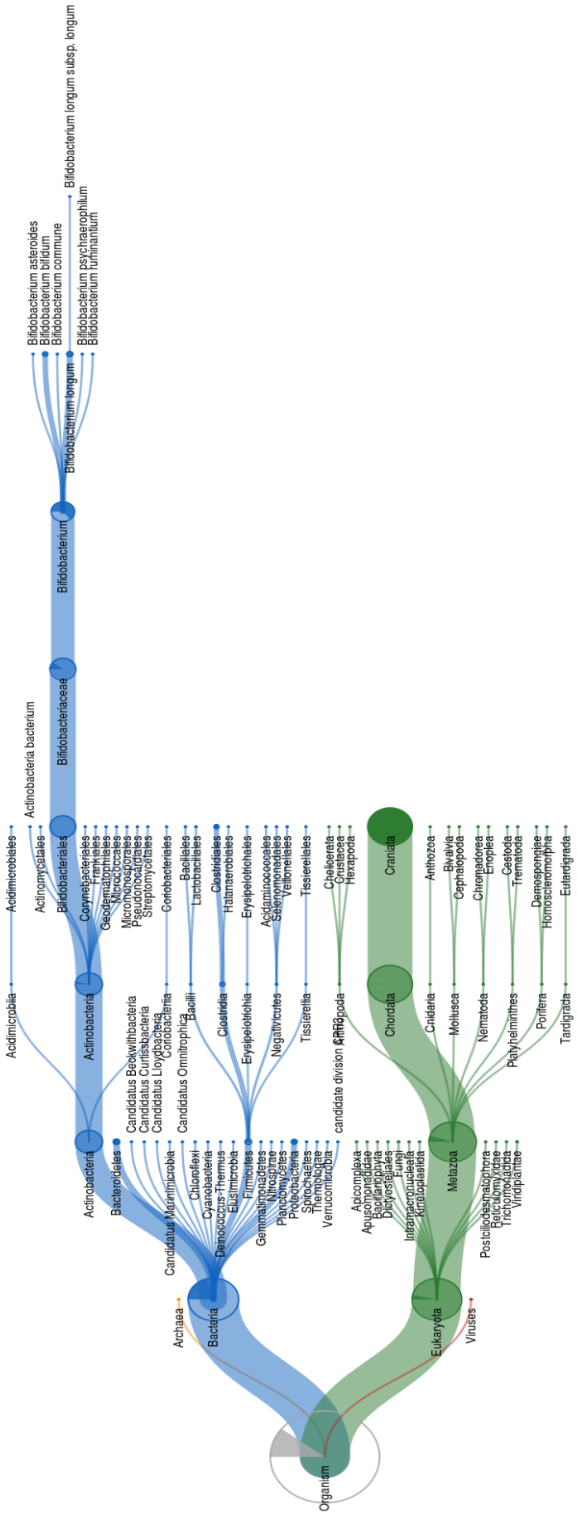


Figure 16: Tree map generated on UniPept platform of the organism present in the HEU-birth infants Mesuere et al. (2015) Proteomics doi:10.1002/pmic.201400361

Day 4-7 HUU

At day 4-7 the microbiome of the HUU infant is more diverse with a decrease in the number of peptides matched to Eukaryotes down to 37 % and an increase in the number of peptides matched to the Bacteria kingdom up to 54%. This change in the composition suggests that the microbiome of the HUU infant is primarily colonised after birth.

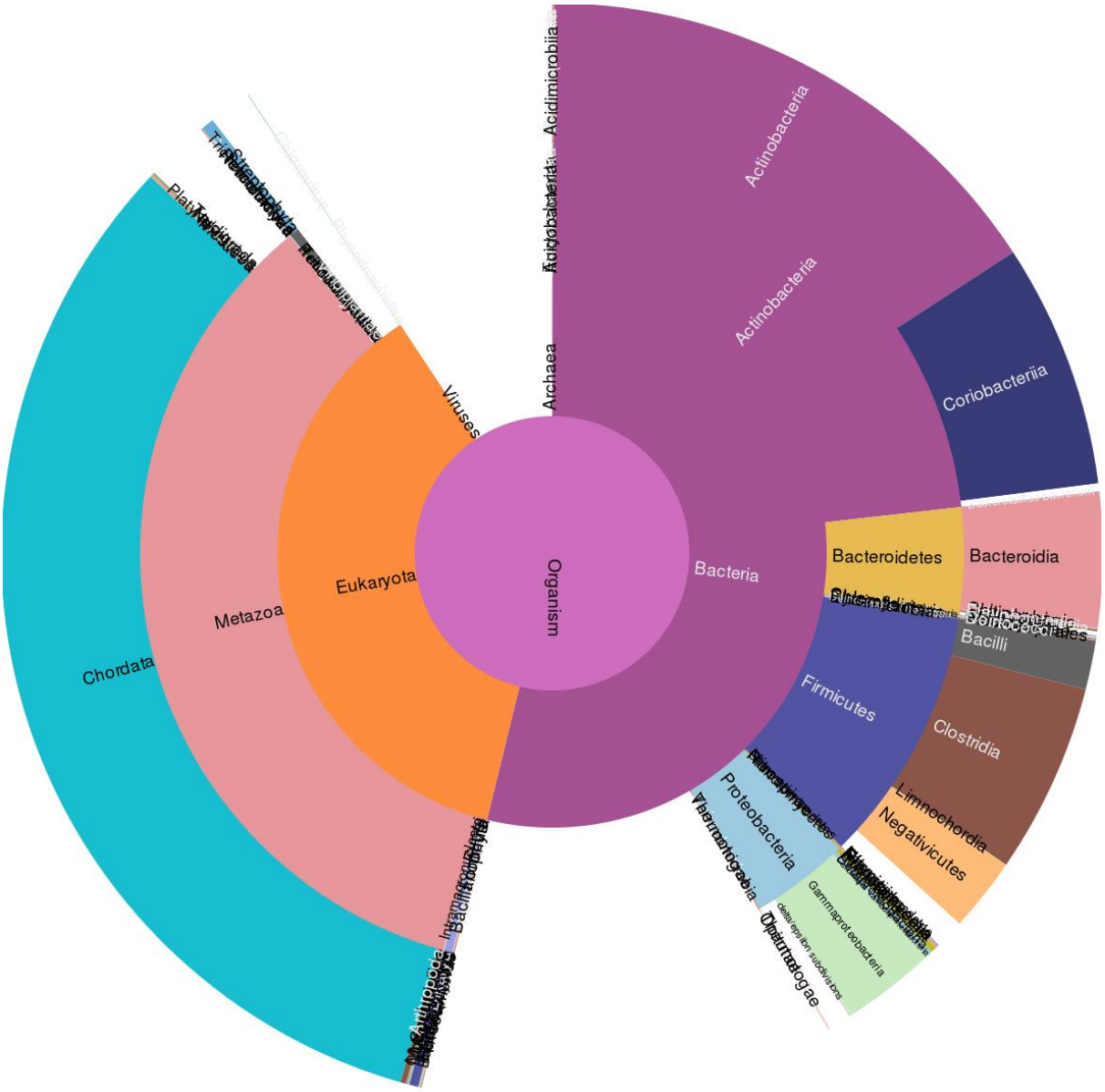


Figure 17: Sunburst chart providing an overview of the organisms identified in HUU D4-7 infants , with 37.% of the peptides identified belonging to Eukaryotes Mesuere et al. (2015) Proteomics doi:10.1002/pmic.201400361

The domain Bacteria is a large contributor to the infant microbiome Figure17

A large portion of the bacteria identified belonged to the class Actinobacteria. The largest contributor to this class was the order Bifidobacteriales.

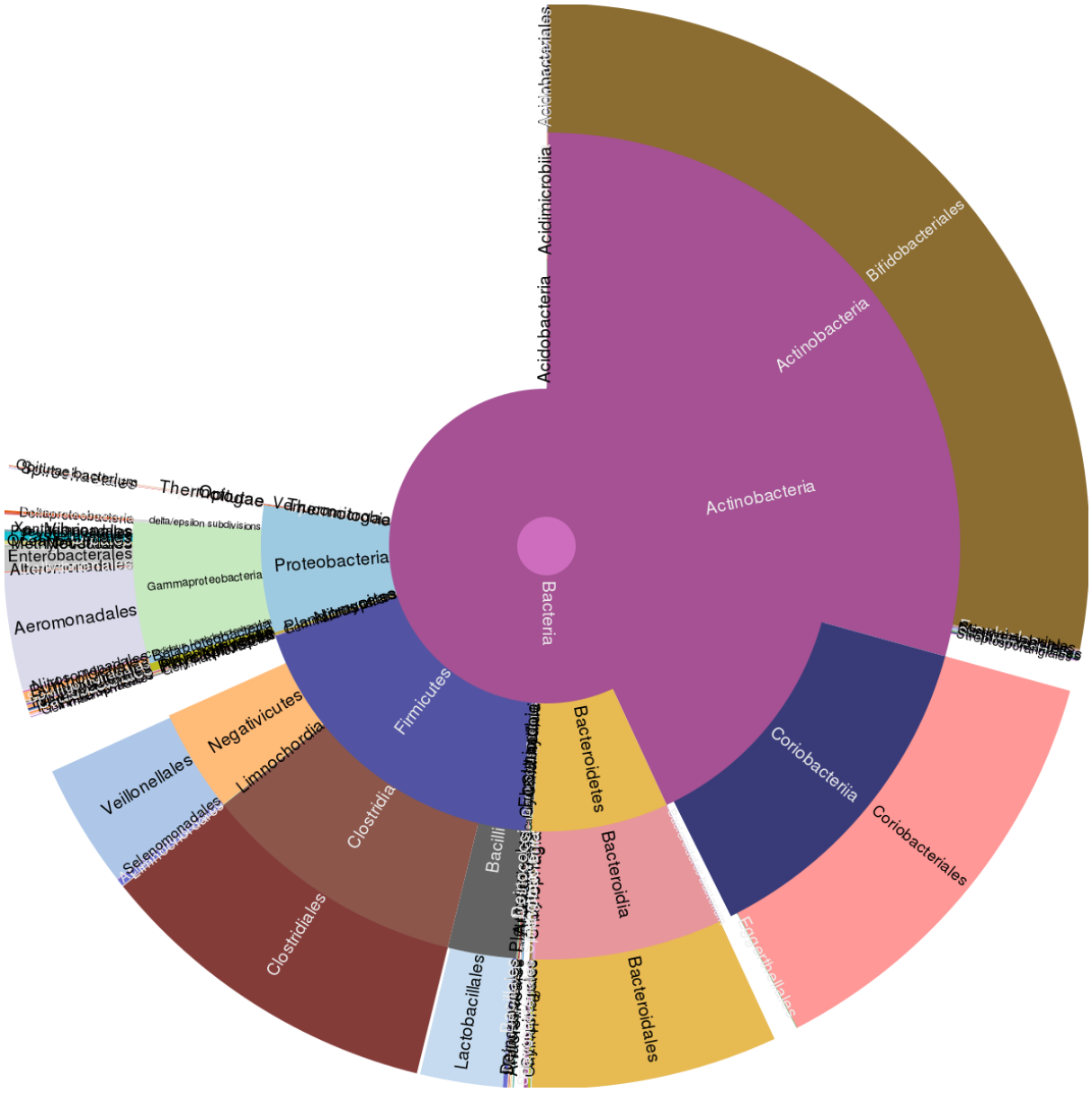


Figure 18: Expanded sunburst chart of the HUUD4-7, with details on the bacteria. 54% of the peptides identified belong to the bacteria superkingdom Mesuere et al. (2015) Proteomics doi:10.1002/pmic.201400361

The HUU microbiota is also more diverse than at birth, with the major contributors to the bacteria belonging to the Phyla Actinobacteria, Firmicutes and Proteobacteria. However the majority of the peptides identified belong to the order Bifidobacteriales , with an increase in the number of Bifidobacterium species from two as at birth to twelve species : *B. adolescentis, B. asteroides , B. bifidum, B. bohemicum, B. breve, B. callitrichos,*

B. longum with subspecies *B. longum, B.infantis B. mongoliense, B. pseudocatenulatum., B. psychraaerophilum, B. huminantium* and *B. scardovii*

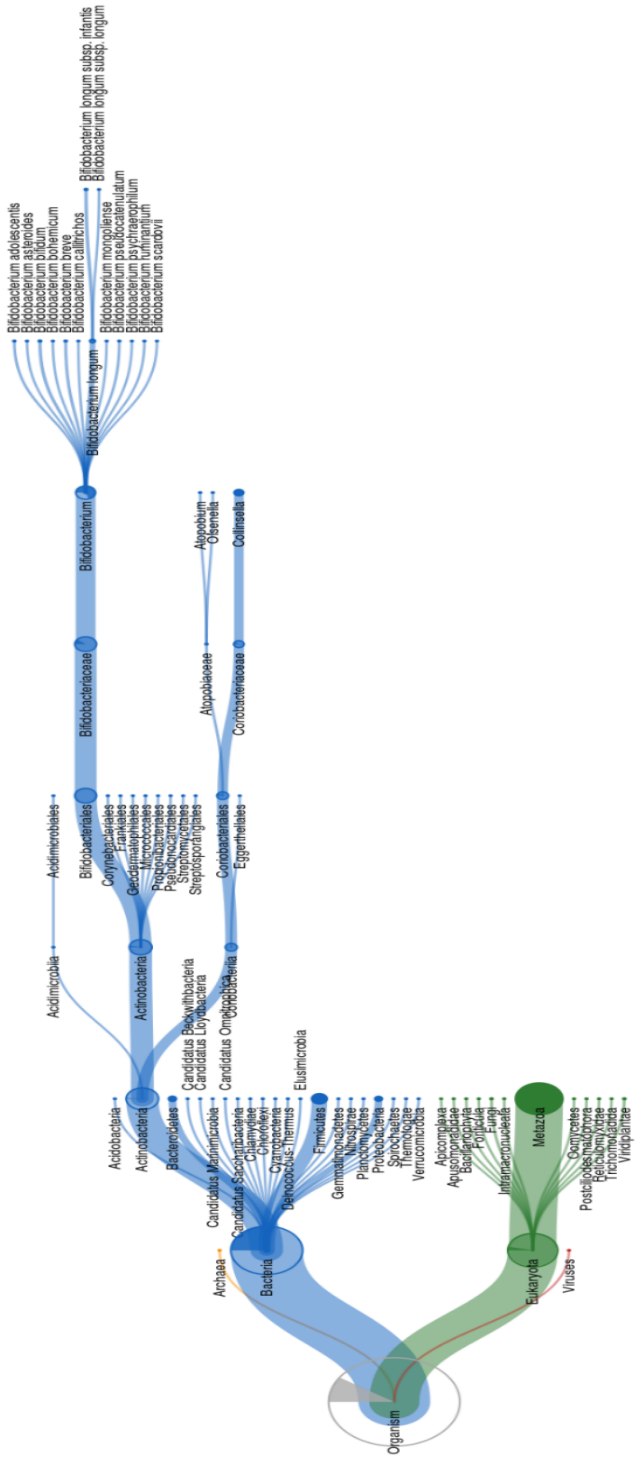


Figure19:Treeview of the organism identified in HUU d4-7.Mesuere et al. (2015) Proteomics doi:10.1002/pmic.201400361

Day4-7 HEU

At Day 4-7 in the HEU most peptides identified map to the kingdom bacteria, with only 27% of the peptides identified belonging to Eukaryotes. 61% of the peptides identified are bacterial derived and of that 61% the majority belong to the following phyla: Actinobacteria, Firmicutes and Proteobacteria.

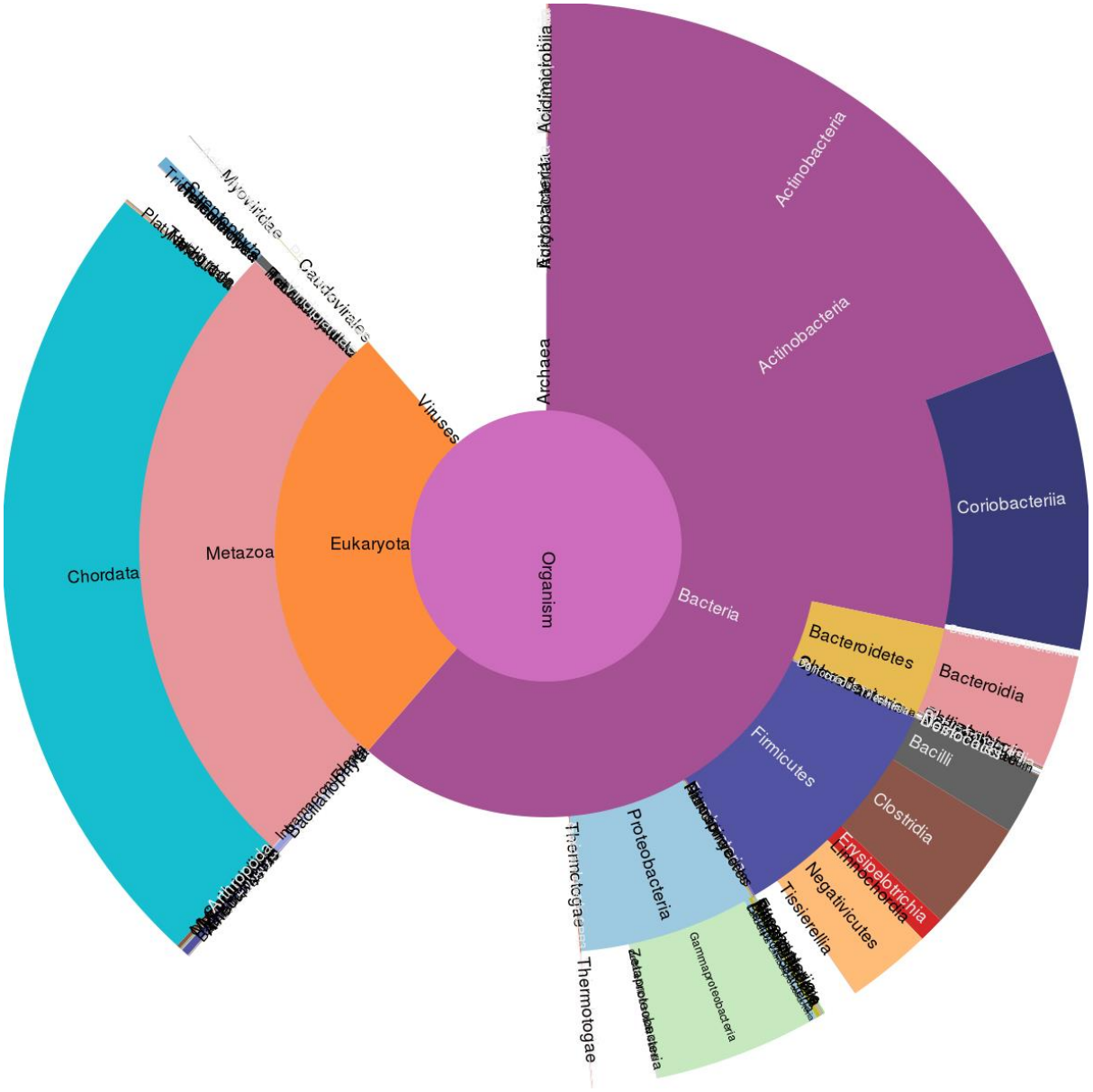


Figure 20: Sunburst chart providing an overview of the organism identified in the HEU. D4-7, 27% of peptides identified belonged to Eukaryotes Mesuere et al. (2015) Proteomics doi:10.1002/pmic.201400361

As with the previous groups many peptides identified belonged to the genus Bifidobacterium with the following species identified: *B. bifidum, B. bohemicum, B. breve, B. callitrichos, B. catenulatum, B. commune, B kashiwanohense. B. longum, B. psuedocatenulatum, B. psychraerophilum* and *B. huminatium.*. Interestingly

Mycobacterium tuberculosis complex was also detected, and this was also detected in the HUU birth stool samples

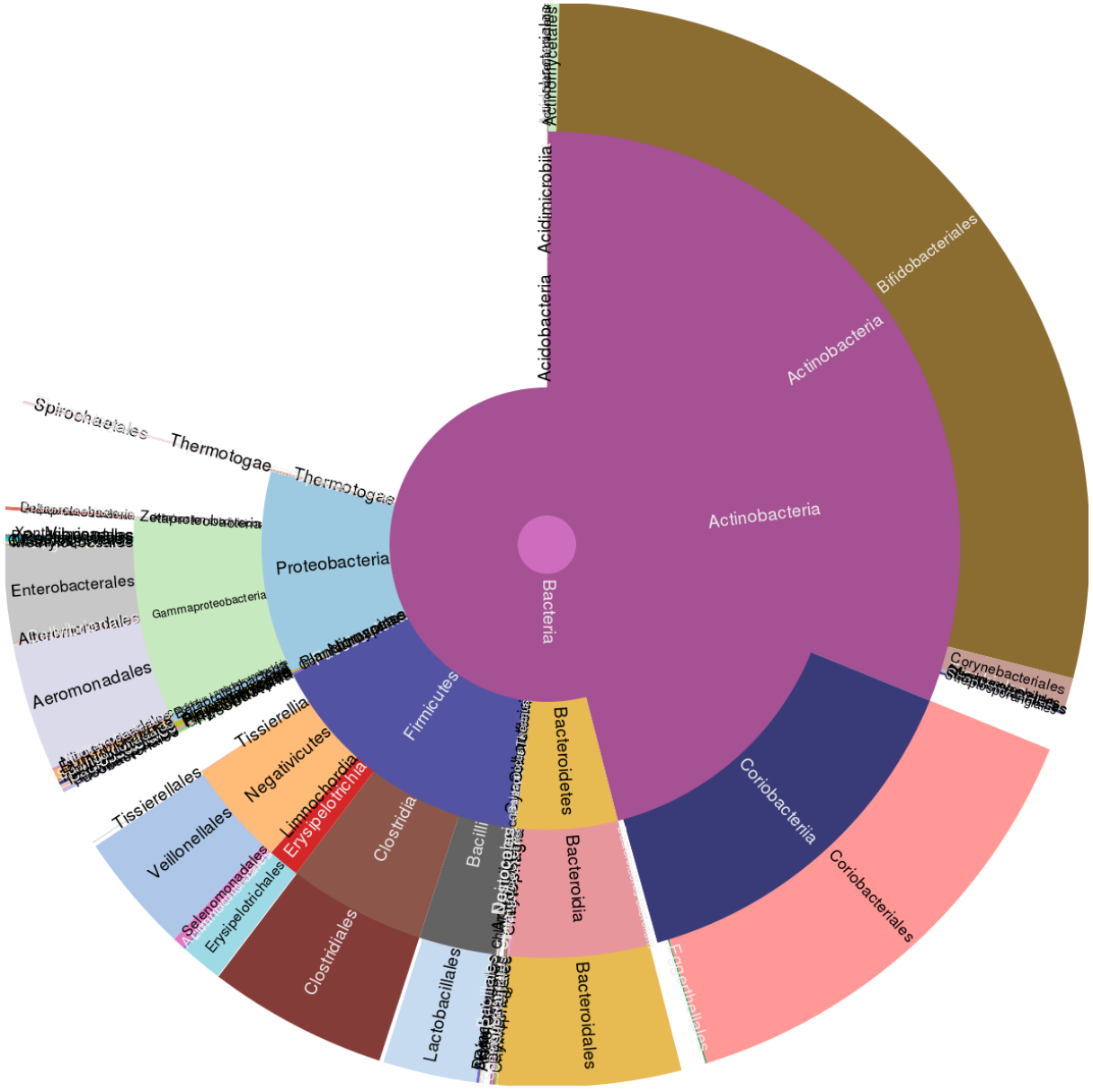


Figure 21: Expanded sunburst chart of the HEUD4-7 microbiota , having details on the bacteria, with 61% of the peptides identified belonging to the bacteria superkingdom (Mesuere et al. (2015) Proteomics doi:10.1002/pmic.201400361

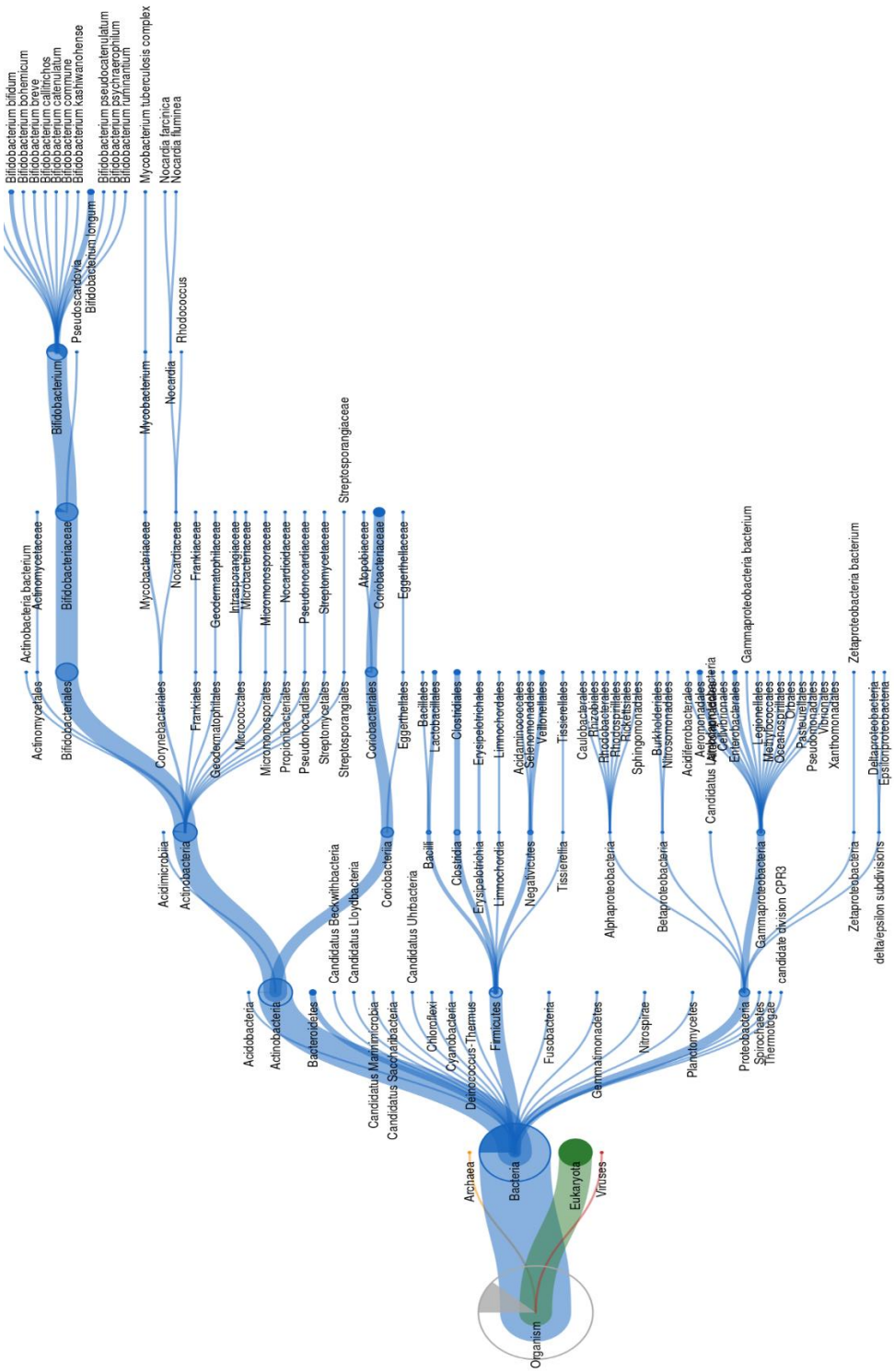


Figure22:Treemap of the organisms present in the HEU D4-7 infants (Mesuere et al. (2015) Proteomics doi:10.1002/pmic.201400361)

Table 6 Number of peptides identified that are associated with Mycobacterium Tuberculosis

Sample Group	Number of peptide sequences associated with mycobacterium tuberculosis
Birth HEU	0
Birth HUU	111
Day4-7 HEU	54
Day4-7 HUU	0

The following phylum that are associated with parasitic infections were also identified. Platyhelminthes, Mollusca, Nematoda and Cnidaria .

Table 7: Parasitic phyla identified

Phylum	Number of peptide sequences identified			
	Birth HEU	**Birth HUU**	**D4-7HEU**	**D4-7HUU**
Platyhelminthes	2	3	3	4
Mollusca	3	1	3	2
Nematoda	6	1	6	7
Cnidaria	1	1	0	0

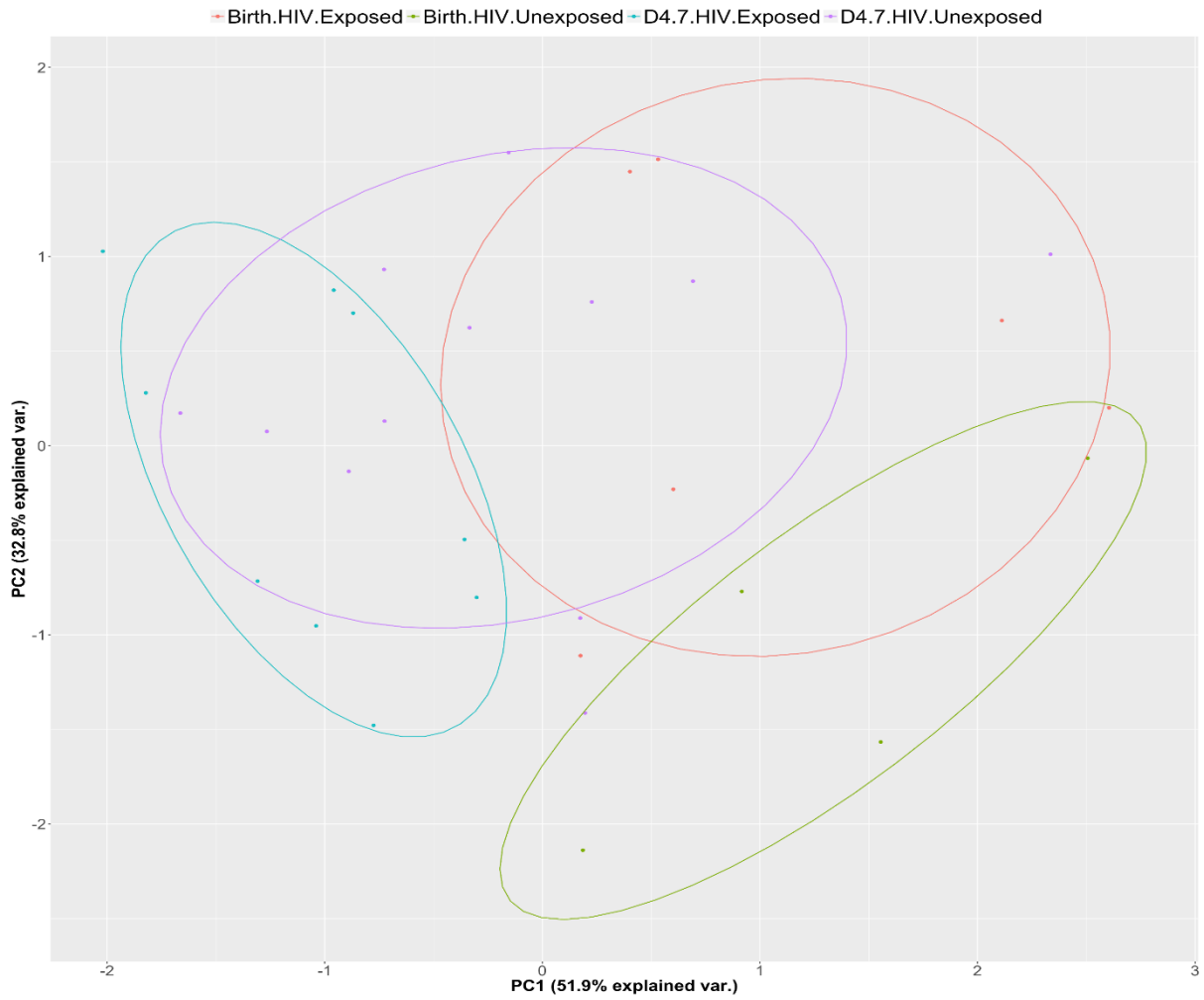


Figure 23: A PCA plot of how the samples cluster according to the phyla present within both groups at each timepoint. The PCA plot demonstrates the overlap in the samples, with the largest overlap occurring between the HEU birth and Day 4-7 HUU infants.

The overlap between the phyla present in the HEU and HUU infants at both time points reveals that the samples overlap greatest based on the timepoint and not on exposure status. The PCA also reveals that the samples cluster primarily according to time point.

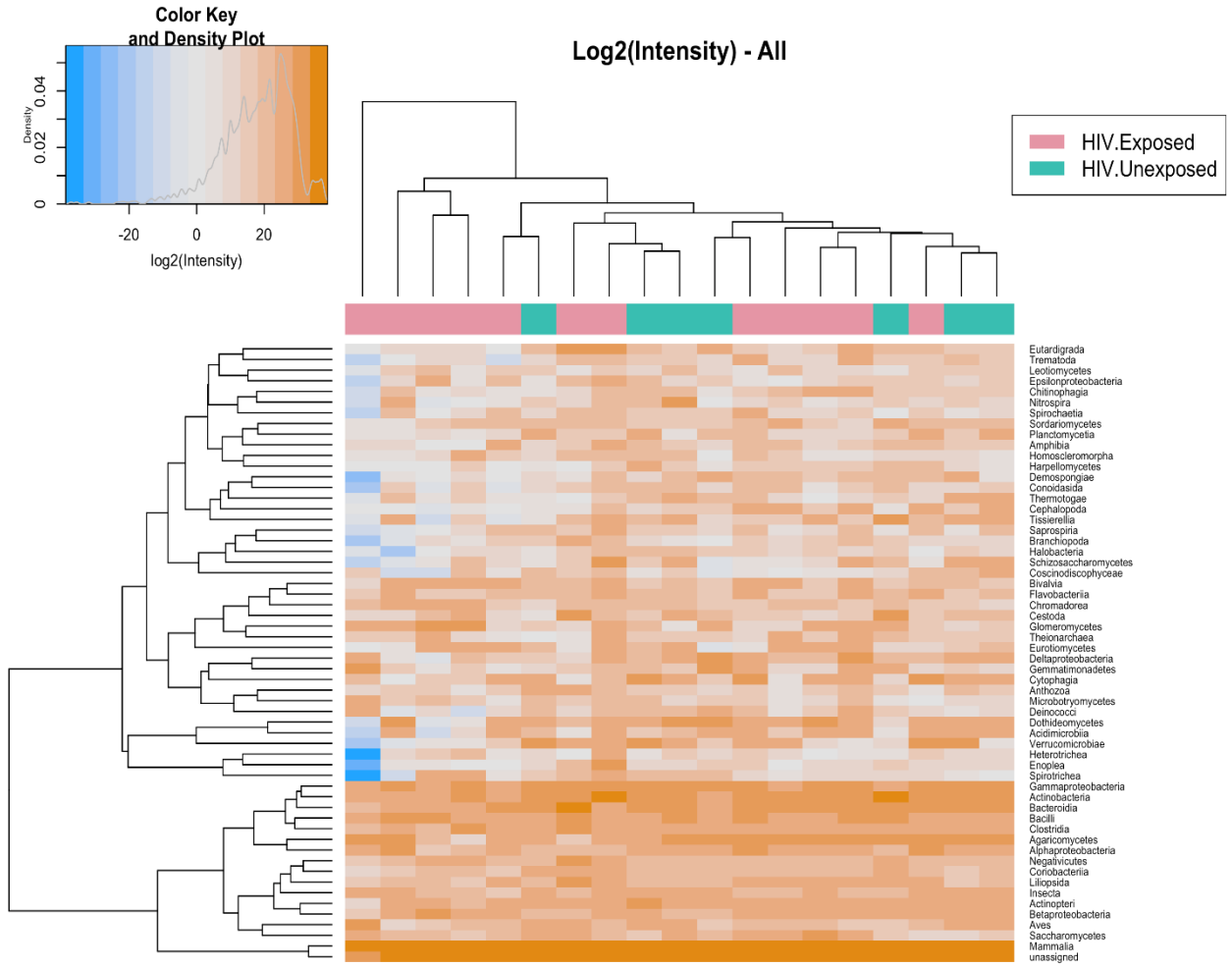


Figure 24: Hierarchical cluster chart of all the expressed classes of organisms in the HEU and HUU at birth

The hierarchical cluster chart at birth is based on the intensity of the peptides that were used to identify the class of organisms. From Figure 24 it is evident that there are a large number of classes, and interestingly the HEU and HUU samples do not cluster according to exposure state. This reiterates the interindividual microbiota of each infant as well as reaffirms the diversity within each sample. The HEU samples tend to cluster a bit more than the HUU samples.

The number of classes are less than at day 4-7, because meconium has a low biomass.

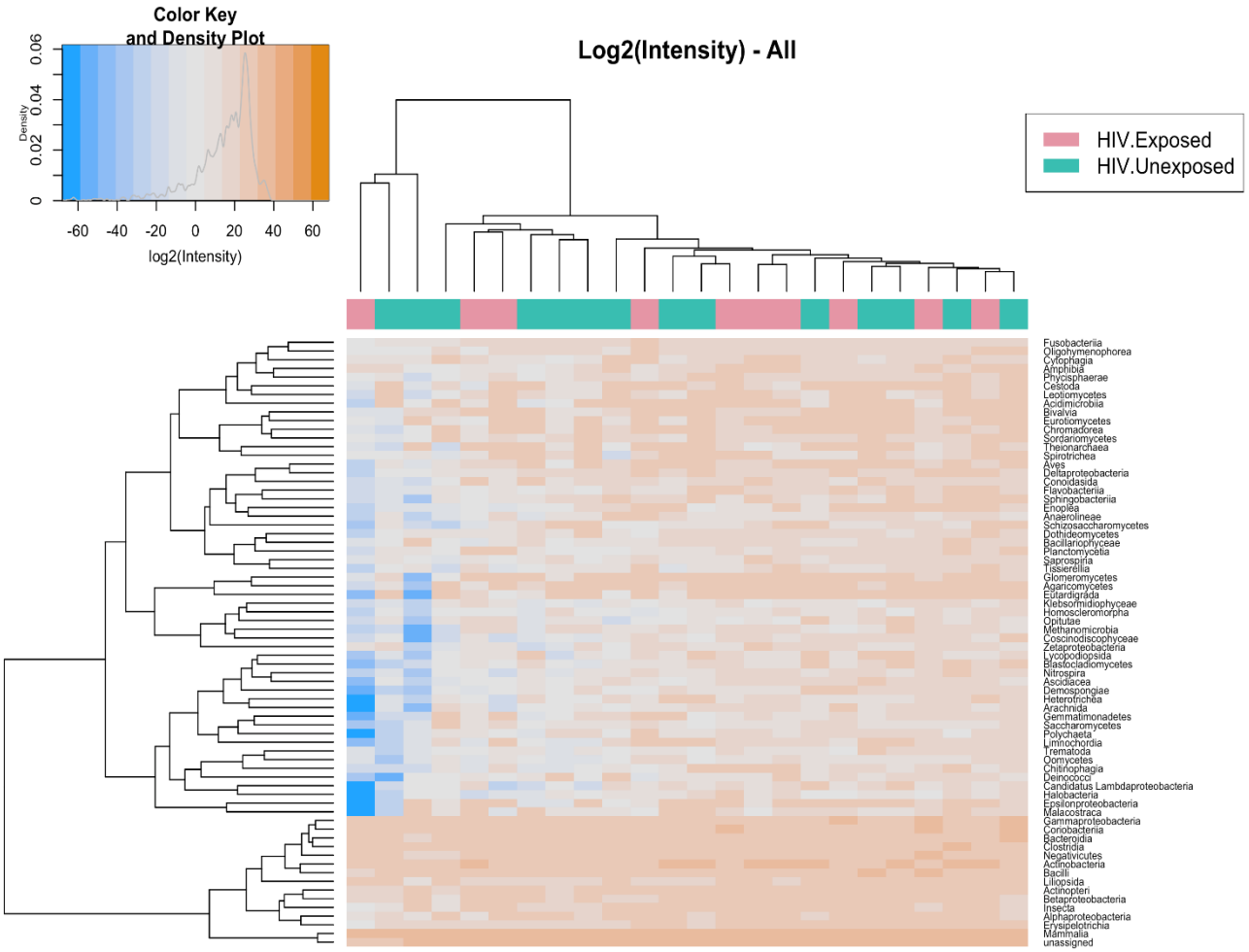


Figure 25: Hierarchical cluster chart of all the expressed classes of organism in HEU and HUU at day 4-7.

Whereas the Cluster map for the samples at day 4-7 reveals a large increase in the number of expressed classes it also reveals greater variation in the how the samples cluster. At day 4-7 the demarcation between the samples is more complex. The intensity of the signals is also not as strong as it is at birth, which may be due to the normalisation and the increase in the number of classes thus no one class is overexpressed.

3.3 Functional Analysis of metaproteomic data:

UniPept pathway analysis

Biological processes

The GO terms associated with the Chordata peptides were analysed on UniPept using the QuickGO plugin. Most of the peptides in the HEU at birth were associated with neutrophil degranulation, innate immune response and immune responses whereas in the HUU at birth the most abundant peptides matched to neutrophil degranulation, proteolysis and innate immune response.

Table 8 3: Top 10 Biological process for Chordata peptides identified on UniPept online in HEU birth.

Peptides	GO term	Name
341	GO:0043312	neutrophil degranulation
204	GO:0045087	innate immune response
197	GO:0006955	immune response
191	GO:0018149	peptide cross-linking
191	GO:0006958	complement activation, classical pathway
178	GO:0070268	Cornification
168	GO:0006508	Proteolysis
152	GO:0031424	Keratinization
148	GO:0042742	defense response to bacterium
146	GO:0006954	inflammatory response

Table 9: Top 10 Biological processes for Chordata peptides identified on UniPept online HUU birth.

Peptides	GO term	Name
249	GO:0043312	neutrophil degranulation
151	GO:0006508	Proteolysis
124	GO:0045087	innate immune response
119	GO:0018149	peptide cross-linking
115	GO:0006955	immune response
112	GO:0005975	carbohydrate metabolic process
110	GO:0030855	epithelial cell differentiation
105	GO:0031424	Keratinization
103	GO:0006958	complement activation, classical pathway
103	GO:0043066	negative regulation of apoptotic process

At day 4-7 the peptides matched to neutrophil degranulation, innate immune response and immune response in the HEU infant whereas in the HUU infant the top peptides matched to the carbohydrate metabolism, translation and glycolytic process.

Table 10: Top 10 Biological processes for Chordata peptides identified on UniPept online HEU day 4-7

Peptides	GO term	Name
322	GO:0043312	neutrophil degranulation
264	GO:0045087	innate immune response
253	GO:0006955	immune response
248	GO:0006958	complement activation, classical pathway

193	GO:0006911	phagocytosis, engulfment
189	GO:0042742	defense response to bacterium
168	GO:0005975	carbohydrate metabolic process
168	GO:0019731	antibacterial humoral response
167	GO:0006508	Proteolysis
163	GO:0050900	leukocyte migration

Table 11: Top 10 Biological processes for Chordata peptides identified on UniPept online HUU day 4-7.

Peptides	GO term	Name
526	GO:0005975	carbohydrate metabolic process
483	GO:0006412	Translation
308	GO:0006096	glycolytic process
203	GO:0008152	metabolic process
176	GO:0006094	Gluconeogenesis
120	GO:0055085	transmembrane transport
111	GO:0006006	glucose metabolic process
104	GO:0006520	cellular amino acid metabolic process
90	GO:0006457	protein folding
85	GO:0015986	ATP synthesis coupled proton transport

3.4 Protein analysis results

Pantherdb analysis

Further pathway analysis was carried out on the Pantherid, using the protein-level identifications. The proteinGroups.txt file for each of the groups at each time point was analysed and all non-human/Homosapien proteins were filtered out

The most abundant protein class in all the groups at both time points was hydrolase, followed by enzyme modulator/immunity proteins.

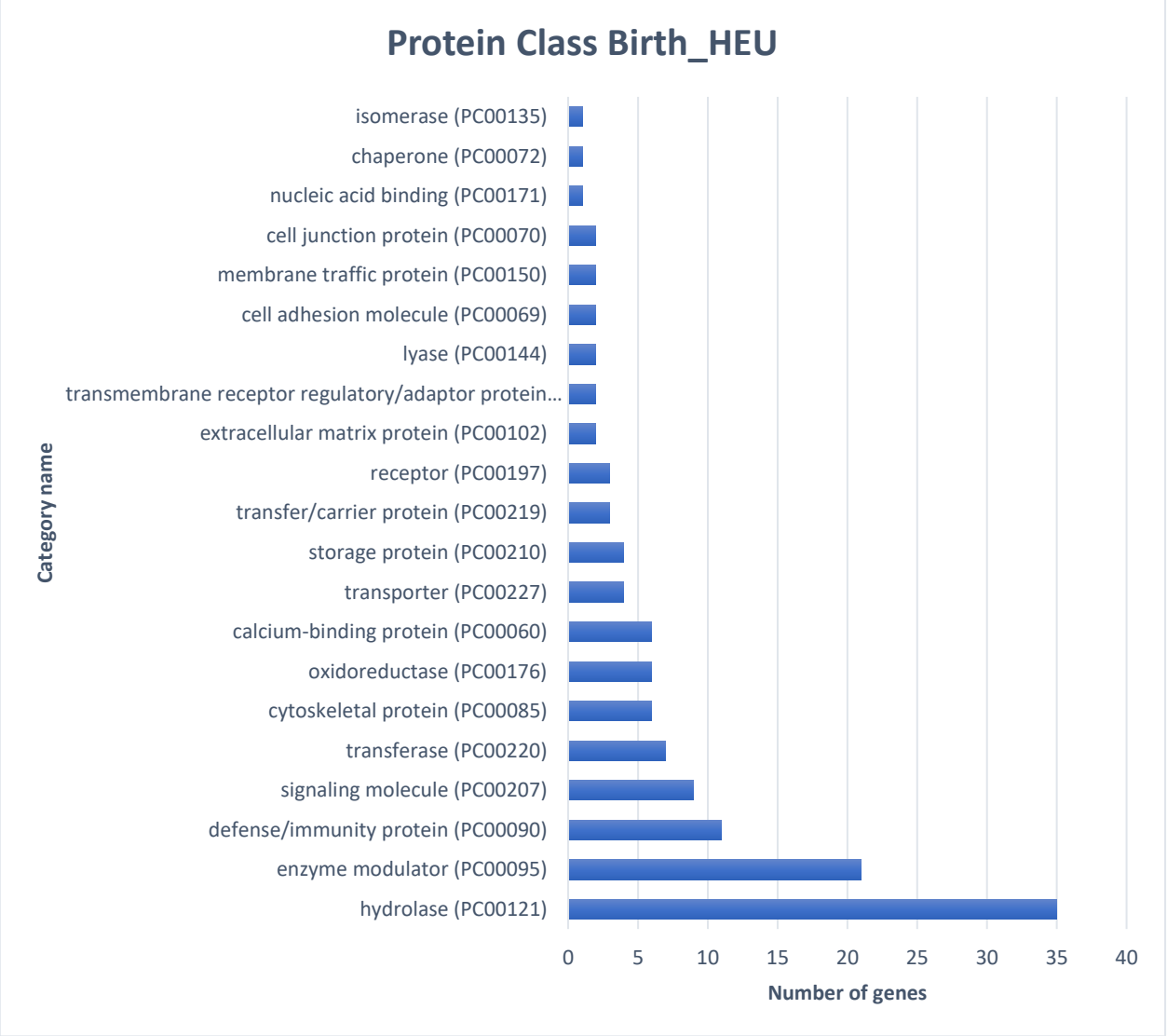


Figure 26: All protein classes identified in the Birth HEU infant group

The top protein classes identified in the Birth HEU group were hydrolases, enzyme modulator and defence/ immunity protein. The defence /immunity protein class is within the top three classes of protein identified

by Panther dB in the Birth HEU group. Whereas the defence /immunity protein class is the fourth most abundant protein class in the Birth HUU group

Table 12: Top 10 protein classes identified for Birth HEU infant group

Category name (Accession):	# genes	Percent of gene hit against total # genes	Percent of gene hit against total # Protein Class hits
hydrolase (PC00121)	35	18.1%	26.9%
enzyme modulator (PC00095)	21	10.9%	16.2%
defense/immunity protein (PC00090)	11	5.7%	8.5%
signaling molecule (PC00207)	9	4.7%	6.9%
transferase (PC00220)	7	3.6%	5.4%
cytoskeletal protein (PC00085)	6	3.1%	4.6%
oxidoreductase (PC00176)	6	3.1%	4.6%
calcium-binding protein (PC00060)	6	3.1%	4.6%
transporter (PC00227)	4	2.1%	3.1%
storage protein (PC00210)	4	2.1%	3.1%

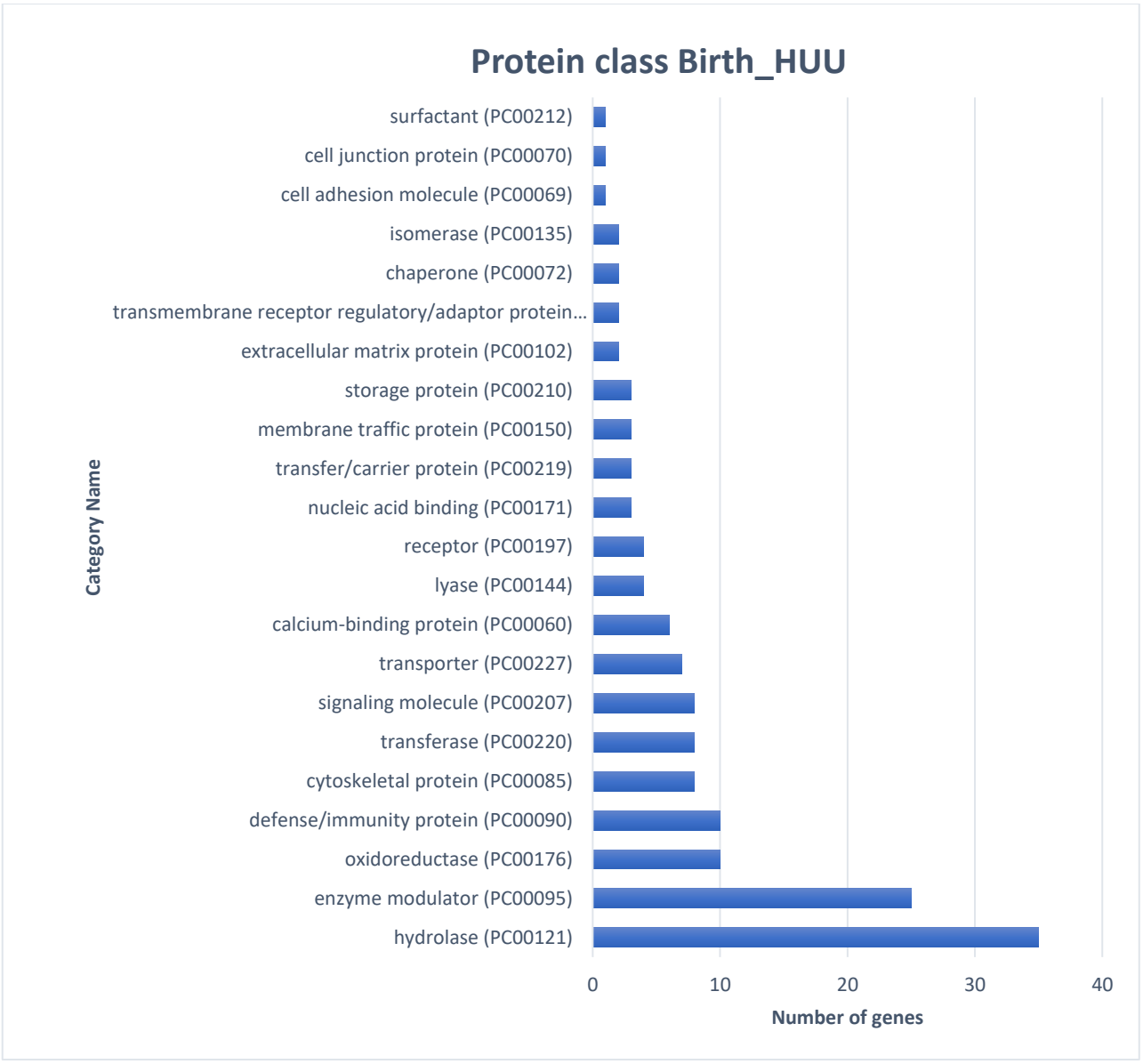


Figure 27 : All protein classes identified for Birth HUU infant group

Table 13: Top 10 protein classes identified for Birth HUU infant group

Category name (Accession):	# genes	Percent of gene hit against total # genes	Percent of gene hit against total # Protein Class hits
hydrolase (PC00121)	35	17.9%	23.6%
enzyme modulator (PC00095)	25	12.8%	16.9%
oxidoreductase (PC00176)	10	5.1%	6.8%

defense/immunity protein (PC00090)	10	5.1%	6.8%
cytoskeletal protein (PC00085)	8	4.1%	5.4%
transferase (PC00220)	8	4.1%	5.4%
signaling molecule (PC00207)	8	4.1%	5.4%
transporter (PC00227)	7	3.6%	4.7%
calcium-binding protein (PC00060)	6	3.1%	4.1%
lyase (PC00144)	4	2.1%	2.7%

At the Day4-7 timepoint the most abundant protein classes for both the HEU and HUU are the same . With the most abundant being hydrolase followed by enzyme modulator and then defence/immunity protein class

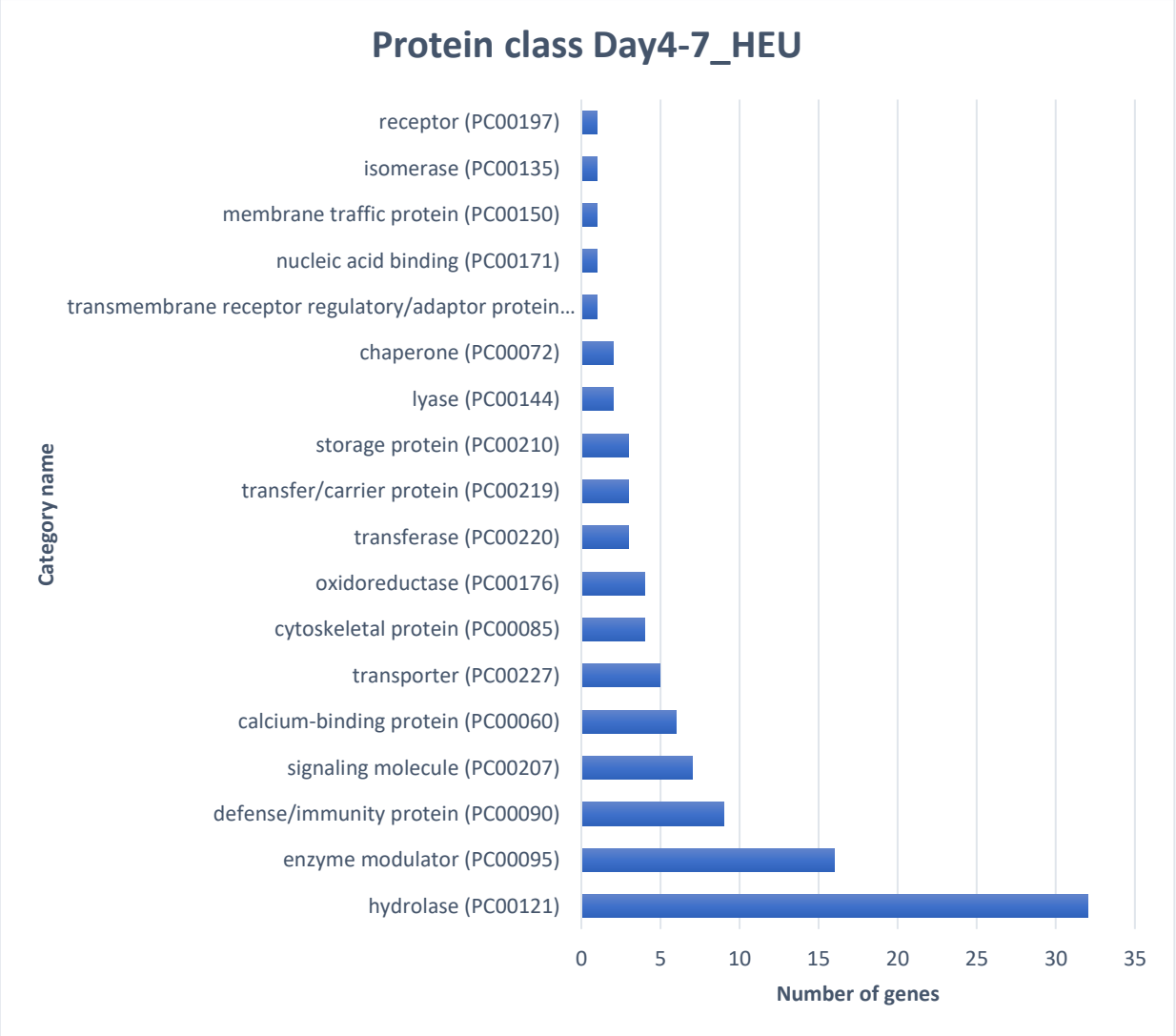


Figure 28: All protein classes identified for Day 4-7 HEU infant group

Table 14: Top10 Protein Classes identified Day4-7_HEU

Category name (Accession):	# genes	Percent of gene hit against total # genes	Percent of gene hit against total # Protein Class hits
hydrolase (PC00121)	32	20.6%	31.7%
enzyme modulator (PC00095)	16	10.3%	15.8%
defense/immunity protein (PC00090)	9	5.8%	8.9%
signaling molecule (PC00207)	7	4.5%	6.9%

calcium-binding protein (PC00060)	6	3.9%	5.9%
transporter (PC00227)	5	3.2%	5.0%
cytoskeletal protein (PC00085)	4	2.6%	4.0%
oxidoreductase (PC00176)	4	2.6%	4.0%
transferase (PC00220)	3	1.9%	3.0%
transfer/carrier protein (PC00219)	3	1.9%	3.0%

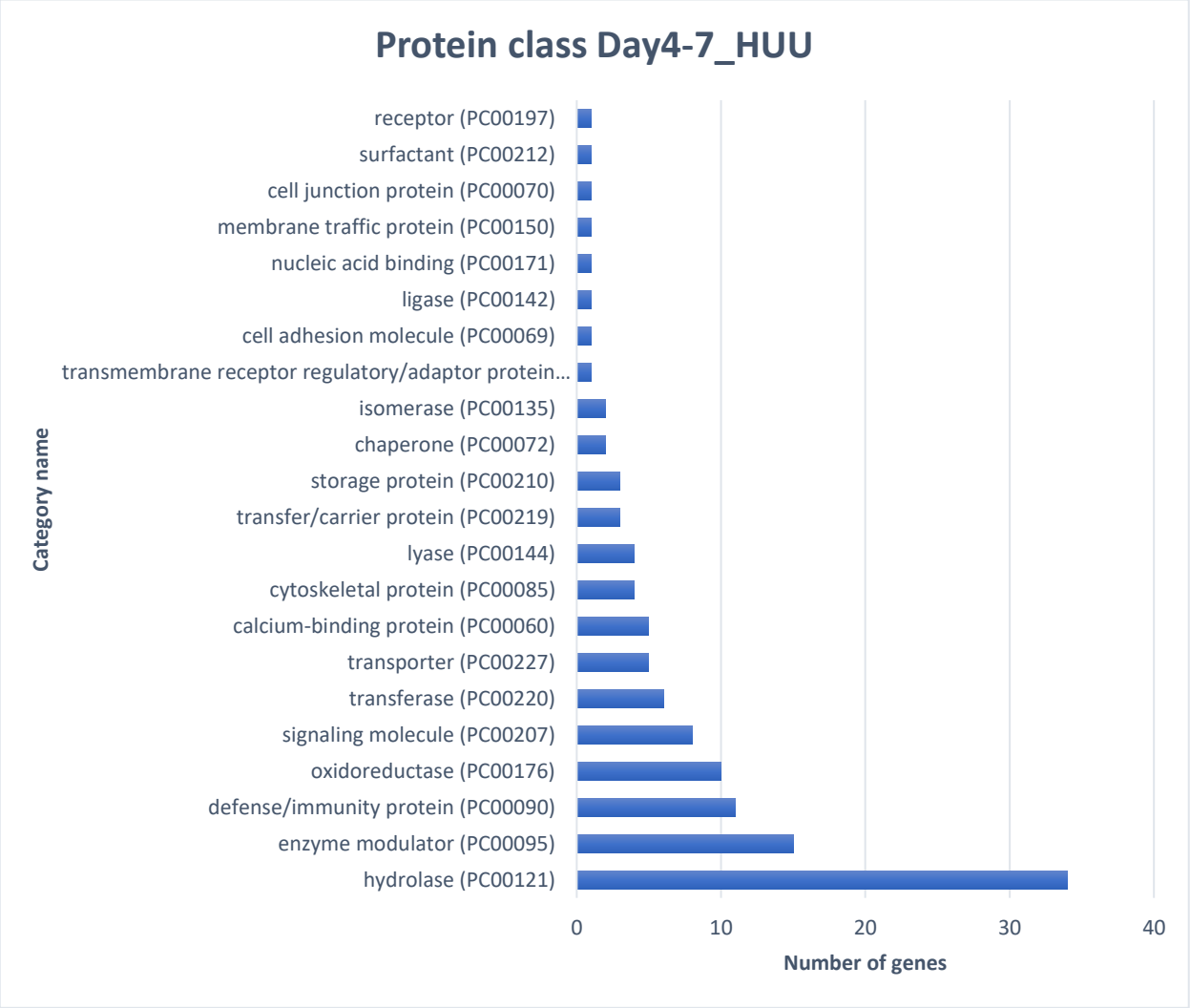
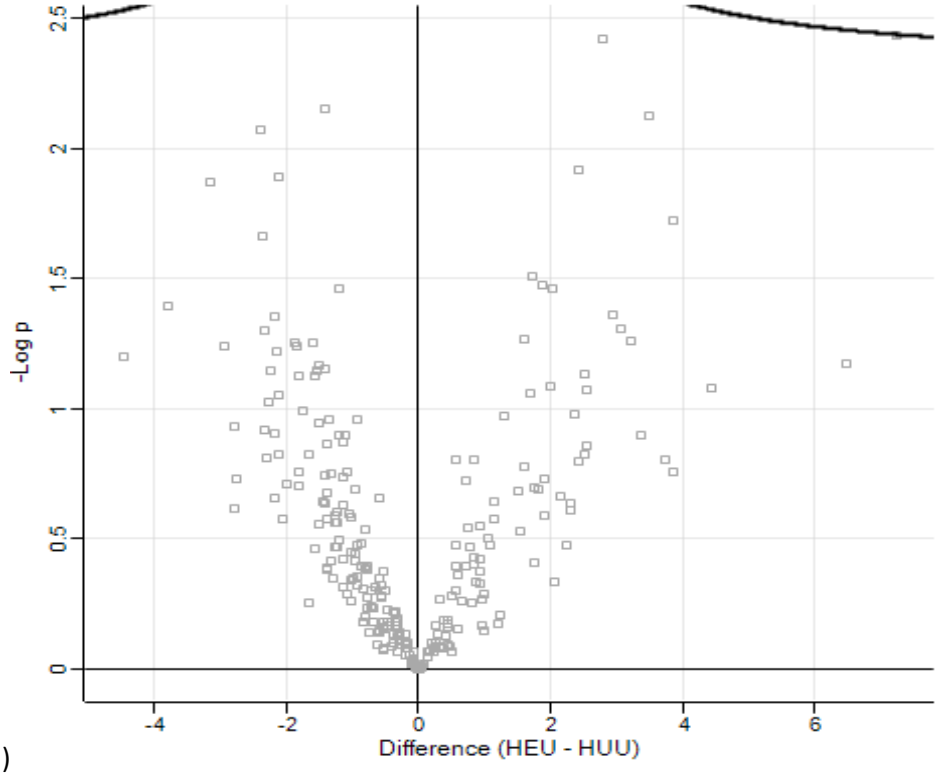


Figure 29: All protein classes identified for Day 4-7 HUU infant group

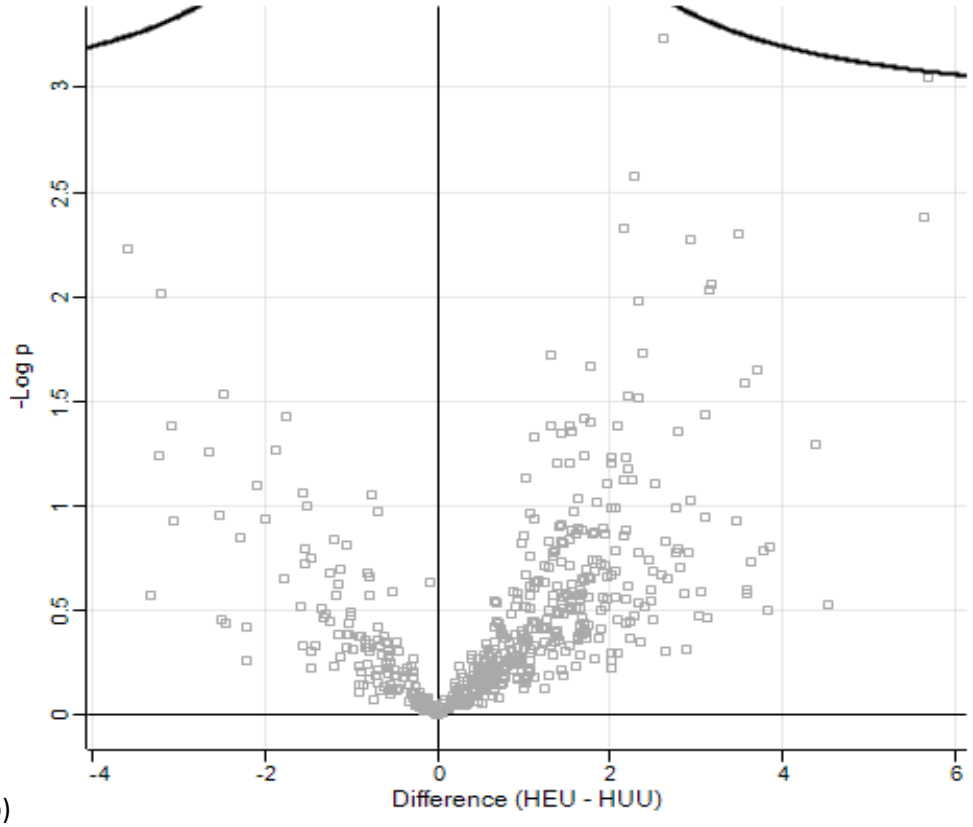
Table 15: Top 10 protein class identified Day4-7 HUU

Category name (Accession):	# genes	Percent of gene hit against total # genes	Percent of gene hit against total # Protein Class hits
hydrolase (PC00121)	34	19.7%	28.3%
enzyme modulator (PC00095)	15	8.7%	12.5%
defense/immunity protein (PC00090)	11	6.4%	9.2%
oxidoreductase (PC00176)	10	5.8%	8.3%
signaling molecule (PC00207)	8	4.6%	6.7%
transferase (PC00220)	6	3.5%	5.0%
transporter (PC00227)	5	2.9%	4.2%
calcium-binding protein (PC00060)	5	2.9%	4.2%
cytoskeletal protein (PC00085)	4	2.3%	3.3%

Volcano plots (Figure30) for proteins identified were generated on the Perseus platform; no significantly expressed proteins were identified in either one the groups, at each timepoint after multiple testing correction – likely reflecting the relatively high variability between metaproteomes from different individuals within each group and the relatively small number of samples per group. Thus a presence-absence analysis was carried out.



a)



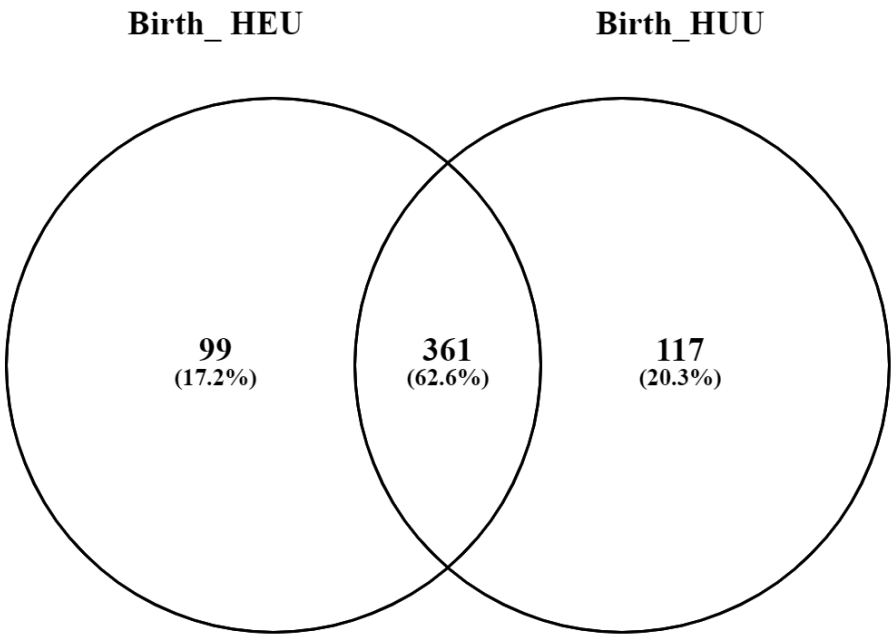
(b)

Figure 30: Volcano plots generated on Perseus (a)Birth HEU vs HUU (b) day 4-7 HEU vs HUU

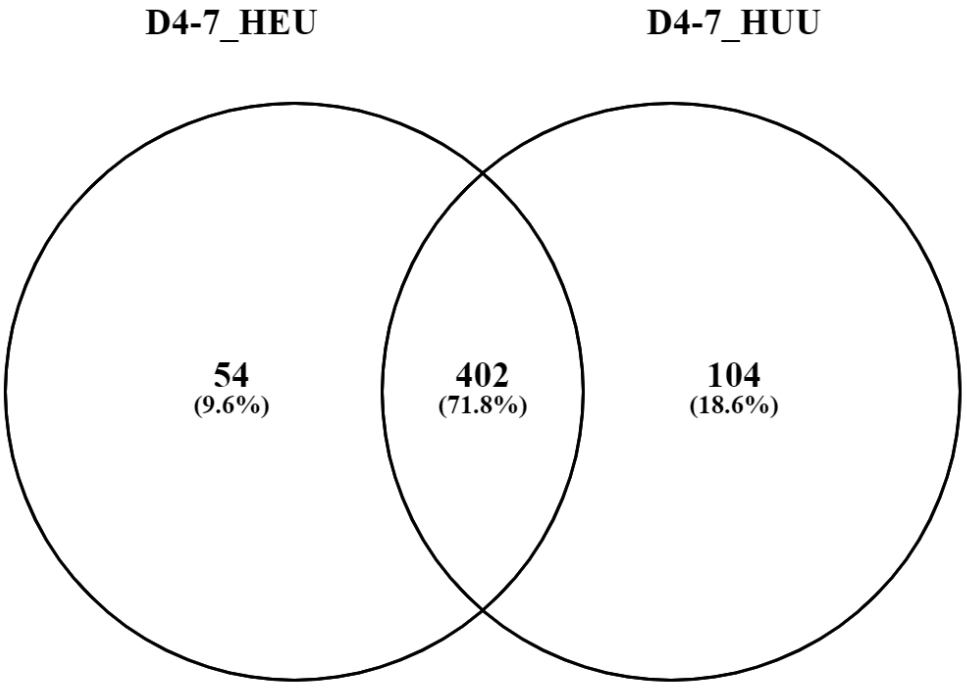
Protein presence absences analysis

A presence-absence analysis was carried out on the proteins that were identified from the *ProteinGroups.txt* file, on the Venny 2.0 web-based platform.

A total of 577 Human proteins were identified at the Birth timepoint: 361 (62.6%) of the proteins identified were shared by both the HEU and HUU infants at birth; 99 (17.2%) were exclusively found in the Birth HEU group; and 117 (20.3%) were exclusively found in the Birth HUU group. A total of 560 human proteins were identified in the Day 4-7 group. Of the proteins identified, 402 (71.8%) were shared amongst the HEU and HUU infants, with 104 (18.6%) of the proteins identified exclusively found in the HUU group and 54 (9.6%) were exclusive to the HEU group.



(a)



(b)

Figure 31: Venn diagram of the protein groups that overlap at the different timepoints (a) Birth timepoint (b) Day 4-7 timepoint . (*Oliveros, J.C. (2007-2015) Venny. An interactive toolfor comparing lists withVenn'sdiagrams.* https://bioinfogp.cnb.csic.es/tools/venny/index.htm)

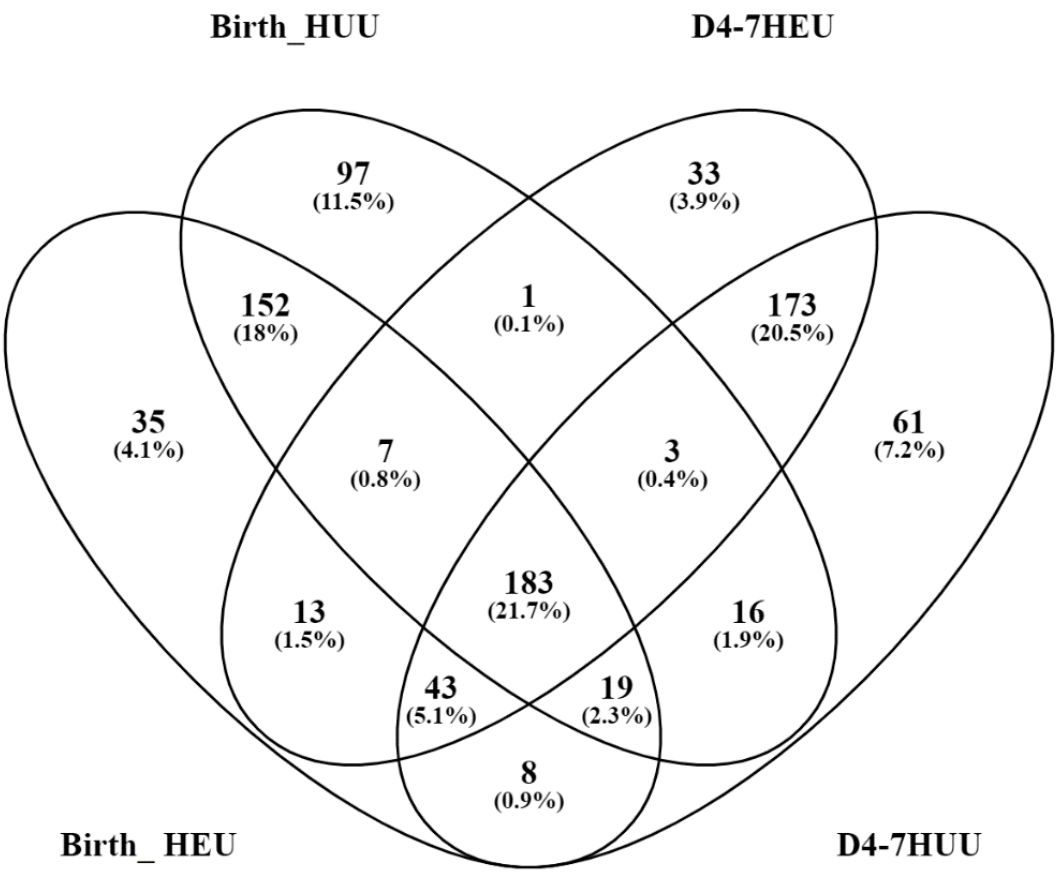


Figure 32: Venn diagram of all the Human proteins shared in HEU and HUU at both the Birth and Day 4-7 timepoint. (Oliveros, J.C. (2007-2015) Venny. An interactive toolfor comparing lists withVenn'sdiagrams. https://bioinfogp.cnb.csic.es/tools/venny/index.htm)

Combined Venn diagram was generated of all the Human proteins identified .183 (21.7%) of the proteins uploaded were common in both HEU and HUU groups at both time points. 97 (11.5%) were exclusive to the Birth HUU and 35 (4.1%)) were exclusive to Birth HEU. Whereas 33 (3.9%) were exclusive to the Day4-7 HEU and 61 (7.2%) were exclusive to the Day4-7 HUU .

From the data generated in the Venn diagram the human proteins that were exclusively found in each group at each timepoint was uploaded into String online platform, to identify the top biological processes. At birth HEU 21 of the 35 proteins were identified, For Birth HUU 56 of 97 proteins were identified. Day4-7_HEU 8 of 33, At Day 4-7 HUU 21 of 61 proteins were identified

Table 16: Top 3Go enrichment Biological process exclusive Birth HEU

#term ID	term description	observed gene count	background gene count	false discovery rate
GO:0043312	neutrophil degranulation	6	485	0.0071
GO:0043588	skin development	4	373	0.0244
GO:0008544	epidermis development	4	403	0.0294

Table 17: Top 10 Go enrichment Biological process exclusive to Birth HUU

#term ID	term description	observed gene count	background gene count	false discovery rate
GO:0016999	antibiotic metabolic process	9	124	2.76e-07
GO:0017144	drug metabolic process	13	622	2.06e-05
GO:0065008	regulation of biological quality	28	3559	4.84e-05
GO:0006091	generation of precursor metabolites and energy	10	388	6.77e-05
GO:0016192	vesicle-mediated transport	19	1699	6.77e-05
GO:0051234	establishment of localization	30	4248	6.79e-05
GO:0006810	transport	29	4130	0.00013
GO:0045055	regulated exocytosis	12	691	0.00014
GO:0051186	cofactor metabolic process	10	467	0.00019
GO:0090066	regulation of anatomical structure size	10	464	0.00019

Table 18: Top 10 Biological process exclusive to day4-7HEU

#term ID	term description	observed gene count	background gene count	false discovery rate
GO:0006959	humoral immune response	4	252	0.0019
GO:0019730	antimicrobial humoral response	3	143	0.0109
GO:0065008	regulation of biological quality	7	3559	0.0146
GO:0042742	defence response to bacterium	3	250	0.0280
GO:0010823	negative regulation of mitochondrion organization	2	50	0.0335

GO:0019731	antibacterial humoral response	2	47	0.0335
GO:0050832	defence response to fungus	2	49	0.0335
GO:0043687	post-translational protein modification	3	365	0.0349
GO:1903201	regulation of oxidative stress-induced cell death	2	65	0.0349
GO:0090559	regulation of membrane permeability	2	79	0.0383

Table 19 Top 7 Biological process exclusive to day4-7HUU

#term ID	term description	observed gene count	background gene count	false discovery rate
GO:0006887	exocytosis	9	774	1.87e-05
GO:0043312	neutrophil degranulation	8	485	1.87e-05
GO:0006880	intracellular sequestering of iron ion	2	5	0.00092
GO:0051651	maintenance of location in cell	3	83	0.0034
GO:0097237	cellular response to toxic substance	3	195	0.0353
GO:0009636	response to toxic substance	4	468	0.0414
GO:0006810	transport	11	4130	0.0434

Table 20: Top 10 Biological processes common in all four groups .

#term ID	term description	observed gene count	background gene count	false discovery rate
GO:0046903	secretion	34	1070	1.03e-19
GO:0002446	neutrophil mediated immunity	26	498	1.37e-19
GO:0045055	regulated exocytosis	29	691	1.37e-19
GO:0043312	neutrophil degranulation	25	485	8.08e-19
GO:0032940	secretion by cell	31	959	1.02e-18
GO:0001775	cell activation	29	1024	5.40e-16
GO:0045321	leukocyte activation	27	894	2.15e-15
GO:0002252	immune effector process	27	927	4.93e-15
GO:0002376	immune system process	39	2370	7.86e-15
GO:0006955	immune response	32	1560	4.26e-14

Chapter 4 Discussion

The purpose of this study was to elucidate the microbiome in HEU and HUU infants using metaproteomic approaches. The study was designed to assess the microbial proteome as well as the host proteomes.

Although the use metaproteomics allows for the identification of novel species in the gut, the study space is still plagued by a low rate of spectral matches identified, with only 7.6% and 4.9% of the MS/MS spectra identified in the birth and Day4-7 respectively. The loss of data is further exacerbated by the database generated by MetaNovo. Of the peptides uploaded, 134, 837,1401 and 2160 in Birth HUU, Birth HEU D4-7HUU and D4-7HEU respectively were unassigned. Thus, a wealth of information is left out of the downstream analyses.

4.1Taxonomic analysis of metaproteomic data

From the results it is evident that there is a difference in the composition of the HEU infant microbiome compared to their HUU peers (Figure 9**)**.Whilst it is not unexpected that a chronic inflammatory condition would result in an altered maternal microbiome, it is of interest to see how this affects the composition of the HEU infants compared to their HUU peers.

This study revealed that the microbiome at birth of the HUU infants is primarily comprised of human proteins (Figure 11), whereas the HEU has a larger amount of microbiota present (Figure 14). The study also revealed that although the relative abundance of bacteria is less in HUU, the infant does not develop in a sterile environment, as the birth stool sample, the meconium, is often used as a proxy for the in-utero environment. Similar findings have been found and reported by (Jiménez *et al.*, 2008; Collado *et al.*, 2016; Zhu *et al.*, 2018).

The findings suggest that the healthy in utero environment is not heavily colonized by bacteria at birth. The increased bacterial presence in the HEU infant may be an indication of a compromised in utero environment.

This study identified a large abundance of bacteria in the HUU at birth that belonged to the Family Bifidobacteriacae and Corynebacterceae, whereas Zhu and colleagues (Zhu *et al.*, 2018) identified a majority of bacteria that belonged to the family Enterobactericae in healthy Chinese infants. However, this study

investigated the microbiota of faecal samples in infants in South Africa, whereas Zhu and colleagues, 2018 examined amniotic fluid. Nevertheless a study of infant's stool microbiome by (Nazmul Huda *et al.*, 2014) also identified Actinobacteria , specifically Bifidobaceriaceae as the most abundant family in their cohort of infants in Bangladesh, similar to the family identified in this study. This suggests that the infant meconium has a unique microbiome that is independent of the amniotic fluid or the placenta and may not be the most accurate proxy of the in-utero environment as was previously believed. This then raises the question of what the origin of the gut infant microbiome in-utero is.

The HEU and the HUU infant stool samples at birth both (Figure13 and Figure 16**)** revealed the presence of Metazoa, namely *Cnidaria, Mollusca, Nematoda, Platyhelminthes, Portifera and Tardigrada*, further affirming that the in-utero environment is not sterile. *A* study that was recently published indicated that infants inherited an immunologically derived protein that protected them from Helminth infections; these proteins were found to be passed on from the mother and were inherited across generations in murine models (Darby *et al.*, 2019). This may explain the presence of these phyla derived peptides in the meconium.

The hierarchical cluster analysis in Figure 24 shows that there are a large number of classes, and interestingly the HEU and HUU samples do not cluster according to exposure state. This reiterates the interindividual microbiota of each infant as well as reaffirms the diversity within each sample. The number of classes identified are less than at day 4-7 in both groups, because meconium has a low biomass.

The results also show that there is a rapid shift in the microbial composition from birth (meconium) to Day 4-7 in both the HEU and HUU infant. This shift may be due to the HBM microbiome at day 4-7, as there is an increase in bacterial biomass identified as well as in increase in the number of species of bacteria in the infant stool samples for the HEU and HUU infant. In the HUU at day 4-7 the majority of the peptides identified belong to the order Bifidobacteriales , with an increase in the number of Bifidobacterium species, Bifidobacteria , from two species at birth to twelve species at D4-7: *B.adolescentis, B.asteroides, B.bifidum,*

B.bohemicum, B.breve, B.callitrichos, B.longum with subspecies B.longum infantis B.mongoliense, B.pseudocatenulatum., B.psychraaerophilum, B.huminantium and B.scardovii. The *B.scardovii* strain has been associated with recurrent urinary tract infections (Barberis *et al.*, 2012).

In the HEU infants group the majority of peptides identified also belonged to the genus Bifidobacterium with an increase from six species at birth to eleven species identified at D4-7: *B. bifidum, B. bohemicum, B. breve, B. callitrichos, B. catenulatum, B. commune, B.kashiwanohense, B. longum, B. psuedocatenulatum, B. psychraerophilum and B. huminatium*.

Of the Bifidobacterium species identified, both HEU and HUU share the presence of several species: *B.bifidum, B.bohemicum, B.breve, B.callitrichos, B.longum , B. psychraerophilum and B. huminatium* . Interestingly the subspecies *B.longum infantis* was only identified in the HUU infants , unlike other members of *B .longum* , the subspecies consumes a large number of Human milk oligosaccharides and is optimised for the infant gut microenvironment (Ward *et al.*, 2007)

The shift in the number of classes is probably due to the presence of HBM, as well as contact with the mother's areolar, skin and the post uterine environment. The heatmap shows that no one species is strongly associated with the microbiota. It reinforces the idea that the microbiome varies amongst individuals and that whilst these HEU infants have been shown to have adverse outcomes, there is no microbial signature as in the cases with other inflammatory diseases such as IBS or Crohn's disease. Any underlying differences at birth may be mitigated against by the introduction of HBM into the diet.

The abundance of Bifidobacterium in the gut of infants has been linked to the digestion and processing of short chain fatty acids (SCFA). HBM has been shown to harbour a large amount of SCFA and thus Bifidobacterium presence in the infant gut is expected (Quigley *et al.*, 2013).

A likely source of the Bacterial biomass in the stools at day 4-7 is HBM. HBM was the only source of nutrition for these infants as they were exclusively breast fed. Whilst HBM is believed to have a low biomass , it is established that HBM is not sterile (Soto *et al.*, 2014; Kumar *et al.*, 2016) and that it would be the primary source of bacteria for the infant (Heikkilä and Saris, 2003). The presence of Bifidobacteria in HBM has

previously been established by Soto and colleagues (Turroni *et al.*, 2012; Jost *et al.*, 2014; Soto *et al.*, 2014) Recently HBM microbiota was characterised and was found to be largely comprised of *Acinetobacter*, *Enterobacter Staphylococcus and Streptococcus* (Kordy *et al.*, 2019). Various studies have suggested that *B. breve and B.longum* are transferred directly from HBM into the infant's gut (Solís *et al.*, 2010; Soto *et al.*, 2014). Therefore, HBM may explain the presence of both species in the HEU and HUU infant.

However, other studies have found that the microbial communities present in the mother's breast milk and matched infants are different in composition (Pannaraj *et al.*, 2017). Regardless it is likely the composition of the infant gut microbiome may be influenced by a combination of the mother's areolar skin and breast milk (Pannaraj *et al.*, 2017).

Irrespective of the differences in the stool microbiota identification, the identification of Bifidobacteria in the infant's stool samples was expected, as Bifidobacterium were first isolated from HBM fed infant's stool in 1899 by Tissier (Weiss and Rettger, no date). The two species *B.bifidum and B.longum* have been identified as the most abundant species in infants (Turroni *et al.*, 2012).

Some of the differences in the HEU microbiome could be associated with the mother's HIV status (Bender *et al.*, 2016).) The average CD4 count for mothers in the HEU D4-7 samples was higher than the CD4 count in the HEU birth samples. Whilst most(47%) Appendix 3 of the HIV positive mothers were on the same HAARt regiment, 26% of the Mothers HAARt treatment was unknown, thus the variation may be due to the HAARt regimen as well as affected by how long the mother has been on ARV treatment. Biological confounders such as gender have been shown to influence the gut microbiome in infants(Cong *et al.*, 2016), and thus the number of males and females within in group at each timepoint may influence the microbial composition.

It also important to acknowledge that there is high interindividual variability within infants (Falony *et al.*, 2016), even within the same clinical group classification. The infant microbiome is known to be very unstable and thus is modulated frequently.

Whilst there may be a difference in the abundance of certain taxa that are represented in this study, it needs to be determined whether these differences might explain the difference in the altered immune function

that was previously described by (Garcia-Knight *et al.*, 2015; Smith *et al.*, 2017). Interestingly the introduction of HBM shifts the microbiome of both the HEU and HUU infant, making the difference in the diversity and abundance of the bacterial classes less prominent.

4.2 Functional analysis of Metaproteomic data

4.2.1 Peptide analysis.

The UniPept online web-based platform has a Quick go plugin that allows the viewing of the pathways that are associated with the peptides that have been identified. The platform allows the navigation over the sunburst charts and generating pathway information related to: Biological process, Molecular Function and Cellular Component. The chart and tables can be viewed at different hierarchical levels.

For this study, the hierarchical level interrogated was at Chordata level with investigation of the Go Terms associated with biological processes, the purpose being an investigation into the host proteins.

At birth most of the Chordata peptides that were identified in the HEU were assigned to the GO term associated with neutrophil degranulation (GO:0043312), followed by innate immune response (GO:0045087) and immune responses (GO:0006955) (Table 8). The top ten GO terms that are identified in the HEU infant are primarily concerned with the immune system, with six out of the ten identified being linked to immune function. Whereas in the HUU infant only four of the top ten are involved in the immune system responses / activation. This may provide insight into the HEU infants immune responses if the immune activation and activity has already started in the womb (Kidzeru *et al.*, 2014). In the HUU infant the top Go terms were neutrophil degranulation (GO:0043312), Proteolysis (GO:0006508), and Innate immune response (GO:0045087) (Table 9).

Interestingly the top GO terms in the HEU infant at day 4-7 remain associated with immune system activity, with eight of the GO terms out of the top ten being dedicated to the immune system response (Table 10), whereas in the HUU the top biological processes are involved in metabolism activity, with no Go term

associated with the immune response (Table 11) . This finding is of interest as it lends evidence to the theory that the dysbiosis in the HEU infant alters the immune development (Kidzeru *et al.*, 2014; Smith *et al.*, 2017). This finding supports those of other studies and may elucidate the altered immune response of HEU even if they remain uninfected (Kidzeru *et al.*, 2014; Smith *et al.*, 2017). The findings may also explain the lower birthweights of the HEU infants.

4.2.2 Protein analysis of metaproteomic data

Panther protein classes

Further pathway analysis was carried out on the Pantherid, using the protein-level identifications. The protein analysis on Panther dB was used to analyze the protein classes present in all groups at both time points. The most abundant protein class in both groups at both time points was hydrolase (PC000121). Hydrolase is a protein complex that is often found in the Liver and the Kidneys. Whilst the most abundant class was Hydrolase, the protein class associated with defence /immunity /(PC000090) was present in the HEU and HUU infant at both time points, this indicates that immune function is an integral part of the infant's microbiome.

Perseus protein Analysis

The protein analysis revealed that whilst there aren't any significantly expressed proteins in either at each time point after multiple testing correction, there are proteins that are exclusively found in ether one of the groups. The use of multiple testing correction in proteomics is somewhat contentious since it can be overly stringent and has been shown to exclude many true identifications in *in silico* mock studies. However, regardless of that, the proteins that are exclusive to each group at each time point may facilitate an understanding of the microbiome differences.

Further functional analysis of the Human proteins exclusively found in each group at each time point was carried out on the String Db online platform. In HEU infants the most abundant GO Term descriptions were

associated with immune function, with neutrophil degranulation (GO: 0043312) and humoral immune response (GO:000695) in birth and Day 4-7 respectively. However, String. dB was only able to identify three GO terms in the Birth HEU and seven GO terms in the Day 4-7 HUU group.

Both the groups experience an increase in the number of genes that were identified, however unlike the QuickGO analysis on UniPept most of the proteins are not associated with immune response. Instead in all cases most of the proteins were associated with Cellular process.

This difference in the biological process data highlights one the shortcomings of microbiome studies. The platforms that are used to analyze the data can generate very different results for the same set of data. This phenomenon was recently discussed by Peters and colleagues (Peters *et al.*, 2019a), where they found that the same data analysed on UniPept and MEGAN gave different results. One of the reasons these differences are observed in our data is due to the nature of the platforms. The differences observed in the biological process analysis on the UniPept platform vs Pantherdb.org is due to the difference in the searches. The UniPept functions at a peptide-centric level, assigning peptides to the lowest common ancestor (LCA) for identification, whilst the analysis on Panther dB relies on protein level identification with all non-human proteins filtered out. This may have resulted in the loss of some human protein that were erroneously attributed to non-human primates in MetaNovo.

Both methods have their advantages and disadvantages. The advantage in using Pantherdb is the platform's ability to assign proteins to Gene Ontologies as well as to match the proteins to genes. This allows for a more accurate inference of function in the human host. Another advantage of Pantherdb is that the database is curated with genomic data and has been optimized for gene-function association. Whilst the QuickGo plugin is useful in that functional analysis and taxonomic analysis can be carried out on a single platform, the issue with peptide-centric analysis is the possibility of a single peptide sequence belonging to more than one protein. This makes it challenging to accurately assign peptides to functions.

4.3 Limitations

Whilst metaproteomics allows us to study the interaction between the host proteins as well as the microbiota, there are still limitations. One of the greatest limitations is the lack of a standardised protocol for the stool sample preparation and analysis (Lai *et al.*, 2019; Peters *et al.*, 2019a). The methods used for samples preparation has a great bearing on the microbiota identified.

Another limitation is the masking of low abundance proteins in mass spectrometry. Low abundance proteins are often not detected and can be missed with further downstream analysis. Thus, clinically relevant microbial and host interplay may be missed due to this.

Furthermore the reliance of a well curate database to identify the microbiota and the host proteins is a limitation, because it is known that upon downstream analysis two very distinct results can be generated using the exact same samples and files, as was witnessed by (Hamady *et al.*, 2009; Peters *et al.*, 2019b).

There is also batch to batch variation that must be accounted for due to human error and that may affect the quality control.

The database MetaNovo that was used erroneously identified human derived proteins as non-human primates i.e. macaque, and this needs to be taken into consideration as this suggests that there is room for error in the identification of the microbial peptides. Whilst UniPept can identify to LCA (lowest common ancestor) as well as species, the safer method to analyse the data was to look at class of bacteria and phyla instead of species as this would increase the error rate. Taking this into account pathogenically relevant species may be missed upon investigation. One way to overcome the possible misidentification is to manually verify the spectra, to ensure what is identified actually exist within the sample.

A major limitation of the study itself was the number of samples available for analysis a total of 63 samples were analysed as part of this sub study. The limited number of samples means there was limited statistical power. Another limitation that must be accounted for which also relates to the curation of the database, is the number of unassigned peptide sequences, which is one of the disadvantages of using a peptide centric

downstream analysis tool. However, it does allow for the identification of unique microbes that were previously difficult to identify in microbiome studies.

Chapter 5. Conclusion and recommendations

From the above study there are underlying differences in the microbiome of HEU and HUU infants. These differences are most evident at birth suggesting a great influence in the womb. Whilst these differences can be seen at birth it is still not clear what the underlying mechanism for the differences is. It could be a host of factors, such as the mother's HIV status, the presence of ARV's and the vaginal microbiome of the mother being amongst a few of the factors that may explain these differences. Interestingly the microbiome differences in the HUU and HEU groups was less starkly different at day 4-7 indicating the important role of HBM in shaping the infant's microbiome. This finding suggests that the presence of HBM could facilitate correcting the underlying differences in the two groups.

A future study comparing the mother's HBM and vaginal microbiome to the infant's gut microbiome, as well as a longitudinal study would be beneficial in understanding the differences seen in these two groups. Also, a larger number of infant mothers diads would be able to generate a statistically powerful study, however the clinical relevance of this study is still realized. There is a trend towards microbiomes colonised by different microbiota. The amniotic fluid, placenta and vaginal microbiome of infant mother pairs would allow us to investigate the origin of the in-uterine environment. Examining the HBM and the infant stool microbiome in a longitudinal study would also generate data on colonisation and changes in the microbiome in the HEU infant. The microbiota could also be investigated and quantified by use of spectral counting to reveal any quantitative differences between the groups investigated.

A metagenomic approach would also be beneficial to understanding the host and microbe interactions. it would also allow for the comparison of the genetic data and the proteomic data. The metabolomic fractions from the sample prep should be analysed as well, as this would allow for a comparison of the protein pathway data generated from the proteome with the data generated by the metabolites. A metabolomics study of the HBM of the mothers in the study would allow for a comparative analysis of the stool metabolome of both the infant and the mother, as well as the HBM metabolome.

Reference list

Aagaard, K. *et al.* (2014) 'The placenta harbors a unique microbiome.', *Science translational medicine*. American Association for the Advancement of Science, 6(237), p. 237ra65. doi: 10.1126/scitranslmed.3008599.

Adler, C. *et al.* (2015) 'Severe Infections in HIV-Exposed Uninfected Infants Born in a European Country.', *PloS one*. Public Library of Science, 10(8), p. e0135375. doi: 10.1371/journal.pone.0135375.

Aguilar, M.-I. (2004) 'HPLC of Peptides and Proteins: Basic Theory and Methodology', in *HPLC of Peptides and Proteins*. New Jersey: Humana Press, pp. 3–8. doi: 10.1385/1-59259-742-4:3.

Aids, E. C. for the E. M. of (2018) 'HIV/AIDS surveillance in Europe 1994-1996', p. 82. doi: 10.2900/523723.

Arikawa, S. *et al.* (2016) 'Mortality risk and associated factors in HIV-exposed, uninfected children', *Tropical Medicine and International Health*, 21(6), pp. 720–734. doi: 10.1111/tmi.12695.

Armanian, A.-M. *et al.* (2016) 'The effect of neutral oligosaccharides on fecal microbiota in premature infants fed exclusively with breast milk: A randomized clinical trial.', *Journal of research in pharmacy practice*, 5(1), pp. 27–34. doi: 10.4103/2279-042X.176558.

Barberis, C. M. *et al.* (2012) 'Recurrent urinary infection with Bifidobacterium scardovii', *Journal of Clinical Microbiology*, 50(3), pp. 1086–1088. doi: 10.1128/JCM.06027-11.

Bender, J. M. *et al.* (2016) 'Maternal HIV infection influences the microbiome of HIV-uninfected infants', *Science Translational Medicine*, 8(349), pp. 349ra100-349ra100. doi: 10.1126/scitranslmed.aaf5103.

Burton, R., Giddy, J. and Stinson, K. (2015) 'Prevention of mother-to-child transmission in South Africa: an ever-changing landscape.', *Obstetric medicine*. SAGE Publications, 8(1), pp. 5–12. doi: 10.1177/1753495X15570994.

Caporaso, J. G. *et al.* (2011) 'Moving pictures of the human microbiome', *Genome Biology*. BioMed Central, 12(5), p. R50. doi: 10.1186/gb-2011-12-5-r50.

CDC (2016) *CROI Press Release: HIV Risk | 2016 | Newsroom | NCHHSTP | CDC*, *Website*. Available at:

https://www.cdc.gov/nchhstp/newsroom/2016/croi-press-release-risk.html (Accessed: 10 February 2019).

Cerea, S. *et al.* (2015) 'Valutazione di aspetti psicologici in un gruppo di studenti universitari con dislessia', *Psicoterapia Cognitiva e Comportamentale*. John Wiley & Sons, Ltd (10.1111), 21(1), pp. 49–72. doi: 10.1113/jphysiol.2009.174136.

Clark, A. E. *et al.* (2013) 'Matrix-Assisted laser desorption ionization-time of flight mass spectrometry: A fundamental shift in the routine practice of clinical microbiology', *Clinical Microbiology Reviews*, 26(3), pp. 547–603. doi: 10.1128/CMR.00072-12.

Collado, M. C. *et al.* (2016) 'Human gut colonisation may be initiated in utero by distinct microbial communities in the placenta and amniotic fluid', *Scientific Reports*. Nature Publishing Group, 6(1), p. 23129. doi: 10.1038/srep23129.

Cong, X. *et al.* (2016) 'Gut Microbiome Developmental Patterns in Early Life of Preterm Infants: Impacts of Feeding and Gender', *PLOS ONE*. Edited by F. Cappello. Public Library of Science, 11(4), p. e0152751. doi: 10.1371/journal.pone.0152751.

Consortium, T. H. M. P. *et al.* (2012) 'Structure, function and diversity of the healthy human microbiome', *Nature*. Nature Publishing Group, 486(7402), pp. 207–214. doi: 10.1038/nature11234.

Costello, E. K. *et al.* (2009) 'Bacterial community variation in human body habitats across space and time.', *Science (New York, N.Y.)*. NIH Public Access, 326(5960), pp. 1694–7. doi: 10.1126/science.1177486.

Coutsoudis, A. (2005) 'Breastfeeding and the HIV positive mother: The debate continues', *Early Human Development*, 81(1 SPEC. ISS.), pp. 87–93. doi: 10.1016/j.earlhumdev.2004.10.010.

Darby, M. G. *et al.* (2019) 'Pre-conception maternal helminth infection transfers via nursing long-lasting cellular immunity against helminths to offspring', *Science Advances*. American Association for the Advancement of Science, 5(5), p. eaav3058. doi: 10.1126/sciadv.aav3058.

Dewhirst, F. E. *et al.* (2010) 'The human oral microbiome.', *Journal of bacteriology*. American Society for Microbiology (ASM), 192(19), pp. 5002–17. doi: 10.1128/JB.00542-10.

Dietert, R. R. (2017) 'The microbiome-immune-host defense barrier complex (microimmunosome) and developmental programming of noncommunicable diseases', *Reproductive Toxicology*, 68, pp. 49–58. doi: 10.1016/j.reprotox.2016.04.026.

Dingle, T. C. and Butler-Wu, S. M. (2013) 'MALDI-TOF Mass Spectrometry for Microorganism Identification', *Clinics in Laboratory Medicine*, 33(3), pp. 589–609. doi: 10.1016/j.cll.2013.03.001.

Doherty, T. *et al.* (2010) 'HIV and Infant Feeding in South Africa Implications of the New WHO Guidelines on HIV and Infant Feeding for Child Survival in South Africa', *Bulletin of the World Health Organization.* World Health Organization, 89(1), pp. 62–67. doi: 10.2471/BLT.10.079798.

Doherty, T. *et al.* (2011) 'Implications des nouvelles lignes directrices de l'OMS relatives au VIH et à la nutrition infantile sur la survie des enfants en Afrique du Sud', *Bulletin of the World Health Organization*, 89(1), pp. 62–67. doi: 10.2471/BLT.10.079798.

Dominguez-Bello, M. G. *et al.* (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), pp. 11971–5. doi: 10.1073/pnas.1002601107.

Dominguez-Bello, M. G. *et al.* (2013) 'Association of Cesarean Delivery and Formula Supplementation With the Intestinal Microbiome of 6-Week-Old Infants.', *Pharmacological Research*. Elsevier B.V., 451(1), pp. 1–8. doi: 10.1016/j.phrs.2012.09.001.

Early mother-to-child transmission of HIV stats plunge | South African Medical Research Council (2016). Available at: http://www.mrc.ac.za/media-release/early-mother-child-transmission-hiv-stats-plunge (Accessed: 13 June 2018).

El-Sadr, W. M., Mayer, K. H. and Hodder, S. L. (2010) 'AIDS in America — Forgotten but Not Gone', *New England Journal of Medicine*. NIH Public Access, 362(11), pp. 967–970. doi: 10.1056/NEJMp1000069.

Di Falco, M. R. (2018) 'Mass spectrometry-based proteomics', in *Methods in Molecular Biology*, pp. 93–106. doi: 10.1007/978-1-4939-7804-5_9.

Falony, G. *et al.* (2016) 'Population-level analysis of gut microbiome variation.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 352(6285), pp. 560–4. doi: 10.1126/science.aad3503.

Fan, W. *et al.* (2013) 'Diversity of the intestinal microbiota in different patterns of feeding infants by Illumina high-throughput sequencing.', *World journal of microbiology & biotechnology*, 29(12), pp. 2365–72. doi: 10.1007/s11274-013-1404-3.

Fettweis, J. M. *et al.* (2014) 'Differences in vaginal microbiome in African American women versus women of European ancestry', *Microbiology (United Kingdom)*. Microbiology Society, 160(10), pp. 2272–2282. doi: 10.1099/mic.0.081034-0.

De Filippo, C. *et al.* (2010) 'Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa', *PNAS*, 107(33), pp. 14691–14696. doi: 10.1073/pnas.1005963107.

Fournier, P. E. *et al.* (2015) 'From culturomics to taxonomogenomics: A need to change the taxonomy of prokaryotes in clinical microbiology', *Anaerobe*. Elsevier Ltd, 36, pp. 73–78. doi: 10.1016/j.anaerobe.2015.10.011.

Fredricks, D. N. (2001) 'Microbial ecology of human skin in health and disease.', *The journal of investigative dermatology. Symposium proceedings*, 6(3), pp. 167–9. doi: 10.1046/j.0022-202x.2001.00039.x.

Fung, I. *et al.* (2012) 'Do Bugs Control Our Fate? The Influence of the Microbiome on Autoimmunity', *Current Allergy and Asthma Reports*. Current Science Inc., 12(6), pp. 511–519. doi: 10.1007/s11882-012-0291-2.

Gao, Z. *et al.* (2007) 'Molecular analysis of human forearm superficial skin bacterial biota.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 104(8), pp. 2927–32. doi: 10.1073/pnas.0607077104.

Garcia-Knight, M. A. *et al.* (2015) 'Altered memory T-cell responses to Bacillus Calmette-Guerin and Tetanus Toxoid vaccination and altered cytokine responses to polyclonal stimulation in HIV-exposed uninfected Kenyan infants', *PLoS ONE*, 10(11). doi: 10.1371/journal.pone.0143043.

Gevers, D. *et al.* (2014) 'The treatment-naive microbiome in new-onset Crohn's disease.', *Cell host & microbe*. NIH Public Access, 15(3), pp. 382–392. doi: 10.1016/j.chom.2014.02.005.

Ghanbari, M., Kneifel, W. and Domig, K. J. (2015) 'A new view of the fish gut microbiome: Advances from next-generation sequencing', *Aquaculture*. Elsevier B.V., 448, pp. 464–475. doi: 10.1016/j.aquaculture.2015.06.033.

Gibb, D. M. *et al.* (2012) 'Pregnancy and infant outcomes among HIV-infected women taking long-term art with and without tenofovir in the DART trial', *PLoS Medicine*, 9(5). doi: 10.1371/journal.pmed.1001217.

Goga, A. *et al.* (2017) 'S A H R 2 0 t h E d i t i o n', *SAHR*, (20 Year Anniversary edition), pp. 137–146.

Greub, G. (2012a) 'Culturomics : a new approach to study the human microbiome', *Clinical Microbiology and Infection*. European Society of Clinical Infectious Diseases, 18(12), pp. 1157–1159. doi: 10.1111/1469-0691.12032.

Greub, G. (2012b) 'Culturomics: a new approach to study the human microbiome', *Clinical Microbiology and Infection*, 18, pp. 1157–1159. doi: 10.1111/1469-0691.12032.

Grice, E. A. *et al.* (2009) 'Topographical and temporal diversity of the human skin microbiome', *Science*. American Association for the Advancement of Science, 324(5931), pp. 1190–1192. doi: 10.1126/science.1171700.

Griffiths, W. J. and Wang, Y. (2009) 'Mass spectrometry: From proteomics to metabolomics and lipidomics', *Chemical Society Reviews*. Royal Society of Chemistry, 38(7), pp. 1882–1896. doi: 10.1039/b618553n.

Gundry, R. L. *et al.* (2009) 'Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow', *Current Protocols in Molecular Biology*. NIH Public Access, Chapter 10(SUPPL. 88), p. Unit10.25. doi: 10.1002/0471142727.mb1025s88.

Gundry, R. L. *et al.* (2010) 'Preparation of Protein and Petides for Mass Spectrometry Analysis in a Bottom-Up Proteomics Workflow', *October*, 77(5), pp. 342–355. doi: 10.1002/0471142727.mb1025s88.Preparation.

Hamady, M. *et al.* (2009) 'Tools , techniques , and challenges Microbial community profiling for human

microbiome projects : Tools , techniques , and challenges', (303), pp. 1141–1152. doi: 10.1101/gr.085464.108.

Hargreaves, J. R. *et al.* (2007) *Explaining continued high HIV prevalence in South Africa: socioeconomic factors, HIV incidence and sexual behaviour change among a rural cohort, 2001-2004*. Wolters Kluwer Health | Lippincott Williams & Wilkins AIDS.

Heijtz, R. D. (2016) 'Seminars in Fetal & Neonatal Medicine Fetal , neonatal , and infant microbiome : Perturbations and subsequent effects on brain development and behavior', *Seminars in Fetal and Neonatal Medicine*. Elsevier Ltd, 46, pp. 4–11. doi: 10.1016/j.siny.2016.04.012.

Heikkilä, M. P. and Saris, P. E. J. (2003) 'Inhibition of Staphylococcus aureus by the commensal bacteria of human milk.', *Journal of applied microbiology*, 95(3), pp. 471–8.

Hesseling, A. C. *et al.* (2007) 'The risk of disseminated Bacille Calmette-Guerin (BCG) disease in HIV-infected children', *Vaccine*, 25, pp. 14–18. doi: 10.1016/j.vaccine.2006.07.020.

Hesseling, A. C. *et al.* (2009) 'BCG vaccination in South African HIV-exposed infants--risks and benefits.', *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde*. NIH Public Access, 99(2), pp. 88–91.

HIV and infant feeding Guidelines on (2010).

Hofer, C. B. *et al.* (2016) 'In Utero Exposure to Antiretroviral Drugs: Effect on Birth Weight and Growth Among HIV-exposed Uninfected Children in Brazil.', *The Pediatric infectious disease journal*. NIH Public Access, 35(1), pp. 71–7. doi: 10.1097/INF.0000000000000926.

Hoffmann, E. De and Stroobant, V. (2007) 'Mass spectrometry principles and Applications', in *Mass Spectrometry - Priniples and Applications*, pp. 43–46. doi: 10.1002/mas.20296.

Le Huërou-Luron, I., Blat, S. and Boudry, G. (2010) 'Breast- v. formula-feeding: impacts on the digestive tract and immediate and long-term health effects.', *Nutrition research reviews*, 23(1), pp. 23–36. doi: 10.1017/S0954422410000065.

Huttenhower, C. *et al.* (2012) 'Structure, function and diversity of the healthy human microbiome', *Nature*, 486(7402), pp. 207–214. doi: 10.1038/nature11234.

Jeurink, P. V. *et al.* (2013) 'Human milk: A source of more life than we imagine', *Beneficial Microbes*, 4(1), pp. 17–30. doi: 10.3920/BM2012.0040.

Jiménez, E. *et al.* (2008) 'Is meconium from healthy newborns actually sterile?', *Research in Microbiology*. Elsevier Masson, 159(3), pp. 187–193. doi: 10.1016/J.RESMIC.2007.12.007.

Jost, T. *et al.* (2013) 'Assessment of bacterial diversity in breast milk using culture-dependent and culture-independent approaches.', *The British journal of nutrition*, 110(7), pp. 1253–62. doi: 10.1017/S0007114513000597.

Jost, T. *et al.* (2014) 'Vertical mother – neonate transfer of maternal gut bacteria via breastfeeding', 16, pp. 2891–2904. doi: 10.1111/1462-2920.12238.

Kidzeru, E. B. *et al.* (2014) 'In-utero exposure to maternal HIV infection alters T-cell immune responses to vaccination in HIV-uninfected infants', *AIDS*, 28(10). doi: 10.1097/QAD.0000000000000292.

Kinross, J. M., Darzi, A. W. and Nicholson, J. K. (2011) 'Gut microbiome-host interactions in health and disease', *Genome Medicine*. BioMed Central, 3(3), p. 14. doi: 10.1186/gm228.

Kirk, R. G. (2012) '"Life in a germ-free world": isolating life from the laboratory animal to the bubble boy.', *Bulletin of the history of medicine*. Various, 86(2), pp. 237–75. doi: 10.1353/bhm.2012.0028.

Kong, H. H. *et al.* (2012) 'Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis.', *Genome research*. Cold Spring Harbor Laboratory Press, 22(5), pp. 850–9. doi: 10.1101/gr.131029.111.

Kordy, K. *et al.* (2019) 'Contributions to human breast milk microbiome and enteromammary transfer of Bifidobacterium breve', *bioRxiv*. Cold Spring Harbor Laboratory, p. 688754. doi: 10.1101/688754.

Kuhn, L. *et al.* (2015) 'Oligosaccharide Composition of Breast Milk Influences Survival of Uninfected Children Born to HIV-Infected Mothers in Lusaka, Zambia', *The Journal of Nutrition*, 145(1), pp. 66–72. doi:

10.3945/jn.114.199794.

Kumar, H. *et al.* (2016) 'Distinct Patterns in Human Milk Microbiota and Fatty Acid Profiles Across Specific Geographic Locations', *Frontiers in Microbiology*. Frontiers, 7, p. 1619. doi: 10.3389/fmicb.2016.01619.

Lagier, J.-C. *et al.* (2012) 'Microbial culturomics: paradigm shift in the human gut microbiome study', *Clinical Microbiology and Infection*. Elsevier, 18(12), pp. 1185–1193. doi: 10.1111/1469-0691.12023.

Lagier, J. C. *et al.* (2015) 'The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota', *Clinical Microbiology Reviews*, 28(1), pp. 237–264. doi: 10.1128/CMR.00014-14.

Lai, L. A. *et al.* (2019) 'Metaproteomics Study of the Gut Microbiome', in. Humana Press, New York, NY, pp. 123–132. doi: 10.1007/978-1-4939-8814-3_8.

Larsen, N. *et al.* (2010) 'Gut Microbiota in Human Adults with Type 2 Diabetes Differs from Non-Diabetic Adults', *PLoS ONE*. Edited by S. Bereswill. Public Library of Science, 5(2), p. e9085. doi: 10.1371/journal.pone.0009085.

Lederberg, B. J. and McCray, A. T. (2001) '' Ome Sweet ' Omics-- A Genealogical Treasury of Words', *The Scientist*, 2 April, p. 8.

Leiby, J. S. *et al.* (2018) 'Lack of detection of a human placenta microbiome in samples from preterm and term deliveries', *Microbiome*. BioMed Central, 6(1), p. 196. doi: 10.1186/s40168-018-0575-4.

Leon, L. J. *et al.* (2018) 'Enrichment of Clinically Relevant Organisms in Spontaneous Preterm-Delivered Placentas and Reagent Contamination across All Clinical Groups in a Large Pregnancy Cohort in the United Kingdom.', *Applied and environmental microbiology*. American Society for Microbiology, 84(14), pp. e00483-18. doi: 10.1128/AEM.00483-18.

Li, S. *et al.* (2016) 'Complexities of Gut Microbiome Dysbiosis in the Context of HIV Infection and Antiretroviral Therapy', *Clinical Pharmacology & Therapeutics*. John Wiley & Sons, Ltd, 99(6), pp. 600–611. doi: 10.1002/cpt.363.

Lim, E. S., Rodriguez, C. and Holtz, L. R. (2018) 'Amniotic fluid from healthy term pregnancies does not

harbor a detectable microbial community', *Microbiome*. BioMed Central, 6(1), p. 87. doi: 10.1186/s40168-018-0475-7.

Ling, X. Z. *et al.* (2010) 'Molecular analysis of the diversity of vaginal microbiota associated with bacterial vaginosis', *BMC Genomics*. BioMed Central, 11(1), p. 488. doi: 10.1186/1471-2164-11-488.

Loewenson, R. *et al.* (1997) *SOCIAL AND ECONOMIC ISSUES OF HIV/AIDS IN SOUTHERN AFRICA A Consultancy Report Prepared for SAfAIDS, Harare*.

Ma, B., Forney, L. J. and Ravel, J. (2012) 'Vaginal Microbiome: Rethinking Health and Disease', *Annual Review of Microbiology*, 66(1), pp. 371–389. doi: 10.1146/annurev-micro-092611-150157.

Magne, F. *et al.* (2017) 'The Elevated Rate of Cesarean Section and Its Contribution to Non-Communicable Chronic Diseases in Latin America: The Growing Involvement of the Microbiota', *Frontiers in Pediatrics*, 5, p. 192. doi: 10.3389/fped.2017.00192.

Mandal, M. K. *et al.* (2013) 'Biomolecular analysis and cancer diagnostics by negative mode probe electrospray ionization', *Analyst*. Royal Society of Chemistry, 138(6), pp. 1682–1688. doi: 10.1039/c3an36554a.

McLafferty, F. W. *et al.* (2007) 'Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics', *FEBS Journal*. John Wiley & Sons, Ltd (10.1111), 274(24), pp. 6256–6268. doi: 10.1111/j.1742-4658.2007.06147.x.

Morgan, X. C. *et al.* (2012a) 'Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment.', *Genome biology*. BioMed Central, 13(9), p. R79. doi: 10.1186/gb-2012-13-9-r79.

Morgan, X. C. *et al.* (2012b) 'Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment', *Genome Biology*, 13(9), p. R79. doi: 10.1186/gb-2012-13-9-r79.

Morgan, X. C., Segata, N. and Huttenhower, C. (2013) 'Biodiversity and functional genomics in the human microbiome', *Trends in Genetics*. Elsevier Current Trends, pp. 51–58. doi: 10.1016/j.tig.2012.09.005.

Mueller, N. T. *et al.* (2015) 'The infant microbiome development: Mom matters', *Trends in Molecular*

Medicine. doi: 10.1016/j.molmed.2014.12.002.

Natchu, U. C. M. *et al.* (2012) 'Exclusive breastfeeding reduces risk of mortality in infants up to 6 mo of age born to HIV-positive Tanzanian women.', *The American journal of clinical nutrition*, 96(5), pp. 1071–8. doi: 10.3945/ajcn.111.024356.

Nazmul Huda, M. *et al.* (2014) *Stool Microbiota and Vaccine Responses of Infants:e362-e372*, *Pediatrics*.

NIH HMP Working Group, T. N. H. W. *et al.* (2009) 'The NIH Human Microbiome Project.', *Genome research*. Cold Spring Harbor Laboratory Press, 19(12), pp. 2317–23. doi: 10.1101/gr.096651.109.

Pallen, M. J. (2011) 'The Human Microbiome and Host–Pathogen Interactions', in *Metagenomics of the Human Body*. New York, NY: Springer New York, pp. 43–61. doi: 10.1007/978-1-4419-7089-3_3.

Pammi, M. *et al.* (2017) 'Intestinal dysbiosis in preterm infants preceding necrotizing enterocolitis: a systematic review and meta-analysis', *Microbiome*. BioMed Central, 5(1), p. 31. doi: 10.1186/s40168-017-0248-8.

Pannaraj, P. S. *et al.* (2017) 'Association Between Breast Milk Bacterial Communities and Establishment and Development of the Infant Gut Microbiome', *JAMA Pediatrics*. American Medical Association, 171(7), p. 647. doi: 10.1001/jamapediatrics.2017.0378.

Paster, B. J. *et al.* (2006) 'The breadth of bacterial diversity in the human periodontal pocket and other oral sites', *Periodontology 2000*. John Wiley & Sons, Ltd (10.1111), 42(1), pp. 80–87. doi: 10.1111/j.1600-0757.2006.00174.x.

Patterson, S. D. and Aebersold, R. H. (2003) 'Proteomics: the first decade and beyond', *Nature Genetics*. Nature Publishing Group, 33(3s), pp. 311–323. doi: 10.1038/ng1106.

Peng, J. *et al.* (2003) 'Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: The yeast proteome', *Journal of Proteome Research*, 2(1), pp. 43–50. doi: 10.1021/pr025556v.

Perez-Muñoz, M. E. *et al.* (2017) 'A critical assessment of the "sterile womb" and "in utero colonization"

hypotheses: implications for research on the pioneer infant microbiome', *Microbiome*. BioMed Central, 5(1), p. 48. doi: 10.1186/s40168-017-0268-4.

Perez, P. F. *et al.* (2007) 'Bacterial Imprinting of the Neonatal Immune System: Lessons From Maternal Cells?', *PEDIATRICS*, 119(3), pp. e724–e732. doi: 10.1542/peds.2006-1649.

Peters, D. L. *et al.* (2019a) 'Metaproteomic and Metabolomic Approaches for Characterizing the Gut Microbiome', *PROTEOMICS*. John Wiley & Sons, Ltd, p. 1800363. doi: 10.1002/pmic.201800363.

Peters, D. L. *et al.* (2019b) 'Metaproteomic and Metabolomic Approaches for Characterizing the Gut Microbiome', *PROTEOMICS*. John Wiley & Sons, Ltd, p. 1800363. doi: 10.1002/pmic.201800363.

Pflughoeft, K. J. and Versalovic, J. (2012) 'Human Microbiome in Health and Disease', *Annual Review of Pathology: Mechanisms of Disease*. Annual Reviews , 7(1), pp. 99–122. doi: 10.1146/annurev-pathol-011811-132421.

Potgieter, M. G. *et al.* (2019) 'MetaNovo: a probabilistic approach to peptide and polymorphism discovery in complex mass spectrometry datasets', *bioRxiv*. Cold Spring Harbor Laboratory, p. 605550. doi: 10.1101/605550.

Powis, K. M. *et al.* (2011) 'Effects of in utero antiretroviral exposure on longitudinal growth of HIV-exposed uninfected infants in Botswana.', *Journal of acquired immune deficiency syndromes (1999)*. NIH Public Access, 56(2), pp. 131–8. doi: 10.1097/QAI.0b013e3181ffa4f5.

Qin, J. *et al.* (2010) 'A human gut microbial gene catalogue established by metagenomic sequencing', *Nature*, 464(7285), pp. 59–65. doi: 10.1038/nature08821.

Quigley, L. *et al.* (2013) 'The complex microbiota of raw milk', *FEMS Microbiology Reviews*, 37(5), pp. 664–698. doi: 10.1111/1574-6976.12030.

Ramokolo, V. *et al.* (2014) 'HIV infection, viral load, low birth weight, and nevirapine are independent influences on growth velocity in HIV-exposed South African infants.', *The Journal of nutrition*, 144(1), pp. 42–8. doi: 10.3945/jn.113.178616.

Ravel, Jacques, Gajer, P., Abdo, Z., Schneider, G. M., Koenig, S. S. K., McCulle, Stacey L, *et al.* (2011) 'Vaginal microbiome of reproductive-age women', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 108(Supplement_1), pp. 4680–4687. doi: 10.1073/pnas.1002611107.

Ravel, J. *et al.* (2011) 'Vaginal microbiome of reproductive-age women', *Proceedings of the National Academy of Sciences*, 108(Supplement_1), pp. 4680–4687. doi: 10.1073/pnas.1002611107.

Ravel, Jacques, Gajer, P., Abdo, Z., Schneider, G. M., Koenig, S. S. K., McCulle, S. L., *et al.* (2011) 'Vaginal microbiome of reproductive-age women', *Proceedings of the National Academy of Sciences*, 108(Supplement_1), pp. 4680–4687. doi: 10.1073/pnas.1002611107.

Rodríguez, J. M. (2014) 'The origin of human milk bacteria: is there a bacterial entero-mammary pathway during late pregnancy and lactation?', *Advances in nutrition (Bethesda, Md.)*. American Society for Nutrition, 5(6), pp. 779–84. doi: 10.3945/an.114.007229.

Round, J. L. and Mazmanian, S. K. (2009) 'The gut microbiota shapes intestinal immune responses during health and disease.', *Nature reviews. Immunology*. NIH Public Access, 9(5), pp. 313–23. doi: 10.1038/nri2515.

le Roux, S. M. *et al.* (2016) 'Clinical outcomes of HIV-exposed, HIV-uninfected children in sub-Saharan Africa', *Tropical Medicine & International Health*. Wiley/Blackwell (10.1111), 21(7), pp. 829–845. doi: 10.1111/tmi.12716.

Salter, S. J. *et al.* (2014) 'Reagent and laboratory contamination can critically impact sequence-based microbiome analyses', *BMC Biology*. BioMed Central Ltd., 12(1). doi: 10.1186/s12915-014-0087-z.

Sheehan, D. and Shanahan, F. (2017) 'The Gut Microbiota in Inflammatory Bowel Disease', *Gastroenterology Clinics of North America*. Elsevier Inc., 46(1), pp. 143–154. doi: 10.1016/j.gtc.2016.09.011.

Sherman, G. G. *et al.* (2017) 'Toward elimination of mother-to-child transmission of HIV in South Africa: how best to monitor early infant infections within the Prevention of Mother-to-Child Transmission Program.', *Journal of global health*. Edinburgh University Global Health Society, 7(1), p. 010701. doi:

10.7189/jogh.07.010701.

Singhal, N. *et al.* (2015) 'MALDI-TOF mass spectrometry: An emerging technology for microbial identification and diagnosis', *Frontiers in Microbiology*, 6(AUG), pp. 1–16. doi: 10.3389/fmicb.2015.00791.

Slingerland, A. E. *et al.* (2017) 'Clinical Evidence for the Microbiome in Inflammatory Diseases.', *Frontiers in immunology*. Frontiers Media SA, 8, p. 400. doi: 10.3389/fimmu.2017.00400.

Slogrove, A. *et al.* (2012) 'HIV-exposed uninfected infants are at increased risk for severe infections in the first year of life', *Journal of Tropical Pediatrics*, 58(6), pp. 505–508. doi: 10.1093/tropej/fms019.

Slogrove, A. L. *et al.* (2016) 'Pattern of Infectious Morbidity in HIV-Exposed Uninfected Infants and Children', *Frontiers in Immunology*. Frontiers, 7, p. 164. doi: 10.3389/fimmu.2016.00164.

Slogrove, A. L., Archary, M. and Cotton, M. F. (2016) 'Optimizing Research Methods to Understand HIV-Exposed Uninfected Infant and Child Morbidity: Report of the Second HEU Infant and Child Workshop', *Frontiers in Immunology*. Frontiers, 7, p. 576. doi: 10.3389/FIMMU.2016.00576.

Smith, C. *et al.* (2017) 'Altered Natural Killer Cell Function in HIV-Exposed Uninfected Infants', *Frontiers in Immunology*, 8(April), p. 470. doi: 10.3389/fimmu.2017.00470.

Solís, G. *et al.* (2010) 'Establishment and development of lactic acid bacteria and bifidobacteria microbiota in breast-milk and the infant gut', *Anaerobe*. Elsevier Ltd, 16(3), pp. 307–310. doi: 10.1016/j.anaerobe.2010.02.004.

Song, S. J., Dominguez-Bello, M. G. and Knight, R. (2013) 'How delivery mode and feeding can shape the bacterial community in the infant gut', *Cmaj*, 185(5), pp. 373–374. doi: 10.1503/cmaj.130147.

Soto, A. *et al.* (2014) 'Lactobacilli and Bifidobacteria in Human Breast Milk', *Journal of Pediatric Gastroenterology and Nutrition*, 59(1), pp. 78–88. doi: 10.1097/MPG.0000000000000347.

TeKippe, E. M. *et al.* (2013) 'Optimizing identification of clinically relevant gram-positive organisms by use of the bruker biotyper matrix-assisted laser desorption ionization-time of flight mass spectrometry system', *Journal of Clinical Microbiology*, 51(5), pp. 1421–1427. doi: 10.1128/JCM.02680-12.

Theis, K. R. *et al.* (2019) 'Does the human placenta delivered at term have a microbiota? Results of cultivation, quantitative real-time PCR, 16S rRNA gene sequencing, and metagenomics', *American Journal of Obstetrics and Gynecology*. Mosby, 220(3), pp. 267.e1-267.e39. doi: 10.1016/J.AJOG.2018.10.018.

Thelen, J. J. and Peck, S. C. (2007) 'Quantitative Proteomics in Plants: Choices in Abundance', *THE PLANT CELL ONLINE*. American Society of Plant Biologists, 19(11), pp. 3339–3346. doi: 10.1105/tpc.107.053991.

Turnbaugh, P. J. *et al.* (2006a) 'An obesity-associated gut microbiome with increased capacity for energy harvest', *Nature*, 444(7122), pp. 1027–1031. doi: 10.1038/nature05414.

Turnbaugh, P. J. *et al.* (2006b) 'An obesity-associated gut microbiome with increased capacity for energy harvest', *Nature*. Nature Publishing Group, 444(7122), pp. 1027–1031. doi: 10.1038/nature05414.

Turnbaugh, P. J. *et al.* (2007) 'The Human Microbiome Project', *Nature*, 449(7164), pp. 804–810. doi: 10.1038/nature06244.

Turroni, F. *et al.* (2012) 'Diversity of Bifidobacteria within the Infant Gut Microbiota', *PLoS ONE*. Edited by J. Neu. Public Library of Science, 7(5), p. e36957. doi: 10.1371/journal.pone.0036957.

Urbaniak, C. *et al.* (2016) 'Human milk microbiota profiles in relation to birthing method, gestation and infant gender', *Microbiome*, 4(1). doi: 10.1186/s40168-015-0145-y.

Ursell, L. K. *et al.* (2012) 'Defining the human microbiome', *Nutrition Reviews*. NIH Public Access, 70(SUPPL. 1), pp. S38-44. doi: 10.1111/j.1753-4887.2012.00493.x.

Vaz-Moreira, I. *et al.* (2011) 'Culture-dependent and culture-independent diversity surveys target different bacteria: a case study in a freshwater sample', *Antonie van Leeuwenhoek*. Springer Netherlands, 100(2), pp. 245–257. doi: 10.1007/s10482-011-9583-0.

Victora, C. G. *et al.* (2016) 'Breastfeeding in the 21st century: Epidemiology, mechanisms, and lifelong effect', *The Lancet*. Elsevier Ltd, 387(10017), pp. 475–490. doi: 10.1016/S0140-6736(15)01024-7.

Victora, C. G. and Barros, A. J. D. (2000) 'Effect of breastfeeding on infant and child mortality due to infectious diseases in less developed countries: A pooled analysis', *Lancet*, 355(9202), pp. 451–455. doi:

10.1016/S0140-6736(00)82011-5.

Ward, R. E. *et al.* (2007) 'In vitro fermentability of human milk oligosaccharides by several strains of bifidobacteria', *Molecular Nutrition and Food Research*, 51(11), pp. 1398–1405. doi: 10.1002/mnfr.200700150.

Wassenaar, T. M. and Panigrahi, P. (2014) 'Is a foetus developing in a sterile environment?', *Letters in Applied Microbiology*. Wiley/Blackwell (10.1111), 59(6), pp. 572–579. doi: 10.1111/lam.12334.

Weiser, S. D. *et al.* (2007) 'Food Insufficiency Is Associated with High-Risk Sexual Behavior among Women in Botswana and Swaziland', *PLoS Medicine*. Edited by S. Gillespie. Public Library of Science, 4(10), p. e260. doi: 10.1371/journal.pmed.0040260.

Weiss, J. E. and Rettger, L. F. (no date) *LACTOBACILLUS BIFIDUS'*.

Williams, B. *et al.* (2000) *The Natural History of HIV / AIDS in South Africa*, *Industrial Research*.

Wittig, B. M. and Zeitz, M. (2003) 'The gut as an organ of immunology.', *International journal of colorectal disease*, 18(3), pp. 181–7. doi: 10.1007/s00384-002-0444-1.

Wood, L. F. *et al.* (2017) 'T Cell Activation in South African HIV-Exposed Infants Correlates with Ochratoxin A Exposure.', *Frontiers in immunology*. Frontiers Media SA, 8, p. 1857. doi: 10.3389/fimmu.2017.01857.

Xuan, J. *et al.* (2013) 'Next-generation sequencing in the clinic: Promises and challenges', *Cancer Letters*. Elsevier Ireland Ltd, 340(2), pp. 284–295. doi: 10.1016/j.canlet.2012.11.025.

Zaura, E. *et al.* (2009) 'Defining the healthy "core microbiome" of oral microbial communities', *BMC Microbiology*. BioMed Central, 9(1), p. 259. doi: 10.1186/1471-2180-9-259.

Zhu, L. *et al.* (2018) 'Bacterial Communities in the Womb During Healthy Pregnancy', *Frontiers in Microbiology*. Frontiers, 9, p. 2163. doi: 10.3389/fmicb.2018.02163.

Appendix

Appendix A.1 Demographic data of Infant cohort for sub study

Sample ID	Status	Timepoint	Infants gender	Infants gestational age in weeks	Age of the mother	Education	Occupation	Live	Electricity	Refrigerator	Running Water	Status
004	HEU	Birth	Male	40	21	Elementary/Junior	Unemployed	Shack	Yes	No	No	Single/ never married
008	HEU	Birth	Female	38	26	Secondary/ High School	Unemployed	Shack	Yes	Yes	No	Single/ never married
009	HEU	D4-7	Female	36	30	Secondary/ High School	Unemployed	Shack	No	No	No	Married
010	HEU	D4-7	Male	37	36	Elementary/Junior	Unemployed	Shack	Yes	Yes	No	Single/ never married
011	HEU	D4-7	Male	37	27	Secondary/ High School	Unemployed	Formal House	Yes	Yes	No	Single/ never married
013	HEU	D4-7	Male	39	32	Secondary/ High School	Salaried Private	Formal House	Yes	Yes	No	Married
016	HEU	D4-7	Male	37	28	Higher Education(College/ Diploma/etc)	Salaried Private	Formal House	Yes	Yes	Yes	Single/ never married
017	HEU	D4-7	Female	38	26	Secondary/ High School	Unemployed	Shack	Yes	Yes	No	Single/ never married
018	HEU	D4-7	Female	40	20	Higher Education(College/ Diploma/etc)	Student	Shack	Yes	Yes	Yes	Single/ never married
020	HEU	Birth	Female	38	35	Secondary/ High School	Unemployed	Shack	Yes	Yes	No	Single/ never married
026	HEU	Birth	Female	40	30	Secondary/ High School	Unemployed	Formal House	Yes	Yes	Yes	Single/ never married
038	HEU	D4-7	Female	38	21	Secondary/ High School	Unemployed	Shack	Yes	Yes	No	Single/ never married

054	HEU	Birth	Female	40	28	Secondary/ High School	Unemployed	Shack	Yes	Yes	No	Single/ never married
055	HEU	Birth	Female	37	21	Secondary/ High School	Unemployed	Shack	Yes	Yes	No	Single/ never married
061	HEU	D4-7	Male	40	25	Secondary/ High School	Self Employed	Formal House	Yes	Yes	Yes	Single/ never married
062	HEU	Birth	Male	40	24	Secondary/ High School	Self Employed	Shack	Yes	Yes	Yes	Single/ never married
063	HEU	Birth	Female	38	22	Secondary/ High School	Unemployed	Shack	Yes	Yes	No	Married
066	HEU	Birth	Female	37	24	Secondary/ High School	Unemployed	Shack	Yes	Yes	No	Single/ never married
071	HEU	Birth	Female	40	29	Secondary/ High School	Domestic/ housekeeper	Shack	Yes	Yes	No	Single/ never married
078	HEU	D4-7	Female	39	35	Secondary/ High School	Unemployed	Shack	Yes	Yes	No	Living together
081	HEU	Birth	Male	40	21	Secondary/ High School	Unemployed	Formal House	Yes	Yes	Yes	Single/ never married
088	HEU	Birth	Female	40	24	Secondary/ High School	Unemployed	Shack	Yes	Yes	No	Married
092	HEU	D4-7	Female	40	29	Secondary/ High School	Salaried Private	Shack	Yes	Yes	No	Single/ never married
100	HEU	Birth	Female	39	23	Higher Education(College/ Diploma/etc)	Student	Formal House	Yes	Yes	Yes	Single/ never married
121	HEU	D4-7	Male	38	29	Secondary/ High School	Unemployed	Shack	Yes	No	No	Single/ never married
127	HEU	D4-7	Male	37	32	Secondary/ High School	Unemployed	Formal House	Yes	Yes	Yes	Single/ never married
138	HEU	D4-7	Male	39	29	Secondary/ High School	Unemployed	Formal House	Yes	Yes	Yes	Single/ never married
140	HEU	D4-7	Female	37	33	Secondary/ High School	Salaried Private	Formal House	Yes	Yes	Yes	Married

142	HEU	D4-7	Male	37	25	Secondary/ High School	Salaried Private	Formal House	Yes	Yes	Yes	Single/ never married
143	HEU	D4-7	Female	37	27	Secondary/ High School	Unemployed	Shack	Yes	Yes	No	Single/ never married
156	HEU	D4-7	Male	38	30	Secondary/ High School	Domestic/ housekeeper	Shack	Yes	Yes	No	Single/ never married
157	HEU	D4-7	Male	39	29	Secondary/ High School	Domestic/ housekeeper	Formal House	Yes	Yes	Yes	Single/ never married
159	HEU	D4-7	Male	39	19	Secondary/ High School	Unemployed	Formal House	Yes	Yes	Yes	Single/ never married
161	HEU	D4-7	Male	39	26	Secondary/ High School	Salaried Private	Shack	Yes	Yes	No	Single/ never married
302	HU	Birth	Male	36	27	Secondary/ High School	Salaried Private	Formal House	Yes	Yes	Yes	Single/ never married
305	HU	D4-7	Male	38	26	Secondary/ High School	Unemployed	Shack	No	No	No	Single/ never married
307	HU	Birth	Male	39	19	Secondary/ High School	Unemployed	Shack	Yes	No	No	Single/ never married
308	HU	D4-7	Male	38	27	Secondary/ High School	Unemployed	Formal House	Yes	Yes	Yes	Single/ never married
309	HU	D4-7	Female	40	32	Secondary/ High School	Unemployed	Shack	Yes	Yes	No	Single/ never married
312	HU	D4-7	Female	41	36	Secondary/ High School	Unemployed	Shack	No	No	No	Single/ never married
314	HU	D4-7	Male	39	29	Secondary/ High School	Unemployed	Shack	Yes	Yes	No	Single/ never married
315	HU	D4-7	Male	39	23	Higher Education(College/ Diploma/etc)	Student	Formal House	Yes	Yes	Yes	Single/ never married
316	HU	D4-7	Female	37	19	Secondary/ High School	Student	Shack	Yes	Yes	No	Single/ never married
318	HU	D4-7	Female	39	30	Secondary/ High School	Unemployed	Shack	Yes	Yes	No	Married

319	HU	D4-7	Female	38	35	Secondary/ High School	Unemployed	Formal House	Yes	Yes	Yes	Living together
320	HU	D4-7	Female	39	32	Secondary/ High School	Unemployed	Shack	Yes	Yes	No	Single/ never married
325	HU	D4-7	Female	39	24	Secondary/ High School	Domestic/ housekeeper	Shack	No	No	No	Living together
332	HU	D4-7	Male	38	28	Secondary/ High School	Domestic/ housekeeper	Formal House	Yes	Yes	No	Single/ never married
337	HU	D4-7	Male	37	28	Secondary/ High School	Salaried Private	Formal House	Yes	Yes	Yes	Single/ never married
345	HU	D4-7	Female	39	24	Secondary/ High School	House Wife	Shack	Yes	Yes	Yes	Married
347	HU	D4-7	Male	39	19	Secondary/ High School	Unemployed	Shack	Yes	Yes	No	Single/ never married
350	HU	D4-7	Male	37	24	Secondary/ High School	Unemployed	Shack	Yes	Yes	No	Single/ never married
354	HU	Birth	Female	38	24	Secondary/ High School	Salaried Private	Formal House	Yes	Yes	Yes	Single/ never married
357	HU	Birth	Male	39	22	Secondary/ High School	Salaried Private	Formal House	Yes	Yes	Yes	Single/ never married
359	HU	Birth	Male	40	26	Secondary/ High School	Unemployed	Shack	Yes	Yes	Yes	Single/ never married
360	HU	D4-7	Female	40	26	Secondary/ High School	Domestic/ housekeeper	Shack	Yes	Yes	No	Single/ never married
365	HU	Birth	Female	38	28	Secondary/ High School	Unemployed	Formal House	Yes	Yes	Yes	Married
374	HU	Birth	Male	37	24	Secondary/ High School	Self Employed	Formal House	Yes	Yes	No	Married
377	HU	D4-7	Female	40	22	Secondary/ High School	Student	Formal House	Yes	Yes	Yes	Single/ never married
381	HU	D4-7	Female	39	26	Secondary/ High School	Unemployed	Formal House	Yes	Yes	Yes	Single/ never married

383	HU	D4-7	Male	38	23	Secondary/ High School	Unemployed	Formal House	Yes	Yes	No	Single/ never married
387	HU	D4-7	Male	39	19	Secondary/ High School	Unemployed	Shack	Yes	No	No	Single/ never married
389	HU	D4-7	Male	38	23	Secondary/ High School	Salaried Private	Formal House	Yes	Yes	Yes	Single/ never married

UNIVERSITY OF CAPE TOWN
Faculty of Health Sciences
Human Research Ethics Committee

Room E53-46 Old Main Building
Groote Schuur Hospital
Observatory 7925
Telephone [021] 406 6492
Email: sumayah.ariefdien@uct.ac.za
Website: www.health.uct.ac.za/fhs/research/humanethics/forms

24 January 2018

HREC REF: 062/2018

Prof J Blackburn
Division of Integrative Biomedical Sciences
N3.06, IDM

Dear Prof Blackburn

PROJECT TITLE: INNATE, ADAPTIVE AND MUCOSAL IMMUNE RESPONSES IN HIV-1 EXPOSED UNINFECTED INFANTS: A HUMAN MODEL TO UNDERSTAND CORRELATES OF IMMUNE PROTECTION (MSc. Med Candidate - Ms P Khomunala) SUB-STUDY LINKED TO 285/2012

Thank you for submitting your study to the Faculty of Health Sciences Human Research Ethics Committee (HREC) for review.

It is a pleasure to inform you that the HREC has **formally approved** the above-mentioned study.

Approval is granted for one year until the 30 January 2019.

Please submit a progress form, using the standardised Annual Report Form if the study continues beyond the approval period. Please submit a Standard Closure form if the study is completed within the approval period.
(Forms can be found on our website: www.health.uct.ac.za/fhs/research/humanethics/forms)

We acknowledge that the student: P Khomunala will also be involved in this study.

Please quote the HREC REF in all your correspondence.

Please note that the ongoing ethical conduct of the study remains the responsibility of the principal investigator.

Please note that for all studies approved by the HREC, the principal investigator **must** obtain appropriate institutional approval, where necessary, before the research may occur.

Yours sincerely

Signature Removed

PROFESSOR M BLOCKMAN
CHAIRPERSON, FHS HUMAN RESEARCH ETHICS COMMITTEE

Federal Wide Assurance Number: FWA00001637.
Institutional Review Board (IRB) number: IRB00001938

This serves to confirm that the University of Cape Town Human Research Ethics Committee complies to the Ethics Standards for Clinical Research with a new drug in patients, based on the Medical Research Council (MRC-SA), Food and Drug Administration (FDA-USA), International Convention on Harmonisation Good Clinical Practice (ICH GCP), South African Good Clinical Practice Guidelines (DoH 2006), based on the Association of the British Pharmaceutical Industry Guidelines (ABPI),and Declaration of Helsinki (2013) guidelines.
The Human Research Ethics Committee granting this approval is in compliance with the ICH Harmonised Tripartite Guidelines E6: Note for Guidance on Good Clinical Practice (CPMP/ICH/135/95) and FDA Code Federal Regulation Part 50, 56 and 312.

HREC :062/2018

A.3 Clinical information of infant mother diads recruited for study

Sample ID	Status	Timepoint	Infants gender	Infants birth weight (g)	infants birth height (cm)	Infants head circumference (cm)	Infants temperature (celsius)	Infants gestational age in weeks	Infants on Arv	Vaccine History of Infant		Mothers CD4 count (cells/mm3)	Mothers HAART treatment	Age of the mother
004	HEU	Birth	Male	3300	47	33	36,5	40	Yes	BCG	OPV	478	Tribuss FDC	21
008	HEU	Birth	Female	3000	45	30	36,5	38	Yes	BCG	OPV	568	Tribuss	26
009	HEU	D4-7	Female	2900	49	34,5	36,8	36	Yes	BCG	OPV	470	unknown	30
010	HEU	D4-7	Male	2510	46	31	37,3	37	Yes	BCG	OPV	830	unknown	36
011	HEU	D4-7	Male	2920	49	31	36,5	37	Yes	BCG	OPV	487	FDC=FTC,EFV, TDF	27
013	HEU	D4-7	Male	3770	48	31	36,2	39	Yes	BCG	OPV	226	FDC=FTC,EFV, TDF	32
016	HEU	D4-7	Male	2520	44	31	36,4	37	Yes	BCG	OPV	526	FDC=FTC,EFV, TDF	28
017	HEU	D4-7	Female	2720	46	31	36,7	38	Yes	BCG		353	unknown	26
018	HEU	D4-7	Female	3160	50	32	36,2	40	Yes	BCG	OPV	470	FDC=FTC,EFV, TDF	20
020	HEU	Birth	Female	3010	49	34	36,6	38	Yes	BCG	OPV	374	FDC=FTC,EFV, TDF	35
026	HEU	Birth	Female	3430	50	32	36	40	Yes	BCG	OPV	456	TDF,FTC,EFV	30
038	HEU	D4-7	Female	3150	46	32	36,2	38	Yes	BCG	OPV	412	FDC=FTC,EFV, TDF	21
054	HEU	Birth	Female	3880	47	35	36,9	40	Yes	BCG	OPV	227	unknown	28

055	HEU	Birth	Female	3130	49	34	36,1	37	Yes	BCG	OPV	272	FDC=FTC,EFV, TDF	21
061	HEU	D4-7	Male	3130	43	33	36	40	Yes	BCG	OPV	165	FTC,EFV,TDF	25
062	HEU	Birth	Male	3610	49	34	36,3	40	Yes	BCG	OPV	445	FDC=FTC,EFV, TDF	24
063	HEU	Birth	Female	3130	46	33	36	38	Yes	BCG	OPV	433	TDF,FTC,EFV	22
066	HEU	Birth	Female	2520	44	34	36,7	37	Yes	BCG	OPV	458	FDC=FTC,EFV, TDF	24
071	HEU	Birth	Female	3030	56	35	36,5	40	Yes	BCG	OPV	595	unknown	29
078	HEU	D4-7	Female	3320	45	32	36	39	Yes	BCG	OPV	350	TDF,FTC,EFV	35
081	HEU	Birth	Male	3040	48	32	36,7	40	Yes	BCG	OPV	497	FDC=FTC,EFV, TDF	21
088	HEU	Birth	Female	3300	48	33	36,2	40	Yes	BCG	OPV	388	unknown	24
092	HEU	D4-7	Female	2750	46	31	36,6	40	Yes	BCG	OPV	795	unknown	29
100	HEU	Birth	Female	3020	48	36	36,3	39	Yes	BCG	OPV	366	unknown	23
121	HEU	D4-7	Male	2720	50	34	36,5	38	Yes	BCG	OPV	319	FTC,EFV,TDF	29
127	HEU	D4-7	Male	3630	50	34	36,6	37	Yes	BCG	OPV	493	TDF,FTC,EFV	32
138	HEU	D4-7	Male	3380	52	35	37	39	Yes	BCG	OPV	368	FDC=FTC,EFV, TDF	29
140	HEU	D4-7	Female	3230	50	35	36,6	37	Yes	BCG	OPV	715	unknown	33
142	HEU	D4-7	Male	2500	46	35	37,4	37	Yes	BCG	OPV	451	FDC=FTC,EFV, TDF	25

143	HEU	D4-7	Female	2780	50	33	37,3	37	Yes	BCG	OPV	685	TDF,FTC,EFV	27
156	HEU	D4-7	Male	3220	46	34	37	38	Yes	BCG	OPV	313	FDC=FTC,EFV, TDF	30
157	HEU	D4-7	Male	2650	48	31	36,3	39	Yes	BCG	OPV		FDC=FTC,EFV, TDF	29
159	HEU	D4-7	Male	3620	53	33	36	39	Yes	BCG	OPV	381	FDC=FTC,EFV, TDF	19
161	HEU	D4-7	Male	3500	47	33	36,4	39	Yes	BCG	OPV	444	FDC=FTC,EFV, TDF	26
302	HU	Birth	Male	3690	51	34	36,7	36	N/A	BCG	OPV	N/A	N/A	27
305	HU	D4-7	Male	3310	42	32	35,5	38	N/A	BCG	OPV	N/A	N/A	26
307	HU	Birth	Male	3040	48	35	36,4	39	N/A	BCG	OPV	N/A	N/A	19
308	HU	D4-7	Male	3160	42	32	34,7	38	N/A	BCG	OPV	N/A	N/A	27
309	HU	D4-7	Female	3130	48	33	36,1	40	N/A	BCG	OPV	N/A	N/A	32
312	HU	D4-7	Female	3320	44	36	36,2	41	N/A	BCG	OPV	N/A	N/A	36
314	HU	D4-7	Male	3000	48	32	37	39	N/A	BCG	OPV	N/A	N/A	29
315	HU	D4-7	Male	3040	47	31	36	39	N/A	BCG	OPV	N/A	N/A	23
316	HU	D4-7	Female	2780	49	34	36,4	37	N/A	BCG	OPV	N/A	N/A	19
318	HU	D4-7	Female	3090	48	33	37	39	N/A	BCG	OPV	N/A	N/A	30
319	HU	D4-7	Female	2770	40	29	36,6	38	N/A	BCG	OPV	N/A	N/A	35

320	HU	D4-7	Female	2930	50	34	36	39	N/A	BCG	OPV	N/A	N/A	32
325	HU	D4-7	Female	2930	49	32	36,8	39	N/A	BCG	OPV	N/A	N/A	24
332	HU	D4-7	Male	2880	50	31	36,5	38	N/A	BCG	OPV	N/A	N/A	28
337	HU	D4-7	Male	2730	47	33	36,9	37	N/A	BCG	OPV	N/A	N/A	28
345	HU	D4-7	Female	3320	48	36	36,8	39	N/A	BCG	OPV	N/A	N/A	24
347	HU	D4-7	Male	3370	51	33	37,3	39	N/A	BCG	OPV	N/A	N/A	19
350	HU	D4-7	Male	3230	48	34	36,2	37	N/A	BCG	OPV	N/A	N/A	24
354	HU	Birth	Female	3110	48	34	36,6	38	N/A	BCG	OPV	N/A	N/A	24
357	HU	Birth	Male	3110	51	36	37,1	39	N/A	BCG	OPV	N/A	N/A	22
359	HU	Birth	Male	3590	54	33,5	36,9	40	N/A	BCG	OPV	N/A	N/A	26
360	HU	D4-7	Female	3730	51	35	37,2	40	N/A	BCG	OPV	N/A	N/A	26
365	HU	Birth	Female	3090	46	32	36,2	38	N/A	BCG	OPV	N/A	N/A	28
374	HU	Birth	Male	3130	48	36	36,5	37	N/A	BCG	OPV	N/A	N/A	24
377	HU	D4-7	Female	3100	44	35	36,6	40	N/A	BCG	OPV	N/A	N/A	22
381	HU	D4-7	Female	3380	48	36	36,2	39	N/A	BCG	OPV	N/A	N/A	26
383	HU	D4-7	Male	3100	44	34	37,7	38	N/A	BCG	OPV	N/A	N/A	23

387	**HU**	**D4-7**	**Male**	**2900**	**45**	**31**	**37,1**	**39**	**N/A**	**BCG**	**OPV**	**N/A**	**N/A**	**19**
389	**HU**	**D4-7**	**Male**	**2620**	**46**	**35,5**	**36,8**	**38**	**N/A**	**BCG**	**OPV**	**N/A**	**N/A**	**23**

B.1) Concentration and TIC of samples that were processed

STOOL LOG	STATUS	Concentration (ug\uL)	Timepoint	TIC
Sample				
4	HEU	3,75	Birth	9,44E+09
8	HEU	2,391	Birth	9,76E+08
9	HEU	4,103	D4-7	5,83E+09
10	HEU	(-)	D4-7	3,66E+09
16	HEU	3,56	D4-7	3,78E+08
17	HEU	8,61	D4-7	3,25E+09
18	HEU	5,452	D4-7	8,29E+09
20	HEU	0,183	Birth	5,43E+09
26	HEU	3,257	Birth	6,98E+09
38	HEU	(-)	D4-7	4,21E+09
54	HEU	(-)	Birth	4,92E+09
55	HEU	(-)	Birth	3,55E+09
62	HEU	3,524	Birth	3,37E+09
63	HEU	1,111	Birth	9,64E+08
66	HEU	0,113	Birth	12200000
71	HEU	0,939	Birth	9,76E+08
81	HEU	(-)	Birth	9,98E+08
88	HEU	3,98	Birth	1,05E+09
92	HEU	1,562	D4-7	5,74E+09
100	HEU	5,77	Birth	1,83E+09
121	HEU	1,699	D4-7	5,81E+09
127	HEU	1,836	D4-7	5,6E+09
140	HEU	5,61	D4-7	1,21E+09
142	HEU	6,34	D4-7	1,38E+08
143	HEU	0,34	D4-7	5,73E+09
156	HEU	6,589	D4-7	1,22E+08
159	HEU	4,69	D4-7	1,05E+09
161	HEU	2,15	D4-7	2,19E+09
302	HU	5,866	Birth	5,13E+09
305	HU	(-)	D4-7	2,81E+09
307	HU	3,404	Birth	7,87E+09
308	HU	5,188	D4-7	4,95E+09
309	HU	1,754	D4-7	9,94E+08
312	HU	1,73	D4-7	3,62E+09
314	HU	1,64	D4-7	1,97E+09
315	HU	7,83	D4-7	1,47E+08
319	HU	5,15	D4-7	1,28E+08
320	HU	1,83	D4-7	1,4E+09
325	HU	0,267	D4-7	2,13E+08
332	HU	3,027	D4-7	8,89E+09
337	HU	1,892	D4-7	1,05E+10

347	HU	1,304	D4-7	2,1E+09
350	HU	(-)	D4-7	9,65E+08
354	HU	2,268	Birth	4,66E+09
357	HU	3,8	Birth	1,63E+09
359	HU	8,99	Birth	1,55E+08
360	HU	6,88	D4-7	1,49E+08
365	HU	8,48	Birth	1,01E+09
374	HU	(-)	Birth	5,6E+09
377	HU	3,98	D4-7	1,29E+09
381	HU	6,54	D4-7	2,2E+09
383	HU	7,86	D4-7	9,61E+08
387	HU	0,836	D4-7	3,06E+09
389	HU	2,91	D4-7	1,49E+08