Placental, oral and vaginal microbiomes and birth outcomes in rural Malawi

Thesis for the degree of

Doctor of Philosophy

UCL

Ronan Doyle

Infection, Immunity, Inflammation and Physiological Medicine

UCL Institute of Child Health

30 Guilford Street

London WC1N 1EH

September 2015

1 Declaration

The following people contributed to the work presented in this thesis:

Participant recruitment, data collection and anthropometric measurements were carried out by relevant members of the iLiNS-DYAD-M clinical trial team. Sample collections of the placenta and vaginal swabs were carried out by study nurses and dental swab collection was carried out by study dental therapists.

Histological analysis of placental tissue was carried out by Steve Kamiza and Kingsley Makwakwa at the Malawi College of Medicine. Manual DNA extractions were carried out by myself, Moses Masewo, David Chaima and John Kamwendo

I, Ronan Doyle, confirm that all PCR, high-throughput sequencing, relevant data cleaning, bioinformatics and statistical analysis presented in this thesis is my own work.

Signed:

Ronan Doyle

2 Abstract

Being born too early or being born too small is the largest cause of neonatal mortality in the world. Compared to the rest of the world, Malawi has one of the largest burdens of preterm birth and neonatal stunting, with infection recognised as an important risk factor. Previous studies have used culture and molecular methods to identify bacteria that could be responsible for triggering labour and foetal growth restriction. The composition of the oral and vaginal microbiome has also been linked as the possible source of these bacteria. However, studies up to this point have been small and have not utilised the full potential of current sequencing technologies. In this thesis, I demonstrated using high-throughput sequencing of the 16S rRNA gene that certain organisms are associated with adverse birth outcomes. Contaminating bacterial taxa, PCR and sequencing error can be filtered post-sequencing to allow reliable reconstruction of microbial communities from low biomass samples such as the placenta. This revealed a specific community structure in the placenta and foetal membranes associated with severe chorioamnionitis. Analysis of communities in both matched vaginal and placental samples increased prevalence of *Peptostreptococcus anaerobius*, *Sneathia sanguinegens* and *Prevotella amnii* were associated with a smaller newborn size. These results provide further evidence of the important role the vaginal microbiome may play in seeding organisms found on placental tissues and therapeutic interventions could be designed to impact these communities with the goal of reducing the risk of preterm birth or intrauterine growth restriction.

3 Table of Contents

4 Abbreviations

95%CI	95 per cent confidence interval
ANOVA	Analysis of variance
AV	Aerobic vaginitis
BLASTN	Nucleotide basic local alignment search tool
BMI	Body mass index
Bp	Base pair
BV	Bacterial vaginosis
BVAB-3	Bacterial vaginosis associated bacteria-3
CI	Confidence interval
Ct	Cycle threshold
DNA	Deoxyribonucleic acid
FASTA	Fast-All format
FASTQ	Fast-Quality format
FDR	False discovery rate
FLASH	Fast length adjustment of short reads
Hb	Haemoglobin
HCA	Histologic chorioamnionitis
HCZ	Head circumference-for-age z-score
HIV	Human immunodeficiency virus
IFA	Iron and folic acid
iLiNS	International lipid-based nutrient supplements
IUGR	Intrauterine growth restriction
LAZ	Length-for-age z-score
LNS	Lipid-based nutrient supplement

Mb	Megabase
MMN	Multiple micronutrients
NR	Non-redundant
OTU	Operational taxonomic unit
PCR	Polymerase chain reaction
PPROM	Preterm premature rupture of membranes
QIIME	Quantitative insights into microbial ecology
qPCR	Quantitative polymerase chain reaction
rDNA	Ribosomal deoxyribonucleic acid
RDT	Rapid diagnostic test
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
SD	Standard deviation
SEM	Standard error of the mean
SNP	Single nucleotide polymorphism
SSU	Small subunit
WHO	World health organisation
µl	Microlitre

5 List of figures

6 List of tables

7 Acknowledgements

I would like to thank my supervisors, Nigel Klein and Kathryn Harris for their support and input throughout this project. Thanks also to office 646 and Vania De Toledo who always ensured I actually got paid.

Special thanks go to Per Ashorn and all my Finnish, Malawian, Australian and American friends who worked on iLiNS-DYAD with me. I found that everyone was happy to help whenever they could and this made me feel very much at home in Malawi.

8 Chapter 1: Introduction

8.1 Adverse birth outcomes

The definition of preterm birth is not fixed and can vary country-to-country. In the UK it is defined as a delivery that occurs before 37 gestational weeks, and being born within this time frame leads to an increased chance of mortality and morbidity. These problems are exacerbated when born before 32 weeks (1); this sub-category of preterm deliveries can be subdivided into very preterm births (32-28 weeks) and extremely preterm births (before 28 weeks).

At the turn of the millennium it was estimated that premature birth was the direct cause 1,080,000 neonatal deaths, world-wide, every year (2). It is considered the largest cause of neonatal mortality in the world (3). Approximately 75% of neonatal deaths occur in premature infants. In developed countries improved standard of care has led to a drop in mortality for neonates born preterm but consequently has led to an increase in morbidity for those extremely preterm neonates who survive. The burden of mortality and morbidity increases again for twin births over singletons with twins more likely to be delivered preterm. In the long term premature infants are more likely to be effected by blindness, deafness, cerebral palsy, stunting, learning disabilities, as well as respiratory and gastrointestinal problems (1,4). A lower gestational age is proportional to the increase in

morbidity with a UK-based study reporting 50% of neonates delivered between 24 and 25 weeks classified as impaired and 25% considered severely disabled (5).

Intrauterine growth restriction (IUGR) resulting in low birth weight and other measures of neonatal size such as neonatal length and head circumference are also associated with both mortality and morbidity (6). Low birth weight itself is defined by any live birth where the infant weighs 2500 grams or less. Although low birth weight is strongly associated to preterm birth, it can also occur in babies born at term due to a number of factors that limit foetal growth in utero. Excluding those born preterm, infants born with low birth weight have a 5-30 greater risk of mortality than those born greater than 2500g (7). In the United States it is estimated that 40% of cases of cerebral palsy are diagnosed in children with a very low birth weight (8). Stunting and small head circumference at birth is also an indicator of IUGR. Stunting is defined by the World Health Organisation (WHO) as being -2 standard deviations lower than the median Z-score for length-for-age and small head circumference is defined as -2 standard deviations lower than the median head circumference-for-age Z-score. In 2011, 165 million children worldwide were stunted (9). Stunting, along with other measures of undernutrition are estimated to cause 3.1 million child deaths annually (9) and neonates with a small head circumference are at a greater risk of impairment of neurological function (10). The knock on effect of this mortality and morbidity can also be measured in state expenditure, with the USA estimated to have spent $26.2

billion in 2005 dealing with the medical and educational costs of these complications (11).

The worldwide preterm birth rate is estimated at 9.6% (11), with the figure particularly low in most developed countries (12). In the UK, infant mortality occurred in 24.3 per 1000 preterm births compared to 1.6 per 1000 term births (13). Where accurate reporting has been available, preterm birth rates have been rising in the developed world (14), however the effect is mitigated by the fact that there is now a near 100% survival rate for neonates born after 32 weeks, dropping only marginally with earlier gestational ages (13).

Although preterm birth and IUGR are still a recognised, serious problem in the developed countries, relatively it is a greater risk and burden in developing countries. Sub-Saharan Africa is known to have the current largest incidence of preterm birth in the world, with preterm birth in Southern Africa approximated at 17.5% (95%CI, 14.6-20.3) of all births (Figure 8-1).

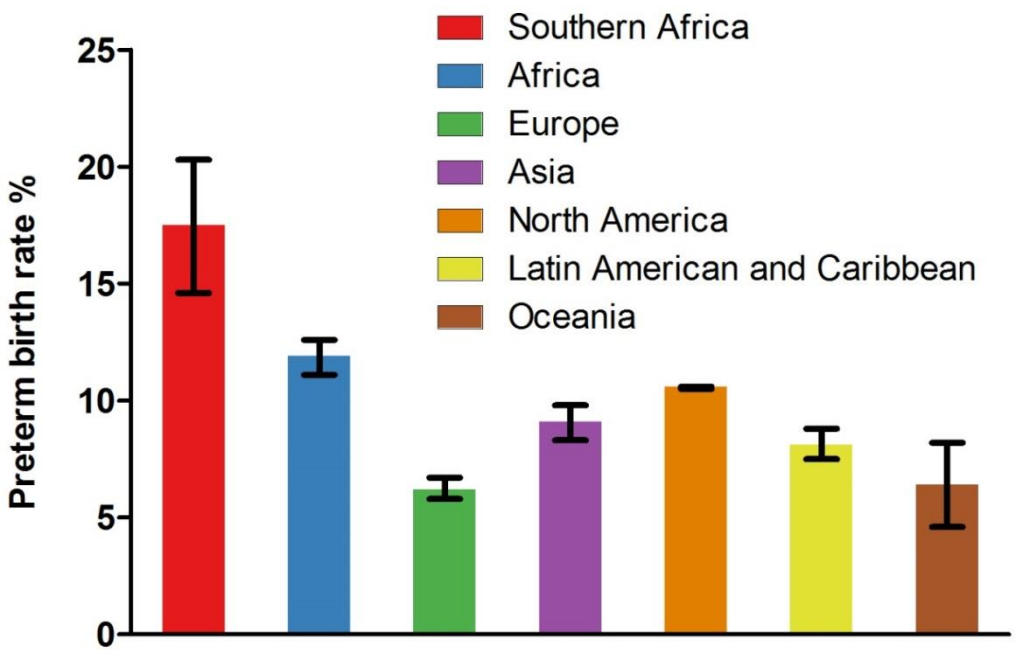

Figure 8-1: Preterm birth rate in Southern Africa in 2005 compared to other geographical regions.
Data used in plot published by *Beck et al* (11).

In Malawi, although figures vary, a recent study placed the incidence of preterm birth (<37 weeks gestation) at 15.1% by ultrasound assessment (15). This is a serious problem for such a resource poor country with current estimates of overall neonatal mortality (including both term and preterm births) ranging between 96.8 (95%CI, 84.3–111.6) per 1000 births (16) and 33 per 1000 births (17) in 2010. This figure rises even further to 132 per 1000 births in children born preterm (18). Complications due to the preterm birth were the largest cause of mortality for babies aged between 0-27 days in Malawi in 2010 (Figure 8-2).

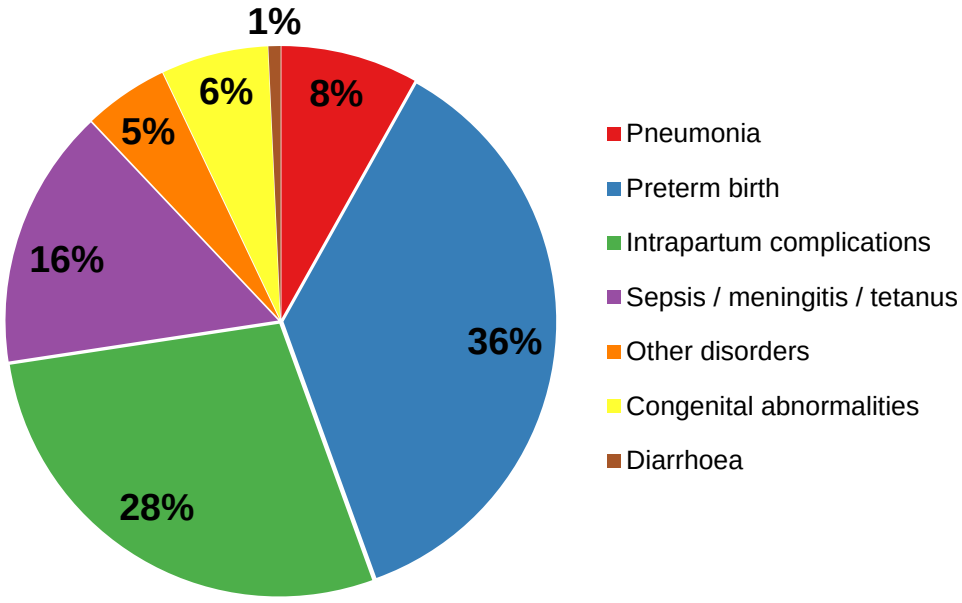

Figure 8-2: Causes of mortality in Malawi in 2010 for infants aged between 0-27 days.
Data used to in plot published by *Liu et al* (19).

Malawi's health infrastructure, especially in rural areas, means that newborns are at a severe risk of mortality even when born after 32 weeks, whereas studies in the UK have just focused on those born very preterm. In Malawi, there is very little data on long-term morbidity of the children who survive being born at 32-37 weeks gestation and so the scale of the problem remains largely unknown. By any measure, it is more dangerous to be born at term in Malawi than preterm in the UK. For a child born very preterm in Malawi the risks are even greater, and strategies are urgently needed to bring down both mortality and morbidity.

8.2 Causes of adverse birth outcomes

Before a clear intervention can be formulated, a better understanding of the pathology is needed. A number of different causes have been associated with preterm birth and intrauterine growth restriction, with many being inter-connected. The mechanisms behind pregnancy and labour are complex and not fully understood, so it is unlikely that preterm birth itself will be a single aetiology but rather a culmination of multiple factors (Figure 8-3).

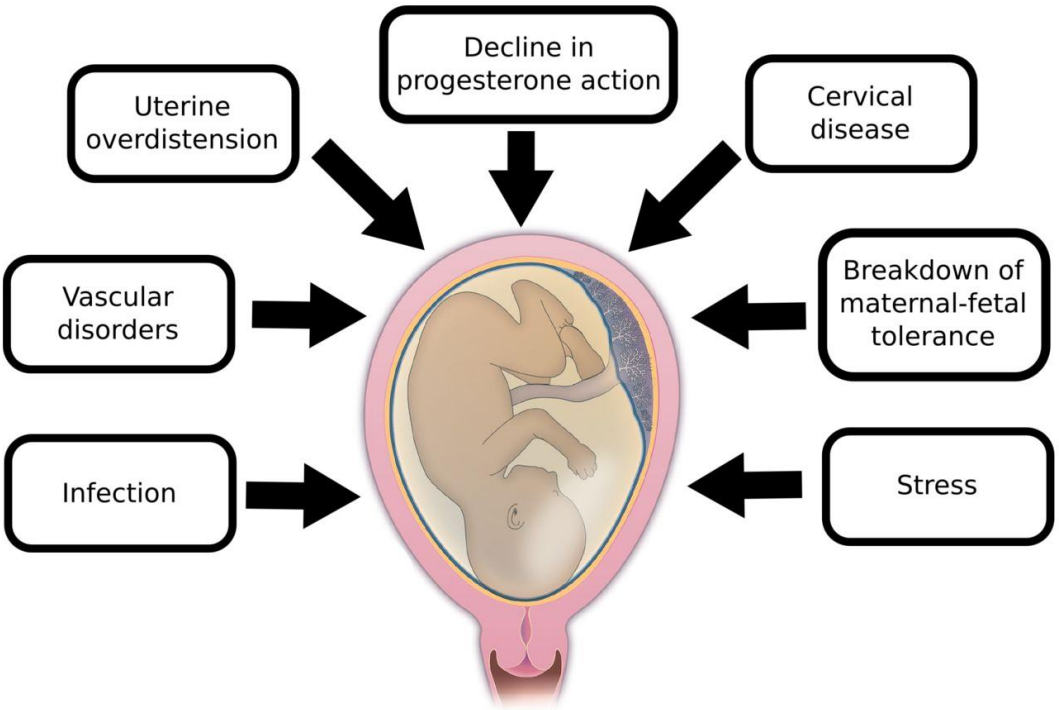

Figure 8-3: Factors linked to an increased risk of preterm birth.
Edited from Romero *et al* (20).

Current knowledge suggests that term and preterm parturition use the same pathways, and so the instigator of labour must differ between the two (21). As well as by gestational age, preterm birth can also be split into 3 distinct types.

Indicated preterm birth is generally due to a maternal or foetal disorder such as pre-eclampsia and can be caused by smoking or obesity and, in countries with good healthcare coverage, can be diagnosed early culminating in an elective caesarean section (4,14). The other types of preterm labour are spontaneous and are distinguishable by delivery with intact membranes with regular contractions or preterm premature rupture of membranes (PPROM) before contractions begin. There are multiple risk factors that have been linked including cervical length, previous history of preterm birth, multiple pregnancy, stress, uteroplacental ischemia, placental abruption, intrauterine infection and up to 30 different SNPs have been identified as potential genetic markers (14,22). These include a mutation in the promotor region of tumor necrosis factor α that leads to oversecretion of the cytokine and increase in risk of preterm birth. Polymorphisms in the genes for other pro-inflammatory cytokines such as interleukin 4, 6 and 10 have also been linked to a higher risk of preterm birth or premature rupture of membranes. Malawi also has serious region specific causes that are relatively uncommon in the developed countries including malnutrition, maternal malaria parasitaemia and HIV infection which have been shown as having a key influence on preterm birth (23–25).

Microbiological studies have indicated infection as a key causative factor in around ~25% of preterm births, with the presence of bacteria involved in ~80% of preterm deliveries depending on the gestational age. Exact mechanisms for triggering labour before 37 weeks are still unknown but they involve many potential risk factors that stimulate inflammatory pathways or

allow the spread of infection. Knowledge of a link between infection and preterm birth has existed since the 1970s (26) and is generally associated with those delivering spontaneously rather than indicated preterm births (27). The presence of intrauterine infection is inversely correlated with gestational age. Depending on the method used it identify microorganisms, amniotic fluid was more likely to be positive for bacteria the early the gestational age at delivery (28,29). This was true even in pregnancies with intact membranes. Viral infections as a cause of IUGR have been well studied. Human cytomegalovirus in particular has been shown to be able to translocate to the placenta and infect fetal tissue (30). Depending on the timing of these infections, this can impair placental development and could be an underlying cause of IUGR. Although work with viruses in terms of fetal growth restriction has been forthcoming, the same associations have not been commonly studied in bacteria.

8.3 Chorioamnionitis

Preterm birth or IUGR caused by infection is mediated by the recognition of microorganisms by the immune system (31). Diagnosis of chorioamnionitis has been extensively used as evidence of intrauterine infection and is defined by inflammation of the chorioamnion tissues. Clinical diagnosis of chorioamnionitis is generally thought to be caused by hyper virulent strains of known human pathogens and diagnosed with the preterm rupture of membranes (32,33). More commonly chorioamnionitis is diagnosed

histologically; this is able to detect less severe but clinically relevant cases of inflammation. Histological chorioamnionitis (HCA) is defined by inflammatory cell infiltration into the placenta and membranes, however multiple semi-quantitative definitions exist leading to discrepancy in interpretation of results (34). Funisitis occurs when this inflammatory process reaches the umbilical cord and foetuses that are exposed to severe chorioamnionitis can develop fetal inflammatory response syndrome (FIRS). This syndrome is defined sub-clinically by funisitis and is a significant risk factor for perinatal mortality (35), fetal cardiac dysfunction (36), lung damage (37) and neurological impairment (38). Chorioamnionitis can put the foetus in direct risk due to migration of pro-inflammatory cells from the site of infection or contact with potentially infected amniotic fluid (39). *In vitro* modelling of the most severe manifestations of chorioamnionitis show an increased incidence of systemic inflammation, preterm birth and foetal loss (40–42). A retrospective histopathological study showed that histological chorioamnionitis and funisitis were more commonly seen in deliveries that had neonatal sepsis, perinatal mortality or low birth weight (43). Chorioamnionitis is significantly more prevalent in preterm births than term births, and this relationship is more pronounced the earlier the delivery. Interestingly, chorioamnionitis is more prevalent amongst term deliveries in black populations (44).

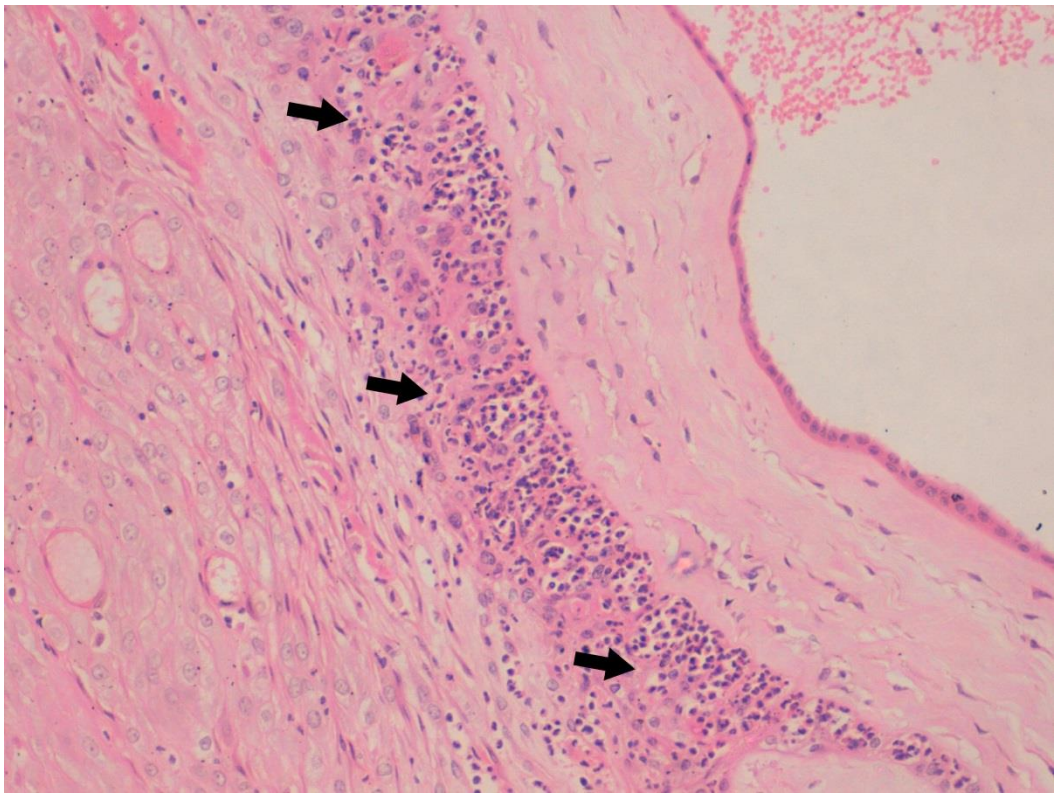

Figure 8-4: Histologic section of chorioamnion (fetal membranes) from a participant showing infiltration of neutrophils (chorioamnionitis) between the membranes and the decidua denoted by the black arrows.

8.4 16S rRNA gene

Early studies used microbial culture techniques to sample and diagnose the presence of bacteria in the amniotic fluid and chorio-decidual space. Many preterm deliveries showed signs of inflammation in a number of the maternal and fetal tissues but there was no evidence of infection by culture. This made diagnosing intrauterine infection as a potential cause of preterm increasingly difficult (45). During this period, researchers found that measuring cytokine levels such as interleukin-6 was a better predictor of preterm birth than the

presence or absence of an infectious agent (46). Either culture techniques were not sensitive enough to elucidate possible infection or microbial invasion of the amniotic cavity is not the only infection site that triggers preterm birth. The limitations of classic bacteriology are well documented and many microbial species are uncultivable or difficult to culture at this time. Analysis of intrauterine infection by this technique is thought to be underestimating the problem. However, when bacteria have been cultured it was found that intra-amniotic infection was often poly microbial with multiple genera recovered including *Ureaplasma, Mycoplasma, Fusobacterium, Bacteroides, Streptococcus* and *Gardnerella* (47–50).

The 16S ribosomal RNA gene (16S rRNA gene) has been adopted as the candidate gene for detecting and differentiating bacteria. It has allowed the application of new molecular techniques in linking infection and adverse birth outcomes. The gene has an important structural role in prokaryotic ribosomal proteins and so it is both ubiquitous and contains highly conserved regions across all bacteria. The primary structure of the 16S rRNA gene subunit also contains variable and hypervariable regions which have allowed researchers, beginning with Fox et al (51), to differentiate bacteria at various taxonomic ranks based purely on a relatively small section of their genomes. Also, the highly conserved secondary structure allows multiple pairwise alignments between these sequences enabling phylogenetic analyses. 16S rRNA sequencing has turned out to be a powerful method of discerning taxa (52,53), as well as being accurate enough to use in a clinical setting (54).

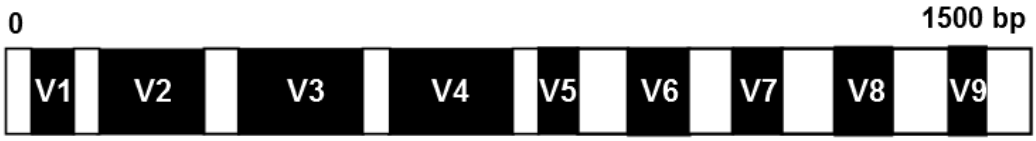

Figure 8-5: Distribution of variable and conserved regions in the 16S rRNA gene.
Variable regions are in black and numbered. Conserved regions are in white.

Studies that compared culture techniques with broad-range primers amplifying the 16S rRNA universally showed an increase in prevalence of bacteria using molecular techniques (55–60). In the context of adverse birth outcomes, the spectrum of genera associated with preterm delivery and the amniotic cavity also increased with these molecular techniques. Sanger sequencing of recovered 16S rDNA amplicon uncovered an increased number of genera recovered from the amniotic fluid of women who delivered preterm, including *Sneathia* spp*.*, *Staphylococcus* spp*.*, *Prevotella* spp*.*, *Peptostreptococcus* spp*.*, *Haemophilus* spp*.* and *Leptotricihia* spp*.* (61).

8.5 Identification of bacteria in the placenta

Much of the focus for intrauterine infection has been on bacteria recovered from amniotic fluid of women who deliver preterm, however it is known that prevalence of bacteria and even the subsequent inflammatory response can be common in term deliveries (62). One study in 2009 by Jones *et al* sequenced the 16S rDNA of bacteria found in the placenta and fetal membranes comparing neonates delivered both term and preterm. It showed

a greater diversity of bacteria in tissues from women who had delivered preterm and increased prevalence of bacteria in those born preterm compared to term. There were also species recovered in the preterm placenta and membranes not found at term such as *Veillonella parvula, Haemophilus influenza, Peptoniphilus lacrimali, Finegoldia magna* and *Enterobacter aerogenes* (63). A later study I carried out using high-throughput sequencing found term deliveries shared many of the same genera, such as *Streptococcus, Microbacterium, Rhodococcus* and *Corynebacterium,* confirmed by two separate primer sets targeting different variable regions of the 16S rRNA gene (64). Six genera (*Fusobacterium, Streptococcus, Mycoplasma, Aerococcus, Gardnerella* and *Ureaplasma)* and one family (*Enterobacteriaceae)* which were either present in greater relative abundances in preterm samples or absent in term deliveries (Figure 8-6). There was a marked difference between the relative levels of *Lactobacillus* in the vaginal deliveries compared to caesarean sections; however this was expected and is easily identifiable. At the species level, *Fusobacterium nucleatum*, *Mycoplasma hominis*, *Aerococcus christensenii*, *Streptococcus anginosus*, *Streptococcus agalactiae*, *Streptococcus spp. (*mitis group) and *Gardnerella vaginalis* were all identified in the placental tissue of women who delivered preterm.

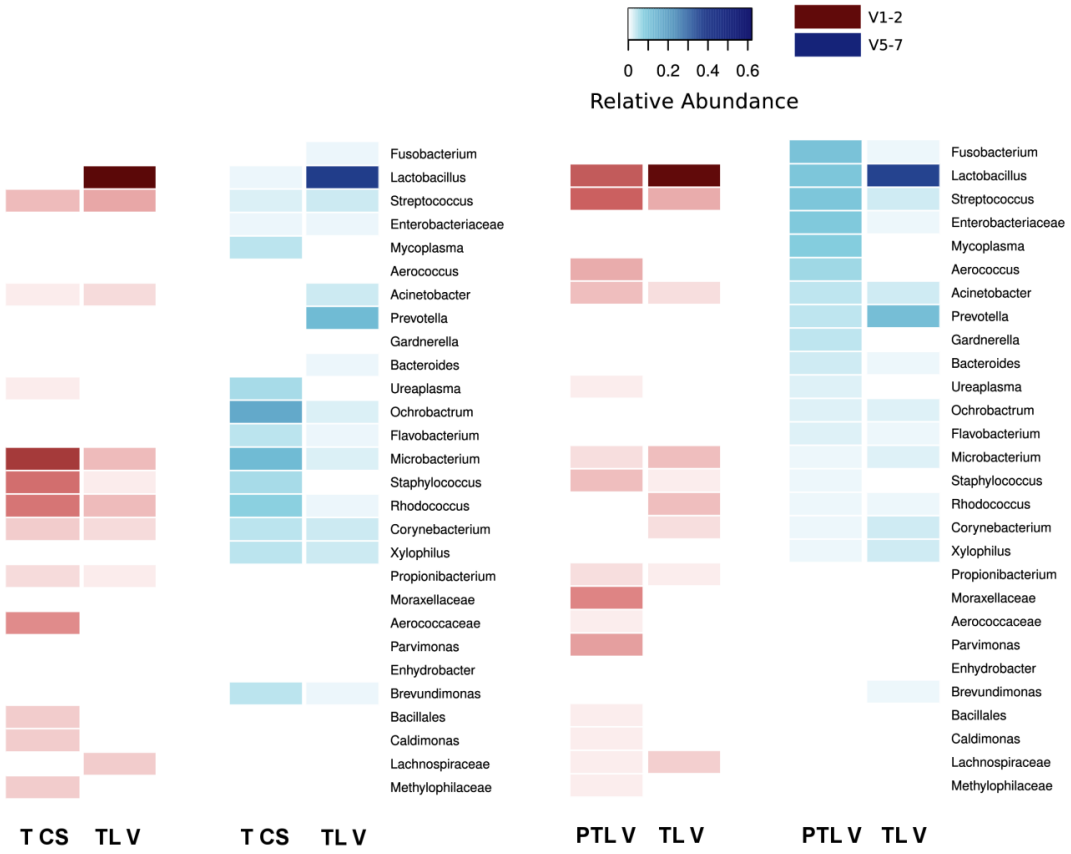

Figure 8-6: Bacteria present in placental membranes from term elective caesarean sections, term vaginal deliveries and preterm caesarean sections.

Relative abundance of each genera are coloured by what 16S rRNA variable regions were used to amplify the sequencing library. The right hand heat map was sorted in descending order by the highest relative abundance genera identified using primers targeting V5-7. T CS: Term delivery by caesarean section. TL V: Term vaginal delivery. PTL V: Preterm vaginal delivery. This figure is the same used by the author in Placenta. 2014 Dec; 35 (12):1099–101.

The study provided further evidence for the role of bacteria in preterm birth. Bacterial DNA is present in the majority of placental membranes whether they are from term or preterm deliveries and irrespective of mode of delivery. However, the relative roles of bacterial diversity, bacterial load and host in the induction of preterm birth still remain to be clarified.

8.6 The source of placental bacteria

Bacteria can colonise the placenta through a number of different routes, including ascending through the cervix from the vagina and haematogenous spread across the placenta. Ascending infection is characterised by a number of distinct stages. Firstly there is a change in the vaginal flora to include microbiota associated with invasion of the intrauterine cavity. This is usually associated with the development of bacterial vaginosis (65,66). There is a link between bacteria commonly found in the vagina and preterm birth (55,66,67). The second stage is the ascension through the cervix and colonisation of the decidua and chorionic membranes. Presence of bacteria on the fetal chorion is associated with thinning and membrane rupture which could lead to foetal infection (68). However, it has also known that bacteria can cross intact fetal membranes as well (69). Bacteria, once inside the amniotic cavity, can be ingested by the foetus through infected amniotic fluid leading to foetal bacteraemia and sepsis.

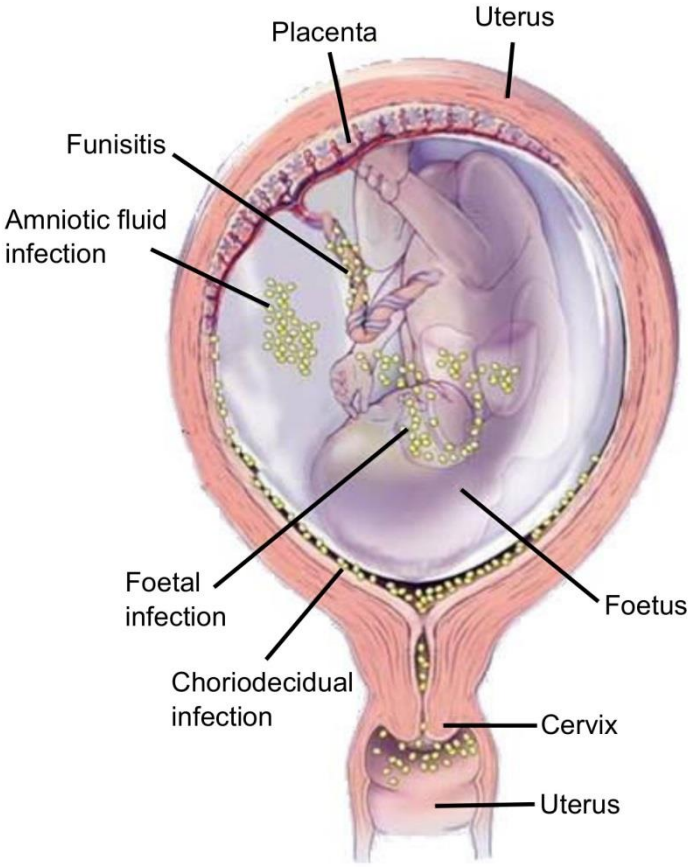

Figure 8-7: Ascending intrauterine infection.
Adapted from Goldenberg *et al* (70).

The classic view that these infectious agents invade and colonise maternal and fetal tissues through an ascending infection from the vagina does not explain the full spectrum of taxonomy recovered. There is increasing evidence for a haematogenous model of transmission to the placenta, notably from the oral cavity (71) and those bacteria can still be recovered from the neonate after delivery (72). A number of mouse model experiments have shown that infections from two organisms associated with periodontal disease, *Porphyromonas gingivalis* and *Campylobacter rectus*, are capable

of causing a number of adverse birth outcomes including low birth weight. If the bacteria reached the murine placenta they could inhibit a number of placental growth factors leading to IUGR (73–76). Using a nested PCR approach one recent study has shown a higher prevalence of *Eikenella corrodens* in placentas of women who had periodontitis and a higher prevalence of *Fusobacterium nucleatum* in women with periodontitis and preterm birth or low birth weight (77). Another study has used primers aimed at the 16S-23S rRNA intergenic transcribed spacer to track a a specific *Fusobacterium nucleatum* strain from a maternal origin to the gastric aspirate of the newborn, providing further evidence for the haematogenous spread of oral bacteria to the newborn (78). A systematic review looking for an association between periodontitis and preterm birth and low-birth weight found a positive correlation in nine out of ten studies (79). A recent study trying to define a common placental microbiome found it resembled the oral microbiome more than any other body site when compared against a large mixed population (80). This included some differences between those placentas recovered from term and preterm deliveries.

The placenta and fetal membranes have been shown not to be sterile sites (62,80) and as techniques have become more sensitive, studies have moved beyond the mere detection of bacteria. Numerous microbial communities exist that have a commensal relationship with the host immune system, as well as others that are pathogenic. Advancements in current technologies for large-scale sequencing platforms allow greater insight and depth into distinct differences between these microbiomes and therefore could provide greater

evidence for bacteria that could be triggering adverse birth outcomes. High-throughput sequencing has many advantages over the conventional Sanger sequencing including a decrease in the cost per sequence, the potential to uncover unlikely pathogens and the depth of coverage to provide a relatively unbiased profile of commensal and pathogenic bacteria (81). Although run time has increased compared to Sanger capillary sequencing, a single Illumina MiSeq run produces ~18 million paired-end reads compared to 0.000096 million reads in a capillary sequencing run. In real terms, this reduces the cost per Megabase from ~£1000/Mb for Sanger sequencing to ~£0.25/Mb for the Illumina MiSeq (82). The umbrella term high-throughput sequencing incorporates a number of different platforms utilising separate methods, however most require short template DNA (200-1000bp) and binding sites for forward and reverse sequencing primers. All current generation sequencers generate millions of sequenced reads in parallel from an input sample allowing the assembly of a draft genome in a single run. However, 16S rDNA high-throughput sequencing studies have widely discussed drawbacks including PCR bias (83) and reagent contamination (84). The lack of standardisation amongst the field also means that even once the data is generated, a variety of the different results can be obtained dependent on what downstream analysis is chosen (85). Still, the use of high-throughput sequencers have already been used successfully in other 16S microbiome studies including antibiotic-associated diarrhoea (86) and bacterial vaginosis (87). It could be that intrauterine infection is also associated with a fluctuating community of bacteria, otherwise termed a dysbioses, rather than the isolation of a single pathogen.

8.7 Aims

The main aim of this project was the characterisation of the range of placental microbiota found in a large cohort of women in rural Malawi and how community composition might associate with birth outcome. To do this I needed to develop and refine a protocol capable of collecting, processing and analysing samples in a low-income environment with greater risk of contamination. A secondary aim of the project was to quantify if the nutritional intervention that the cohort received as part of the clinical trial had an influence on the multiple microbiomes I measured. Additionally, I wanted to define both the oral and vaginal microbiomes in this group to probe possible sources of placental bacteria from other body sites. Lastly, due to the risk of ascending infection I also wanted to study the association between vaginal microbiota and adverse birth outcomes.

9 Chapter 2: General materials and methods

9.1 Study design and enrolment

All subsequent cross-sectional studies in this project were part of a larger clinical trial assessing whether providing Lipid-based Nutrient Supplements (LNS) to mothers during pregnancy and for 6 months postpartum, and to the infant from 6 to 18 months of age, improves child growth to a greater extent than prenatal iron and folic acid (IFA) or multiple micronutrient (MMN) tablets (88).

Participants were enrolled prospectively into the main trial and followed throughout pregnancy, childbirth and beyond. Pregnant women who arrived for antenatal care during the enrolment period were approached for possible inclusion. Inclusion criteria were: ultrasound confirmed pregnancy of no more than 20 completed gestation weeks, residence in the defined catchment area, availability during the period of the study and signed or thumb-printed informed consent. Exclusion criteria were: age less than 15 years, need for frequent medical attention due to a chronic health condition, diagnosed asthma treated with regular medication, severe illness warranting hospital referral, history of allergy towards peanuts, history of anaphylaxis or serious allergic reaction to any substance, requiring emergency medical care, pregnancy complications evident at enrolment visit (moderate to severe oedema, blood Hb concentration < 50 g / l, systolic blood pressure (BP) > 160 mmHg or diastolic BP > 100 mmHg), earlier participation in the iLiNS-

DYAD-M trial (during a previous pregnancy), or concurrent participation in any other clinical trial. All women enrolled as part of the main trial who delivered were included in these analyses.

9.2 Study setting

Enrolment took place in southern Malawi in four separate study sites in the Mangochi district. These included one public district hospital (Mangochi), one semi-private hospital (Malindi) and two public health centres (Lungwena and Namwera). It is estimated that Magochi district hospital serves a semi-urban population of 100,000, while the other sites provide health care to a rural population of around 30,000 people each.

9.3 Collection of birth outcome and baseline data

At enrolment, participants' weight, height and haemoglobin concentration were measured and obstetric history was recorded. Duration of pregnancy was measured using ultrasound. All participants were tested for malarial infection and HIV (unless they were already known to be HIV positive or opted out). At the first home visit to participants 1-2 weeks post-enrolment, information was gathered on demographic, social and economic background. Birth weight was measured as soon as possible after delivery while newborn length and newborn head circumference were measured at the infant's first clinic visit at 1-2 weeks old.

9.4 Sample collection

The placenta was collected after delivery and transferred to a covered sterile container in the hospital, health centre or home (wherever the delivery took place). While in the sterile container awaiting sampling temperatures ranged between 20°C and 40°C. Tissue sampling of the placenta occurred immediately after delivery, unless delivery occurred overnight (in which case the placenta was sampled the following morning), or at home (in which case the container had to be transported to the nearest study clinic first). Two 5 cm x 1 cm pieces of the fetal (chorionic and amniotic) membrane were taken from the edge of the rupture site and two 0.5 cm x 0.5 cm pieces of placental tissue at full thickness were taken from near the umbilical cord insertion. Tissue was cut using individual surgical scissors for each sample. Scissors were autoclaved and wiped with DNA AWAY (Thermo scientific, USA) between uses. One fetal membrane and one placenta sample were placed in separate cryovials. If the sample collection took place in Mangochi district hospital, the cryovials were placed at -80°C. If sample collection took place at an outlying health centre or Malindi hospital, the samples were stored at -20°C for a maximum of two days before being transferred to -80°C storage at Mangochi district hospital. The other two placenta and fetal membrane samples were placed in 10% neutral buffered formalin fixative, processed and embedded in paraffin wax. These were sectioned at 3-5 micron thick and stained with haematoxylin & eosin before being read.

Vaginal mucus samples were taken at a postnatal visit (approx. 1 week after delivery), by a nurse using 4 cotton swabs. She inserted the swabs approximately 7 cm deep into the participant's vagina, without a visual control, and rotated the swab a few times before withdrawing it. After the sample collection, two of these swabs were stored at -80°C for to await DNA extraction.

Dental swabs were collected at Mangochi central site from all mothers who completed the oral health visit at one week after delivery or as soon as possible by specifically trained dental therapists. One sterile plastic swab stick with nylon fiber tip, stored in plain dry tube (microRheologics no. 552, Coban, Brescia, Italy) was used for the sample collection. The dental therapists collected the sample by rubbing the gingival margin of each tooth with the swab. They used dental mirror elevate the cheeks so that the teeth were visible and skin contact was avoided. They started the sample collection from the buccal side of the most posterior (farthest) tooth on the right upper jaw. They repeated the procedure for the palatal sites of the same teeth, continuing then to the lower jaw and repeating it for all lower teeth's buccal and lingual side. They placed the swab immediately to a cold box with ice bricks. The cold box was taken to the laboratory and handed it over to a laboratory technician who removed the swab from the tube and cut the applicator stick with scissors above the fibre tip to fit the swab into cryovials where they were stored. Prior to cutting, they wiped the scissors with disinfectant to avoid contamination. They placed the cyrovial into a -20°Cfreezer and as soon as possible, moved the swab into a -80°C freezer.

9.5 DNA Extraction

DNA extraction was carried out for all sample types using the QIAmp DNA mini kit (Qiagen, Germany) as per the manufacturer's protocol with an additional cell disruption step after lysis with Proteinase K. In the additional step, 0.1mm glass beads (Lysing Matrix B, MP Biomedicals) were added to each sample and the 2ml tubes were shaken on a cell disrupter (Vortex Genie 2, Scientific Industries) for 10 minutes at the highest speed. For every 10 extractions, a negative extraction control was included (200µl buffer AE).

9.6 16S rDNA broad-range qPCR

All DNA samples from placenta and fetal membranes were screened for bacteria using a quantitative PCR (qPCR) SYBR green fluorescent dye assay. The following primer pair targeted the V5-7 regions of the 16S rRNA gene, v785F: 5′-GGATTAGATACCCBRGTAGTC-3′, 1175R: 5′-ACGTCRTCCCCDCCTTCCTC-3′ (64). Each PCR reaction was carried out with the following, 1x Power SYBR Green master mix (Life technologies), 0.4*µ*M of forward and reverse primers, 1µl of template DNA and molecular grade water (Bioline) to give a final volume of 25µl. Amplification took place in an ABI 7300 Real-Time system (Life) under the following conditions: 95°C×10 min, 40 cycles of 95°C×15 sec and 60°C×1 min. Each PCR run included three negative PCR controls (1µl buffer AE from QIAmp DNA mini kit), and a serial dilution of a known amount of positive control from a pure

Escherichia coli culture for quantification used to calculate the bacterial load for each sample. 16S rRNA copy number for each *E. coli* dilution was calculated from the genome molecular weight and DNA concentration as measured by a Qubit 2.0 (Life technologies) and corrected for the seven 16S rRNA copies within the *E. coli* genome. Samples were defined as positive for bacterial DNA if their Ct value was at least 1 Ct lower than the three negative PCR controls (28 Ct ± 3 cycles, depending on variation between runs).

9.7 16S rDNA amplicon high-throughput sequencing

Library preparation was carried out using dual-indexed forward and reverse primers, with barcodes taken from a previous study (89). Both sets of primer sequences can be found in appendix. Each library preparation PCR was carried out with 1X Molzym PCR Buffer, 200 µM dNTPs (Bioline), 0.4 µM forward and reverse primer, 25 mM Moltaq, 5µl template DNA and molecular grade water (Bioline) to give a final reaction volume of 25µl. The reaction was amplified under the following conditions depending on the sample type. Placenta and fetal membrane samples were amplified under the following conditions: 94°C×3 min, 32 cycles of 94°C×30 sec, 60°C×40 sec and 72°C×90 sec, with a final extension cycle of 72°C×10 min. Vaginal and oral swab samples were amplified under the following conditions: 94°C×3 min, 30 cycles of 94°C×30 sec, 60°C×40 sec and 72°C×90 sec, with a final extension cycle of 72°C×10 min. The resulting amplicons were cleaned and pooled using SequalPrep normalization plate kits (Invitrogen) and AMPure XP beads (Beckman Coulter) both as per manufacturer's protocol. Each plate was

pooled into an equimolar final library after quantification using a Qubit 2.0 (Life technologies). Library was loaded onto a MiSeq (Illumina) as per manufacturer's protocol for 500 cycle V2 kits and were multiplexed at either 96 or 384 samples per run. Custom primers were added for read 1 (TACCGGGACTTAGGATTAGATACCCBRGTAGTC), read 2 (AACACGTTTTAACGTCRTCCCCDCCTTCCTC) and index 1 (GAGGAAGGHGGGGAYGACGTTAAAACGTGTT).

9.8 Bioinformatics and statistical analysis

Paired-end 250bp sequenced reads from each MiSeq run were merged using FLASH with the minimum and maximum overlap changed to match the 370bp expected amplicon (90). Merged reads were demultiplexed, pooled and assigned OTUs using QIIME v1.8.0 (91) at 97% similarity against a small custom database of full length 16S rDNA sequences and any sequences that failed to match at 97% were assigned against the full Greengenes database using UCLUST. A representative sequence was picked for each OTU cluster and aligned using PyNAST. Taxonomy information from Greengenes was added to each representative OTU using UCLUST. Parameters for all scripts run through QIIME can be found in the appendix.

Mock communities were sequenced alongside samples and used to filter the dataset for error. Each sample was filtered based on the relative abundance at which the first erroneous OTU appeared in the mock community data for

that specific sequencing run. For most sequencing runs all OTUs occurring ~0.8% relative abundance and below were removed.

Table 9-1: Twenty member mock community control (BEI resources)

Bacterial species	NCBI reference sequence
Acinetobacter baumannii, strain 5377	NC_009085
Actinomyces odontolyticus strain 1A.21	NZ_AAYI02000000
Bacillus cereus strain NRS 248	NC_003909
Bacteroides vulgatus strain ATCC 8482	NC_009614
Clostridium beijerinckii strain NCIMB 8052	NC_009617
Deinococcus radiodurans strain R1 (smooth)	NC_001263, NC_001264
Enterococcus faecalis strain OG1RF	NC_17316
Escherichia coli strain K12 substrain MG1655	NC_000913
Helicobacter pylori strain 26695	NC_000915
Lactobacillus gasseri strain 63 AM	NC_008530
Listeria monocytogenes strain EGDe	NC_003210
Neisseria meningitides strain MC58	NC_003112
Propionibacterium acnes strain KPA171202	NC_006085
Pseudomonas aeruginosa strain PAO1-LAC	NC_002516
Rhodobacter sphaeroides strain ATH 2.4.1	NC_007493, NC_007494
Staphylococcus aureus strain TCH 1516	NC_010079
Staphylococcus epidermidis FDA strain PCI 1200	NC_004461
Streptococcus agalactiae strain 2603 V/R	NC_004116
Streptococcus mutans strain UA159	NC_004350
Streptococcus pneumonia strain TIGR4	NC_003028

Negative controls were sequenced for screening of reagent contamination. SourceTracker (92) was used to predict OTUs in the placental and fetal membrane databases whose sources were the negative controls. These OTUs were removed from further analysis. After filtering, any samples with less than 1000 reads were also removed.

Alpha diversity (number of unique OTUs) and beta diversity (unweighted UniFrac distance) were both calculated as implemented in QIIME after random subsampling without replacement to 1000 reads per sample. Diversity of microbes within an individual's body site was summarised by taking the median value of their intra-individual UniFrac distances with a high value representing a variable community from other individuals and a lower value representing a stable community. The bacterial load of particular phyla, families or species was calculated from participant's relative abundance and their overall bacteria load. Where individual bacterial loads of different species were calculated, 16S rRNA copy number was adjusted for those species accordingly. Heat map comparisons between organism relative abundances were produced using the R package Algorithms and framework for Nonnegative Matrix Factorization (NMF).

Birth weight as measured was used if recorded within 48 h of delivery; if not, birth weight was back calculated from weight measured at 6 or 13 days. If the weight was first measured within 2 to 5 days after delivery, when infants usually lose weight, birth weight was estimated by applying an age-dependent multiplicative factor to the measured weight (93). Length-for-age and head circumference-for-age z scores were calculated using the WHO

Child Growth Standards (94). Statistical analysis was carried out with Stata v13 and R v3.1.0. Adjusted linear regression models were used to measure the change in continuous outcome variables (duration of delivery, birth weight, LAZ and HCZ). Covariates entered into adjusted models were chosen based on previous literature and were entered into the model simultaneously in a one step, forced entry method. For comparisons between OTU abundances and birth outcomes, *q* values were calculated using the Benjamini-Hochberg correction to control for the false discovery rate (FDR).

9.9 Ethical approval and consent

Informed consent was granted from the mother at enrolment. Ethical approval was obtained from College of Medicine Research and Ethics Committee (COMREC), Malawi (Protocol number: P.08/10/972). The trial was registered at clinicaltrials.gov as NCT01239693.

10 Chapter 3: Development of high-throughput 16S rDNA amplicon sequencing method.

10.1 Introduction

An exponential rise in high-throughput sequencing microbiome studies has been possible because of the advances made in wet and dry laboratory techniques. For example, the curation of detailed 16S rDNA sequence databases (95) and algorithms capable of dealing with fast alignment of millions of sequences (96) have allowed accurate recovery of microbial communities from early next generation sequencers (97). Unique short DNA sequences that can be integrated into the 16S rDNA amplicon from specific samples during a PCR step allow the multiplexing of hundreds of samples in a single sequencing run (98). There is enough unique variety in these short “barcode” sequences, especially with dual indexing of the forward and reverse strands, to now allow multiplexing of thousands of samples (89,99). This technique maximises the potential read depth in the latest sequencers to reduce per-sample sequencing costs to their lowest point in history.

There are many challenges facing the interpretation of 16S rRNA gene sequencing results. Primary extraction of genomic DNA can lead to both bias and contamination in later analysis (84,100). The amplicon length studied effects classification (101), diversity of microbes identified and their relative abundance, with smaller amplicons inflating the number of unique community members (102) and the choice of variable region studied has multiple, far-

reaching effects on results. These include underestimation and overestimation of the number of OTUs in a community (103,104) and different taxa recovered depending on region and community analysed (105–107). The type of full length 16S rDNA sequence database used produces different alignment results (85). Chimeric reads produced by PCR generation of the sequencing library can lead to false novel organisms being discovered in the community (108,109). In fact, PCR biases in general have been found to exhibit the greatest effect over community structure (83,110).

Even with these drawbacks, accurate and reliable results can be generated with the use of the right tools and techniques (111,112). These include mechanical disruption of bacterial cells in the extraction and the avoidance of multiple freezing and thawing steps to improve recovery of all taxa (113). Reducing the number of PCR cycles during library preparation to prevent “jackpotting” and stringent quality filtering in the downstream analysis phase both limit the number of erroneous sequences (114,115). Larger, more diverse databases will allow greater breadth of taxonomic classification (116) whereas smaller, curated databases may be better for accurate identification to species level (117). There are also different advantages to using overlapping and non-overlapping paired-end reads (99,118) including uses a complete overlap to check and correct sequencing error, greatly improving quality (119).

There are many bioinformatics tools now available for analysis of 16S amplicon data. The most popular packages combining a number of tools from different sources are currently Quantitative Insights Into Microbial Ecology

(QIIME) (91) and mothur (120). Prior to the initiation of this project there had not been a single study comparing microbial communities in placental tissue. In addition to this there was no consensus in the methods used throughout different microbiome projects. For these reasons I wanted to design and optimise a method that would be ideally suited to addressing the aims of our project. This included the use of a positive control material to provide quality control and standardisation of each method, this has been overlooked in the vast majority of previously published sequencing projects (121,122).

The aim of this chapter was to develop a high-throughput sequencing method for targeting the 16S rRNA gene that identified true constituents compared to contamination. This method would be designed to be able to process thousands of samples easily by a single person and would identify bacteria present at their greatest possible taxonomic depth in the placenta.

10.2 Materials and Methods

10.2.1 Sample collection

For preliminary analysis and method development, nine extra full-thickness placental samples were taken after delivery, in addition to the two samples already taken as part of the study. These samples underwent the same processing and shipment as those stipulated in section 9.4.

10.2.2 DNA Extraction

The nine extra placental samples used in this preliminary work underwent the same extraction protocol as in section 9.5.

10.2.3 16S rDNA amplicon high-throughput sequencing

All PCR and sequencing reactions took place under the same protocols and conditions as stipulated in section 9.7 unless otherwise stated in the results. All gels were loaded with 20µl of amplified product in pre-cast 2% E-gel Agarose SYBR-Safe gels (Invitrogen). All bands on the gel were sized using either 50bp or 100bp DNA ladders (Invitrogen). For results analysing potential contamination of PCR reagents, a single Sanger sequencing run was performed on PCR amplicon using the Big-Dye 3.1 Cycle-sequencing kit (Thermo Fisher scientific) as per manufacturer's instructions and was analysed on a 3130 Genetic Analyser (Thermo Fisher scientific).

10.2.4 Bioinformatics

Primer prospector was used to perform *in silico* comparisons between primer pairs. *Analyze_primers.py, get_amplicons_and_reads.py* and *taxa_assignment_report.py* were used to score primers against the full 16S Greengenes database, generate the amplicon that would be produced and test the possible taxonomic depth recoverable from the amplicon using the RDP classifier. The script *check_primer_barcode_dimers.py* was used to check and filter possible causes of primer dimer after MiSeq compatible 785F/1175R library preparation primers were designed. Downstream analysis and quality filtering of resulting FASTQ files generated from the MiSeq was carried out under the same protocol as stipulated in section 9.8.

10.3 Results

10.3.1 Comparison of 16S rRNA variable region targets for optimum taxonomic coverage

In order to build a representative microbiome from the placenta, fetal membrane, oral and vaginal samples collected in this study I compared three different pairs of broad-range 16S rDNA primers that targeted different variable regions of the 16S rRNA gene. Three potential pairs were compared *in silico* against a large database of full length 16S rDNA gene sequences using primer prospector (123). Two of the primer pairs chosen were 785F/1175R and 13F/343R, both of which are used routinely by the clinical microbiology service at Great Ormond Street Hospital. The third primer pair, 515F/806R, is used regularly in microbiome studies and a large number of individually barcoded primers have already been designed (89). Amplicons were built from successful hits against the entirety of the Greengenes database for each pair of primer sequences entered *in silico*. A scoring system based on gaps, 3' and non-3' mismatches was used to decide whether the primers would successfully generate an amplicon against a target 16S rDNA database sequence. It allowed for the modelling of the potential success or failure of using a specific primer without the costly process of trial and error on a high-throughput sequencer. Figure 10-1 showed, as expected, that all primers lost coverage as they attempted to increase the depth of classification. Both 785F/1175R and 515F/806R generated a taxonomic classification for a similar number of reads in the

database. Primer pair 785F/1175R classified more sequences at phylum, class and order level compared to 515F/806R but this difference was not seen at the family level. 8F/3R was unable to classify a large proportion of taxa within the Greengenes database and the *in silico* distribution of the length of amplicons generated was quite erratic with no clear peak as seen in the other primer pairs. As both 785F/1175R and 515F/806R showed similar high coverage they were chosen for further optimisation using placental samples on the MiSeq.

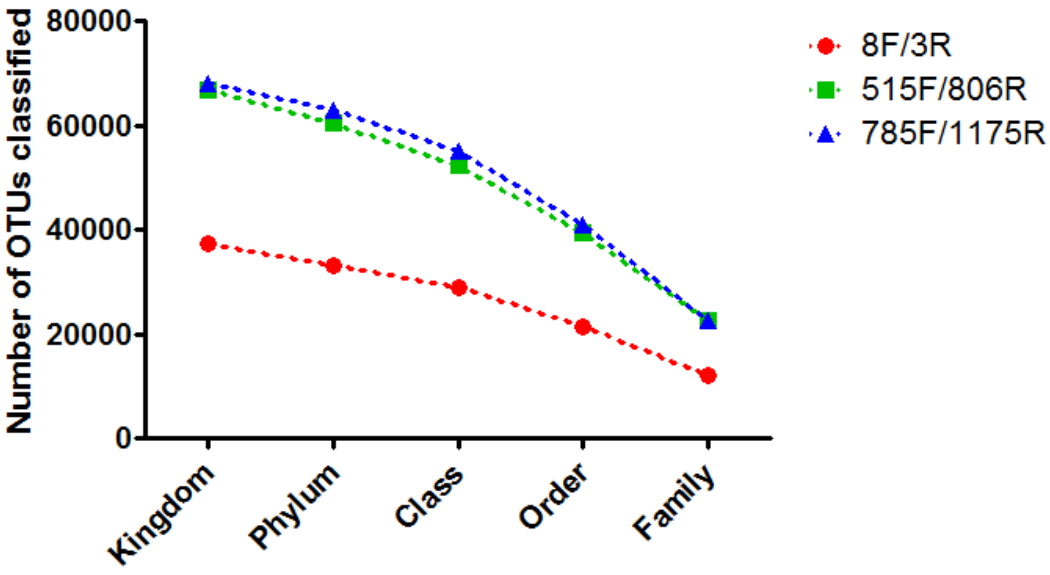

Figure 10-1: *In silico* comparison of taxonomic coverage.
Taxonomic coverage and classification depth predicated against the Greengenes database for three different potential primer pairs. Successful primer hits against the database with low numbers of mismatches were modelled using the Primer Prospector pipeline. Successful amplicons built from the entire database were then assigned taxonomy, then number of successfully assigned amplicons or OTUs are displayed on the y-axis.

Custom sequencing primers were designed with the addition of Illumina adapters, unique barcode sequences and an extra padding sequencing to lessen the likelihood of possible primer dimer. Possibility of secondary structure formation and primer dimer between the primers and the addition of the unique barcodes was checked *in silico* using primer prospector's *check_primer_barcode_dimers.py* script.

After library preparation, I compared the efficiency of the primers in amplifying a serial dilution of a known concentration of *Escherichia coli*. Reactions for both primer pairs were set up using the same input volume of template (5µl), the same number of PCR cycles (32 cycles) and the same unique barcode sequence (rcbc0). The annealing temperature for primers 515F/806R (50°C) differed from 785F/1175R (60°C) as this is what was used previously by Caporaso et al (124). As Figure 10-2 shows, primers 785F/1175R generated a higher concentration of amplified product when compared to 515F/806R.

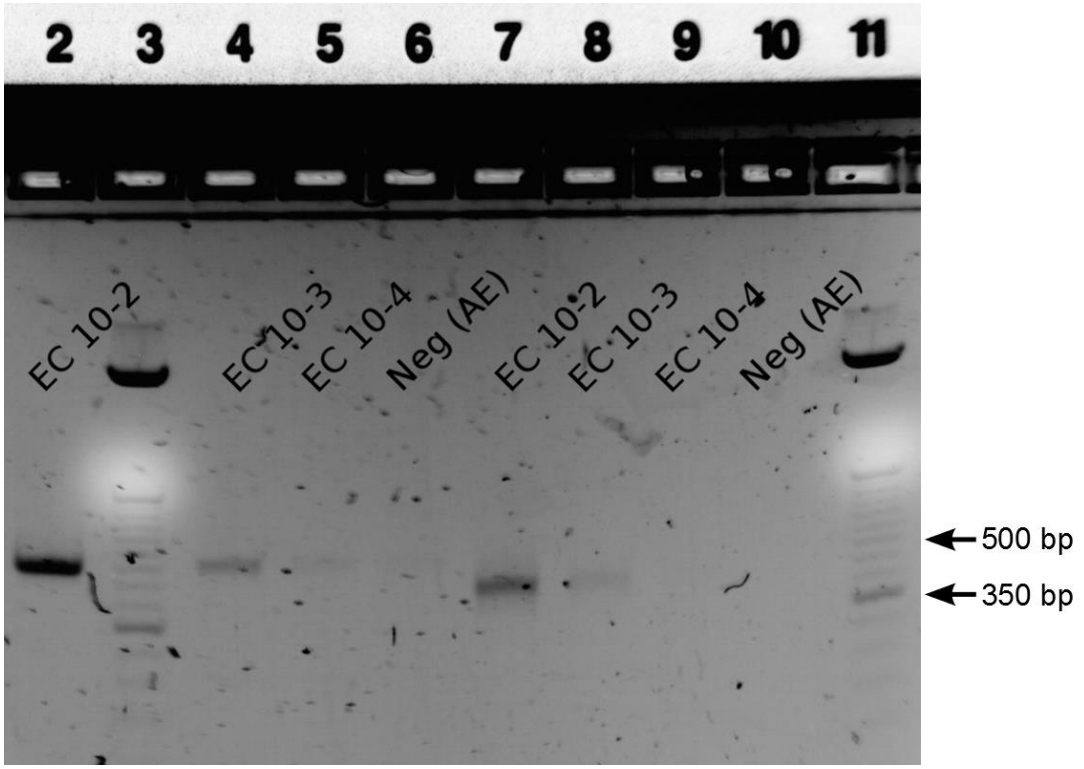

Figure 10-2: Semi-quantitative comparison of product concentration after library preparation between 515F/806R and 785F/1175R primer pairs.
Both sets of primer pairs amplified the same input concentrations of *E. coli* using 5µl of template and 32 cycles of PCR. 20µl of amplified library was visualised on 2% agarose SYBR Safe E-Gels (Invitrogen).

Nine placental samples were sequenced on separate MiSeq runs using both the custom designed 785F/1175R primers and Caporaso et al designed 515F/806R primers. After quality filtering and demultiplexing, the 785F/1175R primers generated 5,834,379 sequences with a median length of 369bp. The 515F/806R primers generated 5,896,952 sequences with a median length of 242bp. Analysis by the QIIME pipeline was used to see whether assembling longer reads allowed for greater confidence in assigning genus and species classifications. The reads generated by the primer pair 785F/1175R generated reads which both had more unique OTUs than 515F/806R and a

higher percentage of reads classified to both genus (77% compared to 68%) and species (31% compared to 23%) level (Figure 10-3)

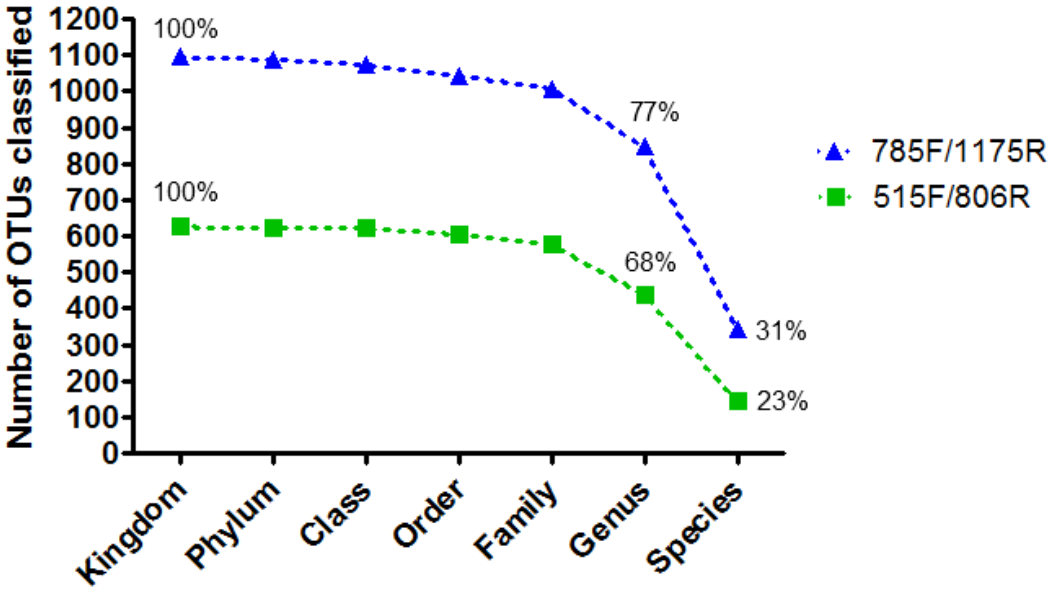

Figure 10-3: Comparison of OTU richness and taxonomic depth between two primer pairs.
The absolute numbers of OTUs that could be confidently assigned to each classification are listed on the y-axis for each of the separate MiSeq runs. Each of the values at Kingdom classification represents 100% of the OTUs identified by that particular primer pair.

10.3.2 Development of library preparation protocol for custom Illumina MiSeq runs

To optimise library preparation for broad-range 16S rDNA specific amplicon I compared *Taq* polymerase, cycle number and reaction template volume. In order to try and reduce error within the PCR reaction I compared high fidelity polymerase (KAPA Biosystems) against an ultra-clean polymerase (Moltaq, Molzym). I compared both *Taq* polymerases using the same PCR conditions

for both as stipulated in section 9.5. The KAPA Taq and buffer were loaded as per manufacturer's instructions. The library after PCR was visualised on a 2% SYBR Safe agarose gel (Invitrogen). Figure 10-4 shows that negative controls in lanes 6 and 7 amplified using KAPA Hifi Taq, containing 5µl PCR-grade water (Bioline), generated a band at the same size as the expected amplicon and matched band amplified in the *E. coli* controls in lanes 2, 3 and 4. This band was missing in the negative controls amplified with the ultra-clean Moltaq polymerase.

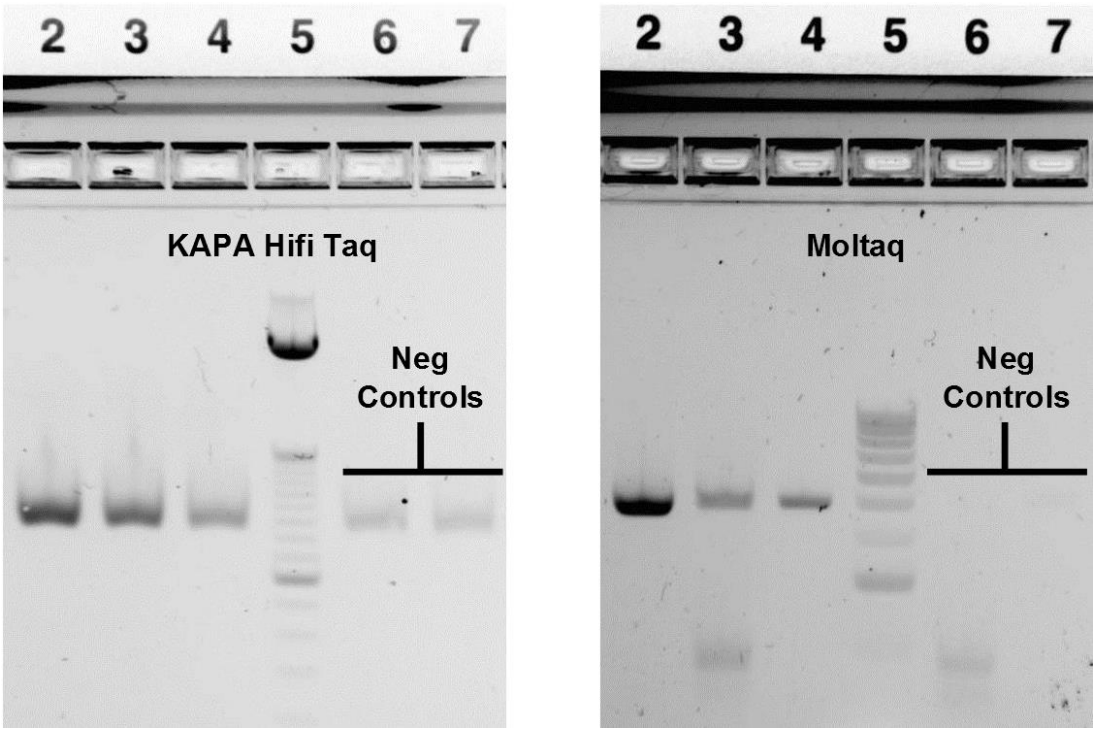

Figure 10-4: Bacterial contamination in Taq polymerase.
20µl of amplified library was visualised on 2% agarose SYBR Safe E-Gels (Invitrogen). Lanes 2, 3 and 4 in both gels contain a serial dilution of a known concentration of *E. coli*. Negative controls in lanes 6 and 7 contained 5µl of PCR-grade water. KAPA Hifi Taq lane 5: 50bp ladder and on the Moltaq gel lane 5: 100bp ladder. 16S rDNA amplicons in both gels appeared at ~500bp.

To reduce the number of errors introduced into the dataset during the PCR, I reduced the number of PCR cycles. A reduction in the number of cycles from 36 to 34, following the addition of 2µl template into a 25µl reaction, resulted in a 10-fold reduction in the sensitivity of the PCR (Figure 10-5).

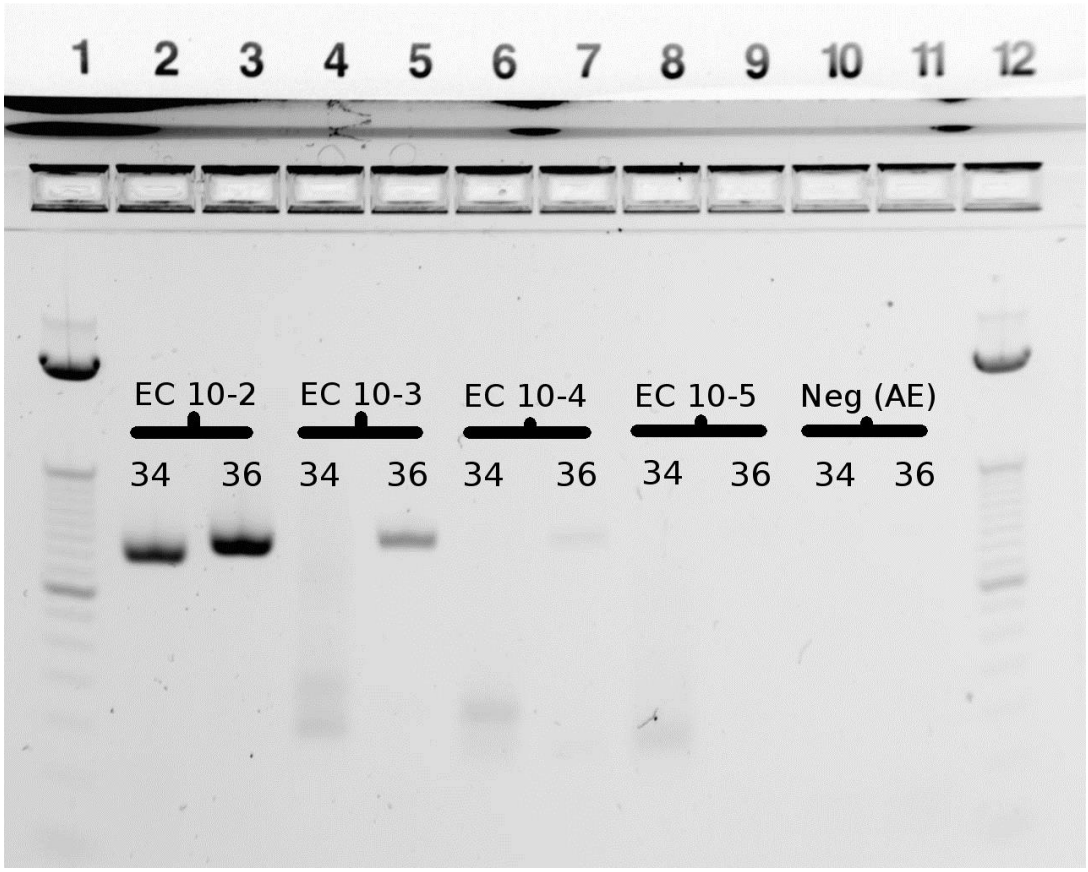

Figure 10-5: Comparison of lowering cycle number on sequencing library generation.
20µl of amplified library was visualised on 2% agarose SYBR Safe E-Gels. EC = positive control *E. coli* DNA. 34 or 36 = 34 cycles or 36 cycles of PCR respectively. Expected amplicon size was observed at ~500bp on a 50bp ladder.

However, by increasing the volume of template added to the reaction I was able to reduce the number of PCR cycles whilst retaining the original sensitivity of the PCR. Figure 10-6 shows in lanes 11 and 12 of the first gel, that successful amplification of *E. coli* was possible once volume of template added to the reaction was increased to 5µl using 30 PCR cycles. However when amplifying the lowest diluted concentration in *E. coli* (right hand gel, lane 2) only elucidated a very weak band compared to when amplified by 32 cycles (left hand gel, lane 7). When 2µl of *E. coli* template was added and amplified for 30 cycles, no band could be seen on the gel (left hand gel, lanes 8, 9 and 10).

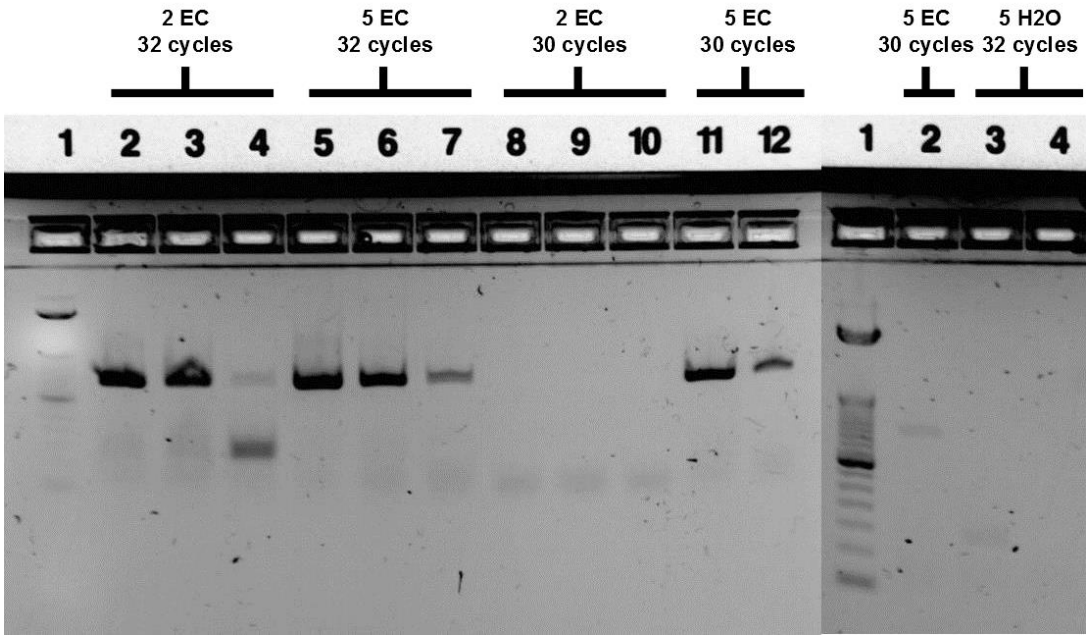

Figure 10-6: Comparison of template volume and cycle number on library preparation.

20µl of amplified library was visualised on 2% agarose SYBR Safe E-Gels. 2 or 5 EC = Either 2µl or 5µl of 1/10 serial dilutions of *E. coli* DNA for each combination of loading volume and cycle number. From left to right: EC 10^{-2}, EC 10^{-3} and EC 10^{-4}. 30 or 32 = 30

cycles or 32 cycles of PCR respectively. Expected amplicon size was ~500bp on a 50bp ladder.

10.3.3 Bacterial contamination of oligonucleotide primers when sequencing low biomass samples

After preliminary analysis of the first 384 samples sequenced I found that the predominate organism recovered from the high-throughput sequencing data was an unknown *Sphingomonas* species (Figure 10-7).

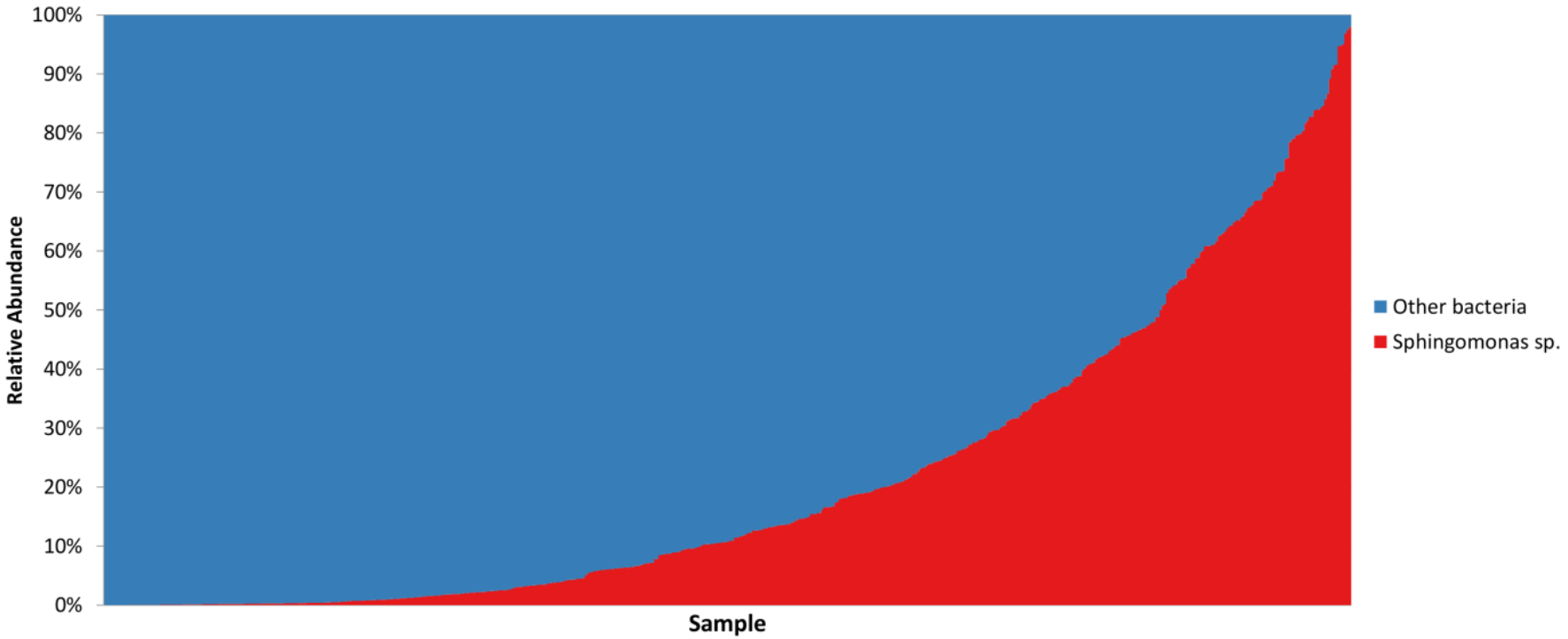

Figure 10-7: Relative abundance of a *Sphingomonas* sp. across 384 placental tissue samples.

I tested for the possibility of reagent contamination and when I used the sequencing primers as template in the library preparation reaction I found that I could detect amplicon at the same size as our expected 16S rDNA product (Figure 10-8Figure 10-8). Sanger sequencing of the PCR product produced was used to confirm whether it was the same organism identified after high-throughput sequencing on the MiSeq and a BLAST search of the Sanger sequencing product also matched it to an uncultured *Sphingomonas* sp. New lyophilised oligonucleotide primers from a different source were also tested for the same contamination and were found to be free of detectable bacterial DNA. These primers were taken forward to be used in all future library preparations.

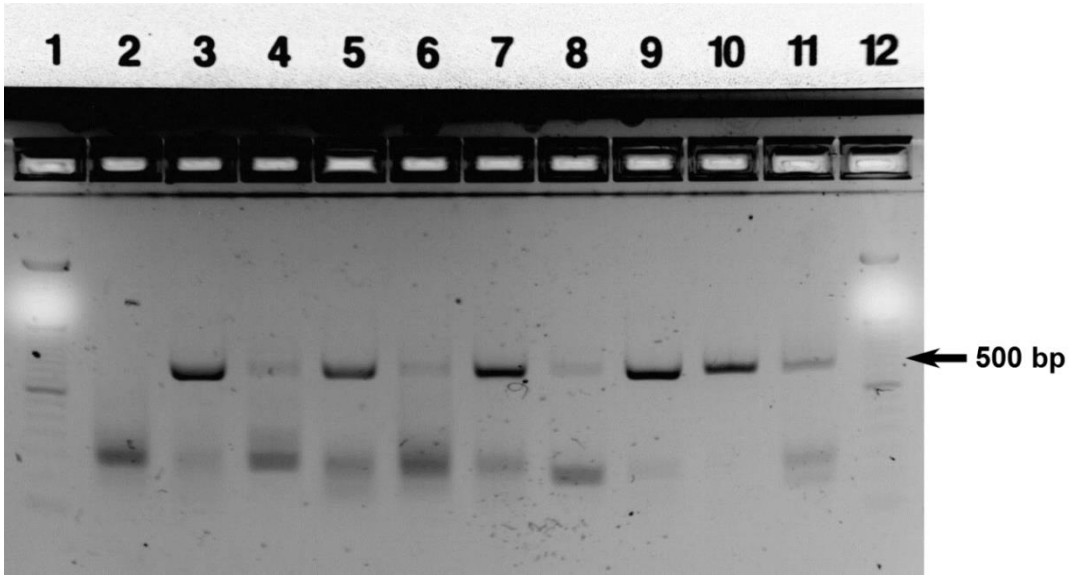

Figure 10-8: Amplification of bacterial DNA contamination found in diluted oligonucleotides.

20µl of amplified library was visualised on 2% agarose. Lane 2 contained amplified product from 5µl PCR grade water and lane 3 contained amplified product from 5µl *E.coli* DNA. Lanes 4-11 contained amplified product from 5µl of an randomly selected individual oligonucleotide used for library preparation. Expected amplicon size was ~500bp on a 50bp ladder.

10.3.4 Development of a concise custom 16S database and rigorous filtering recreates artificial mock communities

To validate the specificity and sensitivity of my 16S rDNA high-throughput sequencing wet lab and dry lab protocols, I sequenced mixtures of known concentrations of genomic DNA from a number of different bacteria in an attempt to describe the exact microbial community present. I used a ten member mock community kindly provided by the Jim Hugget at LGC, Teddington, UK (Table 10-1).

Table 10-1: Ten member mixed bacterial mock community with known relative abundances

Bacterial species	16S rRNA gene copy number adjusted relative abundance (%)
Neisseria meningitides	28.60
Streptococcus pneumoniae	25.83
Klebsiella pneumonia	20.24
Staphylococcus aureus	12.74
Streptococcus pyogenes	7.26
Streptococcus agalactiae	2.72
Escherichia coli	1.50
Enterococcus faecalis	0.67
Pseudmonas aeruiginosa	0.36
Acinetobacter baumanii	0.07

I sequenced the mock community in triplicate on a single MiSeq run and after quality filtering I obtained 244,009, 21,671 and 19,705 high quality reads in each sample. After picking Operational Taxonomic Units at 97% sequence identity against either the Greengenes database or SILVA database, I found a large amount of read artefacts and erroneous assignment of taxonomy leading to an inflation of richness and phylogenetic diversity as sequencing depth increased. Figure 10-9 shows even though there was only a single *Neisseria* species in the sample, read error contributed to five separate OTU bins being assigned *Neisseria* or *Neisseriaceae* taxonomy. This included one OTU that clustered with *Streptococcus spp.* rather than in the *Neisseria* clade. Similar separations of clades due to error were also seen in both OTUs assigned to either *Streptococcus* or *Staphylococcus* genera. I also found at least one example of an erroneous sequence included in the Greengenes database that led to a sequence that was later shown to have 99% similarity to *Streptococcus pneumoniae* being clustered as a separate OTU identified as an uncultured member of the *Caulobacteraceae* family.

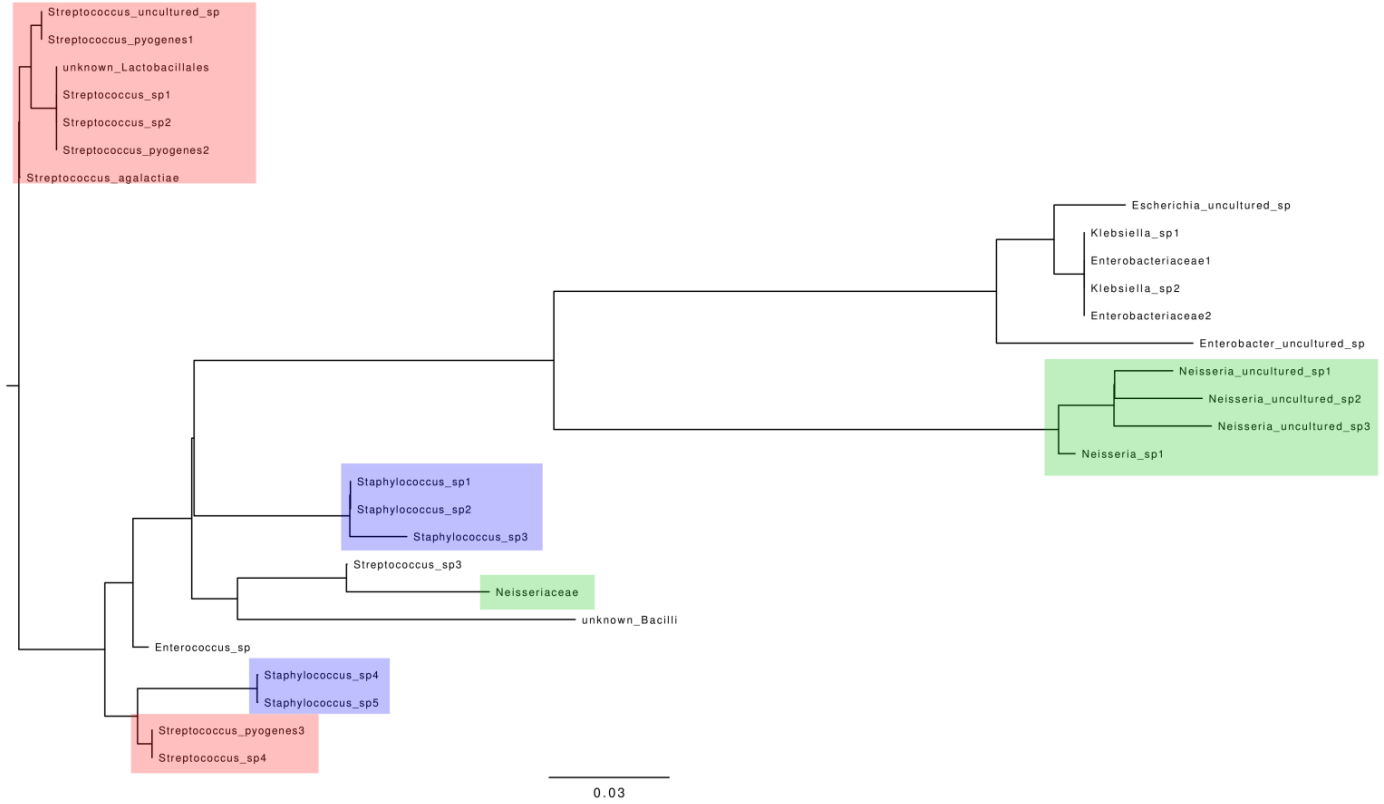

Figure 10-9: Phylogenetic tree of the most abundant 28 OTUs after sequencing a known mock community.
Phylogenetic analyses of ~370 bp OTUs with taxonomy assigned by RDP classifier. Colours highlight similar taxa that have been clustered on separate branches based on nucleotide differences.

I filtered the SILVA SSU Ref NR 111 database from 597,607 full length 16S rDNA sequences to 5,493 sequences. I kept only high quality sequences that had species identification and were from phyla previously found in our dataset after taxonomy assignment against the full Greengenes database. In most cases I limited the database to one full length sequence per species. When assigning global OTUs, I would first pick against the smaller, curated custom database and would pass any sequences that failed to match at 97% similarity through a second round of OTU picking against the full Greengenes database.

An unfiltered dataset with OTUs assigned through the default workflow in QIIME found 166 unique OTUs in the 20 species mixed mock community (Table 10-2). This included all the known community members as well as 69 OTUs whose taxa did not match any in the mock community. The most inflated OTU counts in actual mock community genera were *Streptococcus* (26 unique OTUs compared to 3 species) and *Staphylococcus* (24 OTUs compared to 1 species). Using the two rounds of OTU picking against our custom database and modest filtering of the dataset, removing all OTUs found at 0.07% abundance in the sample and below, I removed a vast majority of spurious OTUs (Table 10-2). The value 0.07% was chosen as it was the most abundant OTU from a taxonomic lineage that I knew was not present in the mock community.

Using the same OTU table, I performed a further rigorous filtering step that removed the erroneous *Neisseria*, *Streptococcus* and *Staphylococcus* OTUs identified after multiple pairwise alignments in Figure 10-9. This led to all

OTUs present at 1% relative abundance and below in the sample removed. While this recreated an ideal OTU-to-species ratio for many of the mock community members it also meant three of the lower abundance organisms were filtered out from the dataset entirely (Table 10-2).

Table 10-2: Comparison of database and filtering effects on mock community members

	Unfiltered, Greengenes database	Contaminating bacteria filtered (0.07%), custom database	Major alignment errors filtered (1%), custom database
Overall number of OTUs	166	29	13
Neisseria meningitides	8	2	2
Streptococcus pneumoniae	21	6	1
Klebsiella pneumonia	8	6	4
Staphylococcus aureus	24	6	2
Streptococcus pyogenes	4	2	2
Streptococcus agalactiae	1	1	1
Escherichia coli	6	1	1
Enterococcus faecalis	14	2	0
Pseudmonas aeruiginosa	7	1	0
Acinetobacter baumanii	5	1	0
Other	68	0	0

10.4 Discussion

A successful method for the characterisation of microbiota across multiple body sites has been designed and optimised and can now be applied to the samples collected in this study. Likely sources of reagent contamination have been identified and removed and methods formulated to filter out external PCR and sequencing errors from the dataset.

In this study I compared three primer sets that covered 16S rRNA gene hypervariable regions V1-V2 (8F/3R), V4 (515F/806R) and V5-V7 (785F/1175R). Much of the variance seen in a microbiome study can be due to the particular human, animal or environmental site studied as well as the technology employed to capture, amplify and sequence DNA. However choosing which of the target variable regions of the 16S rDNA gene to amplify is an important contributing factor (105). Liu *et al* reported that primers covering the V2 and V3 regions performed much better than V6 in taxonomic assignment and community clustering (125). Although I compared similar variable regions, in this study, I merged larger amplicons from overlapping paired-end reads, encompassing multiple variable regions, which could explain why the V5-V7 amplicon in this study yielded the best results. Soergel *et al* proposed a primer pair amplifying a smaller read length from the V4 provided similar coverage to longer reads from other areas of the gene for comparing shifts at the phylum level (105). However phylum level shifts are not always useful and species level identification is ideal in many studies. The microbiome of placental and fetal tissue does not have the same static, complex microbiome found in other sites, for example the gut and so

large scale comparisons of shift in bacteria phyla is unlikely to be as useful in determining underlying problems. *In silico* comparisons have been used to evaluate primer pairs (106,107) and although this kind of computational modelling can identify obvious shortcomings of certain methods, it can be difficult to translate these findings to those of the more complex representative human body sites. There are limitations in every broad-range primer pair chosen and so screening of potential candidates for your study question against “real world” examples is a key advantage (106). I found primers covering the V5-V7 hypervariable region could easily be integrated into a workflow on the MiSeq platform and provided better coverage for our specific samples when compared to the two other primer pairs in this study.

Reagent contamination is now beginning to be recognised as a large confounding factor in microbiome studies (126). In a study by Salter et al, they showed that as the bacterial biomass used as input for your library preparation reduced, the relative amount and diversity of reagent contamination increased (84). In their example they used serial dilutions of known quantities of *Salmonella bongori* to demonstrate this, however I show similar results when sequencing 16S rDNA from placental tissue. The amount of bacterial DNA extracted from the placenta can vary greatly which leaves ample opportunity in some samples to amplify 16S rDNA contamination found in laboratory reagents. In our case it was widespread presence of a *Sphingomonas sp.* in the diluted forward and reverse oligonucleotide primers was the most significant contamination faced. This is due to the fact that primers can be shipped with contaminating bacterial DNA

due to improper sterilisation during the manufacturing process. There have been other published cases of *Sphingomonas spp.* contamination in 16S rDNA studies (127,128) probably because this organism is a recognised as a contaminant of ultra-pure water systems in the manufacturing industry (129). Due to the lower levels of contamination observed in lyophilised primers supplied by the same company it is likely that water used during oligonucleotide synthesis was the source.

The number of PCR cycles has long been recognised as a source of bias and can lead to inflation of unique 16S rDNA sequences in bacterial studies. In a comparison between 35 and 18 amplification cycles, an increase in the percentage of chimeric sequences was seen at 35 cycles (130). A later similar study found 30% more chimeric samples within their dataset when comparing amplification at 30 cycles to 15 cycles (114). Unfortunately, reduction of PCR cycle also leads to reduction in the amount of PCR product recovered. In this study I have tried to strike a balance between lowering PCR cycles but still keeping sensitivity high enough to identify genuine bacterial colonisation of placental tissue. Instead I have focused on the use of ultra-pure reagents (Moltaq) and a number of methods to try and reduce the numbers of erroneous reads post-sequencing, including clustering OTUs and extensive quality filtering (115,131).

Other considerations to take into account when trying to improve the accuracy and reproducibility of high-throughput sequencing data are the quality of the reads produced and the reference database used. Generally the MiSeq can produce millions of high quality reads but the accuracy of

Illumina reads increases from 85% to 99.65% when overlapping paired-end reads are used instead of single-end (132). There are also a host of documented advantages using the current Greengenes database as a training set due to it being largest available repository of curated full length 16S rDNA gene sequences. Trimming this reference database to match a specific amplicon length has also been shown to increase confidence in taxonomic assignment (116). I have chosen to pursue an additional strategy that is the curation of a site-specific reference database, which has been previously described for a set of oral microbiome data (133). It is common for 16S databases to have large numbers of unidentified and poorly annotated sequences and this leads to poor taxonomic assignment. Instead, a site-specific 16S rDNA database can be curated and updated and as more samples are sequenced, the database becomes more accurate, as previously shown by the Vaginal Human Microbiome Project (117).

The sequencing technique has been designed and optimised and can now be applied to the study samples using larger multiplexed MiSeq runs. Jackknifed bootstrapping analysis of the data already collected showed that 1,000 sequences per sample provided enough depth to recover all the diversity in the sample. This means there is the potential to multiplex 500 samples on a single MiSeq run and still be confident the results are accurate and reproducible. Also, as improvements to sequencing methods and downstream processing of data are regularly published (119,134) they should lead to even more reproducible and reliable results from microbiome studies.

11 Chapter 4: Effect of the Lipid-based nutritional supplement on placental, oral and vaginal microbiota.

11.1 Introduction

Nutritional status is a key health determinant, especially during pregnancy. The prevalence of women with low BMI in Africa is still 10% higher than in high-income countries (9) and maternal nutritional deficits such as anaemia, vitamin A deficiency or zinc deficiency during pregnancy can lead to higher incidence of maternal and newborn morbidity and mortality (135–137). Southern Africa has the second highest prevalence of low birth weight and among the highest proportions of stunted, wasted and underweight mothers and neonates (9). Micronutrient supplementation has been used in the attempt to supplement poor diet in low-income countries during pregnancy to improve duration of pregnancy and newborn size. A meta-analysis of these studies found that multiple micronutrients (MMN) had a small effect on birth weight but no effect on duration of pregnancy compared to iron and folic acid (IFA) supplementation (94).

The iLiNS-DYAD-M trial set about to improve maternal and child health in rural Malawi with the addition of Lipid-based nutrient supplement (LNS) to a pregnant mother's and newborn child's diet. A previous systematic review found that additional protein and energy, as well as micronutrients, could reduce intrauterine growth restriction (138). The only previous LNS trial

before iLiNS-DYAD-M that studied the impact of LNS during pregnancy found that it increased mean birth length compared to MMN supplementation (139). iLiNS-DYAD-M trial results were published in 2015 and found that the provision of LNS did not have any statistically significant effect on low birth weight, newborn stunting or newborn small head circumference (88). Despite these results, the change to the diet of these participants could still affect predictors and outcomes used in this thesis.

The effect of nutrition and diet on the microbiome is well established. Microbial composition of the gut has now been associated with obesity, with reduced species richness and a change in phenotype that increases energy extraction contributing to weight gain (140,141). Gut microbiota have also been linked to the widespread problem of undernutrition (142) and in malnourished children there are consistent, long-term changes in community structure (143). A study in Kenya in 2014 even found that a micronutrient supplement, specifically iron, could be adversely altering infant's gut microbiota (144). Due to the action of malnutrition on the immune system (145), it is also plausible that diet and nutritional supplementation could have an effect on other human microbiomes. The vaginal microbiome has been shown to be altered by increased fat, folate, vitamin E and calcium intake that can increase the prevalence of pathogenic bacteria associated with bacterial vaginosis (146). In the mouth, increased sugar consumption has long been linked to increased prevalence of dental caries and enamel erosion (147). In low income, tropical countries where malnutrition and other infections are rife, this can affect the efficiency of saliva as an immune barrier, substantially

changing the microbiota found in the mouth and possibly causing acute necrotising gingivitis (148). One recent study has looked at the specific action of a diet supplement on a rural sub-Saharan African population but found while there were changes in the gut microbiota, there was little difference found in both the oral and vaginal microbiomes (149).

I wanted to study the potential impact the nutritional intervention an individual received on their multiple microbiomes. I wanted to quantify any effects so they could be taken into account later when comparing the possible role of the microbiome has in association with birth outcomes. Secondly, if there was an effect, it would be possible to study the possibility of using nutritional interventions in positively altering microbial community structures in the placenta, mouth or vagina.

11.2 Materials and Methods

11.2.1 Study design and enrolment

See section 9.1. Full details on the design, nutrient and energy contents of each dietary supplement in the trial can be found in the publication of the primary outcomes (88).

11.2.2 Study setting

See section 9.2

11.2.3 Collection of birth outcome and baseline data

See section 9.3

11.2.4 Sample collection

See section 9.4

11.2.5 DNA Extraction

See section 9.5

11.2.6 16S rDNA broad-range qPCR

See section 9.6

11.2.7 16S rDNA amplicon high-throughput sequencing

See section 9.7

11.2.8 Bioinformatics and statistical analysis

Downstream processing of all FASTQ reads collected from placenta, fetal membrane, vaginal and oral samples was carried out as stipulated in section 9.8.

Three separate variables were used to measure changes in microbiota found in the placenta and fetal membranes. Prevalence of bacteria in the placenta and fetal membrane was defined as detectable presence of 16S rDNA in a sample of placenta or fetal membrane tissue from a participant. I also analysed the amount of 16S rDNA present in the tissue as representative of the bacterial load. To represent how bacterial community structure changes between the different arms of the study I calculated the median intra-individual unweighted UniFrac distance for each individual from pairwise comparisons between all individuals. A higher median UniFrac distance represented a different subset of microbiota in that individual compared to all others. The lower the UniFrac distance, the more similar that participant's microbiota was to other participants. To represent bacterial species richness I

calculated the number of unique OTUs present for each intervention arm in each site. For the oral and vaginal microbiome I only reported UniFrac distances and observed number of OTUs.

I carried out the statistical analysis with Stata 13 (StataCorp, College Station, USA). I conducted the statistical analysis according to the analysis plan written and published before the intervention code was opened. I based the analysis on the principle of intention-to-treat. The intervention code was broken after the analysis. The global null hypothesis was tested between intervention groups using Fishers exact test for binary outcomes and one-way ANOVA for continuous outcomes. If either test was significant, pairwise comparisons were then carried out between all intervention groups. All analyses were also adjusted using multivariate models to correct for possible confounding effects, all variables picked for testing in the models were chosen based on previous literature as logically capable of forming an independent interaction with the outcome being measured. All covariates were then entered in a single step into the equation. Continuous outcomes were modelled using linear regression and binary outcomes were modelled using log-binomial regression. I tested variables predefined in the analysis plan as effect modifiers using the likelihood ratio tests for interaction between the intervention and maternal characteristics. Maternal variables tested were age, BMI at enrolment, number of previous pregnancies, anaemia, HIV and malaria status at enrolment, education (completed years), socio-economic status and study site of enrolment. Stratified analysis was then produced for any effect modifiers that were statistically significant.

11.3 Results

11.3.1 Sample collection and baseline characteristics by intervention group

I did not have a predefined hypothesis concerning the effect of LNS on participant's resident microbiomes at inception. However, I was still interested in any potential impact on the microbiome and postulated that LNS might change the prevalence and amount of bacteria recovered from the placenta and fetal membranes and the diversity of bacteria found in the oral cavity and vagina.

Of the 1391 participants recruited into the iLiNS-DYAD-M trial, a sample of placenta was collected at delivery from 1030 (73.3 %) participants and a sample from the fetal membranes was collected from 1095 (78.7 %) participants. For DNA analysis, after excluding twin deliveries, this left 1018 (73.2 %) participants with a placenta sample analysed, with 347, 335 and 336 in the IFA, MMN and LNS groups respectively. Loss to follow up was similar in all of the supplement groups (p=0.535). It also left 1083 (77.9 %) participants with fetal membrane samples analysed with a breakdown of 358, 364 and 361 in the IFA, MMN and LNS groups respectively. Loss to follow up was similar in all of the supplement groups (p=0.811). An oral and vaginal swab was also collected from each participant one week after delivery. After excluding twin deliveries and any samples that did not produce enough reads after sequencing, 1104 (79.4 %) oral swabs and 1107 (79.5 %) vaginal swabs were included in the analysis. Of the 1104 participants with an oral

swab collected, 374 were in the IFA, 374 in MMN and 356 in LNS groups. Loss to follow up was similar in all of the supplement groups (p=0.531). For participants with a vaginal swab collected there were 370, 371 and 366 in the IFA, MMN and LNS groups respectively. Loss to follow up was similar in all of the supplement groups (p=0.952).

Across study arms, all participants with a vaginal swab sample taken had similar health status, demographic and socioeconomic characteristics at enrolment (p>0.05) (Table 11-1).

Table 11-1: Baseline characteristics of participants at enrolment by intervention group

Characteristic	IFA	MMN	LNS	*P* value[b]
Number of participants (n=1107[a])	370	371	366	-
Mean (SD) BMI, kg/m²	22.1 (2.6)	22.1 (3.0)	22.1 (2.8)	0.908
Mean (SD) maternal age, years	25 (5.9)	25 (5.9)	25 (6.2)	0.824
Mean (SD) maternal education, completed years at school	3.9 (3.4)	3.9 (3.4)	4.0 (3.5)	0.809
Mean (SD) proxy for socioeconomic status	-0.08 (0.9)	-0.04 (0.9)	-0.03 (0.9)	0.817
Proportion of anaemic women (Hb < 100 g/l)	21.6 %	18.1 %	20.2 %	0.480
Proportion of primiparous women	19.7 %	20.8 %	19.7 %	0.916
Proportion of women with a low BMI (<18.5 kg/m²)	5.4 %	5.5 %	6.3 %	0.860
Proportion of women with a positive HIV test	16.0 %	10.9 %	14.8 %	0.097
Proportion of women with a positive malaria test (RDT)	21.9 %	24.6 %	23.8 %	0.699

[a] Those participants that had a section of placental tissue taken after delivery and were processed for DNA extraction and sequencing.
[b] P-value obtained from ANOVA (comparison of means) or Fishers exact test (comparison of proportions) .

11.3.2 Association between placental and fetal membrane bacterial load, prevalence and nutritional intervention

From participants included in all intervention arms, bacteria were identified in 46.8 % (n=476) of participants' placental tissue and 64.8 % (n=654) of fetal membranes. The mean (SD) bacterial load was 4.8 (0.7) Log10 16S rDNA copies/µl in the placental tissue and 5.2 (0.8) Log10 16S rDNA copies/µl in fetal membranes. Bacterial load had a J-shaped distribution in both the placenta and the fetal membranes with most tissues having very low bacterial loads.

Analysis of bacterial load in the placenta and fetal membranes by intervention group showed an association in the unadjusted values with the bacterial load in the placenta being lower in the LNS group (Table 11-2). The difference in means (95% CI) was -0.17 (-0.31, -0.02) between the LNS and IFA groups (p=0.023) and -0.21 (-0.35, -0.06) between the LNS and MMN groups (p=0.006). This difference was also seen in the adjusted model at the global level and between the LNS and MMN groups, but not between the LNS and IFA groups. There were no differences seen in the presence of 16S rDNA among individuals in the different groups (Table 11-3).

Table 11-2: Bacterial load in the placenta and fetal membranes by intervention group

	Result by study group				Comparison between LNS and IFA group		Comparison between LNS and MMN group		Comparison between MMN and IFA group	
Outcome	IFA	MMN	LNS	*P* value	Difference in means (95 % CI)	*P* value	Difference in means (95 % CI)	*P* value	Difference in means (95 % CI)	*P* value
Mean (SD) bacterial load in the placenta (Log_{10} 16S rDNA copies/µl) [1]	4.85 (0.7)	4.89 (0.7)	4.68 (0.6)	0.014	-0.17 (-0.31, -0.02)	0.023	-0.21 (-0.35, -0.06)	0.006	0.04 (-0.10, 0.18)	0.583
Placenta bacterial load comparison adjusted model [3]				0.049	-0.14 (-0.28, 0.01)	0.071	-0.18 (-0.34, -0.03)	0.019	0.05 (-0.10, 0.20)	0.536
Mean (SD) bacterial load in the fetal membrane (Log_{10} 16S rDNA copies/µl) [2]	5.21 (0.8)	5.22 (0.9)	5.24 (0.8)	0.896	0.03 (-0.12, 0.17)	0.736	0.03 (-0.12, 0.19)	0.650	0.01 (-0.14, 0.16)	0.901
Fetal membrane bacterial load comparison adjusted model [3]				0.761	0.04 (-0.11, 0.19)	0.588	0.05 (-0.09, 0.20)	0.478	-0.01 (-0.16, 0.13)	0.865

[1] IFA n=168, MMN n=151, LNS n =157
[2] IFA n=247, MMN n=260, LNS n =231
[3] Model was adjusted for mother's height at enrolment, mother's age at enrolment, BMI, HIV, primiparity, anaemia, malaria status at enrolment, completed years in education, socio-economic status, study site, time between delivery and collection of placenta sample, caesarean sections and season at enrolment. Covariates were chosen in a predefined analysis plan on the logical potential to form an association with one of the outcomes based on previous literature.

Table 11-3: Prevalence of bacteria in the placenta and fetal membranes by intervention group

	Result by study group				Comparison between LNS and IFA group		Comparison between LNS and MMN group		Comparison between MMN and IFA group	
Outcome	IFA	MMN	LNS	*P* value	Risk ratio (95 % CI)	*P* value	Risk ratio (95 % CI)	*P* value	Risk ratio (95 % CI)	*P* value
Bacteria in the placenta[1]	157 (46.7%)	151 (45.1%)	168 (48.4%)	0.683	1.03 (0.89, 1.21)	0.659	0.96 (0.82, 1.13)	0.668	1.07 (0.91, 1.26)	0.383
Bacteria in the placenta comparison adjusted model [3]				0.400	1.02 (0.87, 1.19)	0.821	0.92 (0.78, 1.08)	0.298	1.11 (0.95, 1.30)	0.198
Bacteria in the fetal membrane[2]	247 (68.9%)	260 (71.4%)	231 (63.9%)	0.094	1.08 (0.97, 1.19)	0.156	1.12 (1.01, 1.23)	0.033	0.97 (0.88, 1.06)	0.475
Bacteria in the fetal membranes comparison adjusted model [3]				0.112	1.07 (0.96, 1.19)	0.233	1.11 (1.01, 1.23)	0.037	0.96 (0.87, 1.05)	0.363

[1] IFA n=347, MMN n=335, LNS n =336
[2] IFA n=358, MMN n=364, LNS n =361
[3]Model was adjusted for mother's height at enrolment, mother's age at enrolment, BMI, HIV, primiparity, anaemia, malaria status at enrolment, completed years in education, socio-economic status, study site, time between delivery and collection of placenta sample, caesarean sections and season at enrolment. Covariates were chosen in a predefined analysis plan on the logical potential to form an association with one of the outcomes based on previous literature.

11.3.3 Association between placental, fetal membrane, oral and vaginal beta diversity and nutritional intervention

Table 11-4 and Table 11-5 show the analysis between bacterial diversity in placenta, fetal membranes, oral cavity and vagina. Phylogenetic diversity was measured using the unweighted UniFrac distance based on pairwise comparisons between all samples. A single value per participant was generated by taking the median distance value from all comparisons. A higher distance value would suggest that an individual's microbiome had phylogenetically dissimilar organisms to the rest of the cohort and a lower value would suggest the opposite. When the values were grouped by intervention arm there were no statistically significant differences seen at the global level in either the unadjusted values or the adjusted models.

Table 11-4: Bacterial diversity in the placenta and fetal membranes by intervention group

	Result by study group				Comparison between LNS and IFA group		Comparison between LNS and MMN group		Comparison between MMN and IFA group	
Outcome	IFA	MMN	LNS	*P* value	Difference in means (95 % CI)	*P* value	Difference in means (95 % CI)	*P* value	Difference in means (95 % CI)	*P* value
Mean (SD) intra-individual unweighted UniFrac distance in placental tissue [1]	0.69 (0.1)	0.69 (0.1)	0.69 (0.1)	0.869	0.00 (-0.01, 0.02)	0.627	0.00 (-0.01, 0.01)	0.947	0.00 (-0.01, 0.01)	0.680
Placenta unweighted UniFrac distance comparison adjusted model [3]				0.872	0.00 (-0.01, 0.02)	0.700	0.00 (-0.01, 0.01)	0.903	0.00 (-0.01, 0.02)	0.620
Mean (SD) intra-individual unweighted UniFrac distance in fetal membranes[2]	0.72 (0.1)	0.72 (0.1)	0.71 (0.1)	0.156	-0.01 (-0.02, 0.00)	0.141	-0.01 (-0.02, 0.00)	0.067	0.00 (-0.01, 0.01)	0.730
Fetal membrane unweighted UniFrac distance comparison adjusted model [3]				0.219	-0.01 (-0.02, 0.00)	0.136	-0.01 (-0.02, 0.00)	0.124	0.00 (-0.01, 0.01)	0.962

[1] IFA n=167, MMN n=150, LNS n =157
[2] IFA n=247, MMN n=260, LNS n =231
[3] Model was adjusted for mother's height at enrolment, mother's age at enrolment, BMI, HIV, primiparity, anaemia, malaria status at enrolment, completed years in education, socio-economic status, study site, time between delivery and collection of placenta sample, caesarean sections and season at enrolment. Covariates were chosen in a predefined analysis plan on the logical potential to form an association with one of the outcomes based on previous literature.

Table 11-5: Bacterial diversity in the oral cavity and vagina by intervention group

	Result by study group				Comparison between LNS and IFA group		Comparison between LNS and MMN group		Comparison between MMN and IFA group	
Outcome	IFA	MMN	LNS	*P* value	Difference in means (95 % CI)	*P* value	Difference in means (95 % CI)	*P* value	Difference in means (95 % CI)	*P* value
Mean (SD) intra-individual unweighted UniFrac distance in the oral cavity[1]	0.49 (0.1)	0.49 (0.1)	0.51 (0.1)	0.342	0.01 (-0.01, 0.02)	0.347	0.01 (-0.01, 0.02)	0.148	0.00 (-0.02, 0.01)	0.609
Oral cavity unweighted UniFrac distance comparison adjusted model [3]				0.326	0.01 (-0.01, 0.02)	0.366	0.01 (-0.00, 0.02)	0.137	0.00 (-0.02, 0.01)	0.553
Mean (SD) intra-individual unweighted UniFrac distance in the vagina	0.59 (0.1)	0.59 (0.1)	0.60 (0.1)	0.510	0.01 (-0.01, 0.02)	0.295	0.00 (-0.02, 0.02)	0.923	0.01 (-0.01, 0.02)	0.340
Vaginal unweighted UniFrac distance comparison adjusted model [3]				0.438	0.01 (-0.01, 0.03)	0.264	0.00 (-0.02, 0.02)	0.990	0.01 (-0.01, 0.03)	0.270

1 IFA n=370, MMN n=371, LNS n =366
2 IFA n=374, MMN n=374, LNS n =356
3 Model was adjusted for mother's height at enrolment, mother's age at enrolment, BMI, HIV, primiparity, anaemia, malaria status at enrolment, completed years in education, socio-economic status, study site, time between delivery and collection of placenta sample, caesarean sections and season at enrolment. Covariates were chosen in a predefined analysis plan on the logical potential to form an association with one of the outcomes based on previous literature.

11.3.4 Association between placental inflammation and nutritional intervention

Prevalence of intervillositis across all participants included in the analysis was 17.5 % (n=176). Chorioamnionitis was recorded in 26.1 % (n=258) of participants whereas 12.1 % (n=120) of participants had severe chorioamnionitis. Table 11-6 shows the analyses between intervillositis and chorioamnionitis across different nutritional interventions. There were no statistically significant differences between intervention groups for these variables at the global level in either the unadjusted values or the adjusted models.

Table 11-6: Inflammation in the placenta by intervention group

					Comparison between LNS and IFA group		Comparison between LNS and MMN group		Comparison between MMN and IFA group	
Outcome	IFA (n=330)	MMN (n=333)	LNS (n=345)	*P* value	Risk ratio (95 % CI)	*P* value	Risk ratio (95 % CI)	*P* value	Risk ratio (95 % CI)	*P* value
Intervillositis[1]	54 (16.4%)	64 (19.2%)	58 (16.8%)	0.580	0.97 (0.69 to 1.58)	0.876	1.14 (0.83 to 1.58)	0.415	0.85 (0.61 to 1.18)	0.337
Intervillositis, adjusted model[2]				0.264	0.97 (0.68 to 1.38)	0.858	1.25 (0.89 to 1.25)	0.195	0.78 (0.55 to 1.09)	0.144
Chorioamnionitis[1]	89 (27.3%)	85 (26.1%)	84 (24.9%)	0.180	1.09 (0.85 to 1.42)	0.473	1.05 (0.81 to 1.36)	0.718	1.05 (0.81 to 1.35)	0.723
Chorioamnionitis, adjusted model[2]				0.486	1.17 (0.89 to 1.52)	0.245	1.13 (0.87 to 1.47)	0.377	1.04 (0.71 to 1.17)	0.772
Severe chorioamnionitis[1]	45 (13.8%)	38 (11.7%)	37 (10.9%)	0.511	1.26 (0.69 to 1.63)	0.265	1.06 (0.69 to 1.63)	0.773	1.18 (0.79 to 1.77)	0.412
Severe chorioamnionitis, adjusted model[2]				0.476	1.27 (0.84 to 1.93)	0.258	1.05 (0.67 to 1.63)	0.842	1.22 (0.79 to 1.86)	0.364

[1] Unadjusted model.
[2] Adjusted p-values were calculated using log-binomial regression models. Models were adjusted for covariates that had a significant association (p<0.10) with the outcome on bivariate analysis, in addition to the corresponding values at enrolment. All models were adjusted for mother's height at enrolment, mother's age at enrolment, BMI, HIV, primiparity, anaemia, malaria status at enrolment, completed years in education, socio-economic status, study site, time between delivery and collection of placenta sample and season at enrolment.

11.3.5 Effect modifiers

Maternal height, maternal BMI, gestational age at enrolment, maternal age, proxy for socioeconomic status, maternal education, primiparity, season at enrolment, maternal anaemia, malarial infection, enrolment site and maternal HIV status were all tested for interactions with the intervention, with respect to the microbiota. The variables were tested using the likelihood ratio test and were selected in a predefined analysis plan as likely to be able to modulate the effect of the nutritional intervention. Only whether the mother was primiparous or multiparous tested positive for effect modification with at least one outcome.

Table 11-7 shows bacterial load by intervention group and stratified by whether the mother was primiparous or multiparous. The comparison had both a significant interaction with the intervention using the likelihood ratio test and had a global significant *P* value between the intervention groups in multiparous women. When stratified by parity, mothers who were multiparous had a lower bacterial load in the LNS group compared to the MMN group (-0.24 (-0.39, -0.09), p=0.002) but not the IFA group.

Table 11-7: Effect modification by intervention group

		Result by study group				Comparison between LNS and IFA group		Comparison between LNS and MMN group		Comparison between MMN and IFA group	
Outcome	Interaction test *P* value[1]	IFA	MMN	LNS	*P* value[2]	Difference in means (95 % CI)	*P* value	Difference in means (95 % CI)	*P* value	Difference in means (95 % CI)	*P* value
Placenta bacterial load, multiparous	0.035	4.8 (0.6)	4.9 (0.7)	4.7 (0.5)	0.010	-0.12 (-0.27, 0.03)	0.127	-0.24 (-0.39, -0.09)	0.002	0.12 (-0.03, 0.27)	0.113

[1] *P* values were calculated using likelihood-ratio test.
[2] *P* values were calculated using ANOVA comparison of means.

11.4 Discussion

This study's findings suggest that provision of LNS to pregnant women may have an impact on lowering bacteria load in full thickness placental samples when compared against the MMN group, especially in mothers that have delivered more than once. However the same result was not seen when compared to the IFA group or when comparing fetal membrane samples. There were no differences between intervention groups when compared to the presence of bacteria in the placenta and fetal membranes or bacterial diversity in the placenta, fetal membranes, oral cavity and the vagina.

Inflammation associated with intrauterine infection can be harmful to the foetus. Low birth weight infants have been shown to have significantly higher levels of oxidative stress markers in their cord blood, when compared to those of a healthy weight (150). As well as injuring the foetus, reactive oxygen species can also damage the amniotic epithelium potentially leading to early delivery. However this effect has been mitigated *in vitro* after treatment with vitamins C and E. However even with the high levels of vitamin deficiency in Malawian diets, there were no statistical differences seen across the different intervention arms in birth outcomes (88) or chorioamnionitis. Although it did not translate into a quantifiable effect on birth outcomes, the lower bacterial load in the LNS group suggests a possible role for increased energy and fat intake in a rural Malawian diet in protecting

against bacterial proliferation in the placenta. However, this was not reflected in the ability to detect the presence of bacterial 16S rDNA.

Comparative studies looking at the associations between nutrient intake, birth outcomes and the microbiome are scarce. A recent study in Tanzania looked at the effect of probiotic yoghurt with additional micronutrients on maternal vaginal, oral and faecal microbiota and its impact on birth outcomes, however it did not study the effect on placental microbiota (149). Similar to this study, they found very little effect on phylogenetic diversity could be explained by the nutritional supplement. Generally large scale shifts in taxa composition were explained by other influences.

Although I found no difference in bacterial diversity in the oral microbiota, the ability to detect a difference could be influenced by the type of sample obtained in this study. A swab was used to rub the gingival margins, which would not pick up pathogenic bacteria found in deep periodontal pockets that could affect oral health and so I cannot rule out that these interventions had no effect.

More complex analysis could be carried out on this dataset to track more minor shifts in community constituents if this was desired. For example, I know the effect that iron has on certain gut and vaginal taxa (144,151) and so I could look at the effect on abundance of OTUs known to utilise iron sources, rather than focusing on the entire community. The effect might also be more acute in at-risk populations, such as anaemic individuals. However, as I did not find any major shifts in multiple bacterial and inflammatory

variables across the intervention arms I have chosen to focus the subsequent result chapters in this thesis to studying their possible association with birth outcomes.

12 Chapter 5: The placental microbiome, its association with adverse birth outcomes and the potential source of intrauterine bacteria

12.1 Introduction

Sub-Saharan Africa has had consistently higher incidence of preterm deliveries (152) compared with the rest of the world, with recent estimates of between 10.0% (88) and 16.3% (153) of all births in Malawi. While the aetiology of spontaneous preterm labour remains elusive, a role for bacterial infection and colonisation of fetal membranes and the associated maternal inflammatory immune response is now recognized as the probable trigger in some cases. Spontaneous preterm birth (SPTB) is distinguished by a higher frequency of bacterial colonisation and by the nature and diversity of bacterial species (63). There is increasing evidence to suggest that it is the type of bacteria present in preterm deliveries, rather than simply the presence of bacteria, that differs from term deliveries (62,64,80). While bacteria can infect, and sometimes cross the placenta via the maternal circulation, in the majority of cases the common route of spread is through ascending infection from the vagina through the cervical canal (70). It appears that it is these bacteria that leads to inflammation of the chorioamniotic membranes (chorioamnionitis), well recognised as being highly associated with earlier delivery (70,154).

Whereas most studies to date have been conducted in Europe and North America, in this study I wanted to describe the core microbiome found in placental tissue and fetal membranes in a cohort of women in rural southern Malawi. I hypothesised that microbial community structure would be altered in participants with chorioamnionitis and those changes would be associated with duration of pregnancy, birth weight, newborn length and newborn head circumference.

12.2 Materials and Methods

12.2.1 Study design and enrolment

See section 9.1

12.2.2 Study setting

See section 9.2

12.2.3 Collection of birth outcome and baseline data

See section 9.3

12.2.4 Sample collection

Sample collection for genomic DNA extraction was carried out as stipulated in section 9.4. To examine for chorioamnionitis, one fetal membrane and one placenta were placed in 10% neutral buffered formalin fixative, processed and embedded in paraffin wax. These were sectioned at 3-5 micron thick and stained with haematoxylin & eosin before being read.

12.2.5 DNA Extraction

In preparation for extraction of genomic DNA, each frozen tissue sample was cut into smaller pieces using a sterile pair of surgical scissors. 20-50mg of tissue was then transferred to a sterile 2ml screw-cap tube and extracted using same protocol as stipulated in section 9.5.

12.2.6 16S rDNA broad-range qPCR

See section 9.6.

12.2.7 16S rDNA amplicon high-throughput sequencing

Samples positive for bacterial DNA by qPCR were selected for sequencing. Sequencing was carried out as per protocol in section 9.7.

12.2.8 Bioinformatics and statistical analysis

See section 9.8.

12.3 Results

12.3.1 Sample collection

A total of 1391 participants were recruited into the iLiNS-DYAD-M trial and enrolment began in February 2011, with the last delivery taking place in February 2013. 1097 (78.9%) participants had at least one placental or fetal membrane tissue analysed for bacteria and histology. Figure 12-1 shows a flow diagram that documents reasons for the loss to follow-up of the 21.1% of participants who were not included in this microbiome study. Of these, 49 were excluded because they had twins, a placental sample wasn't collected or no histological result was obtained.

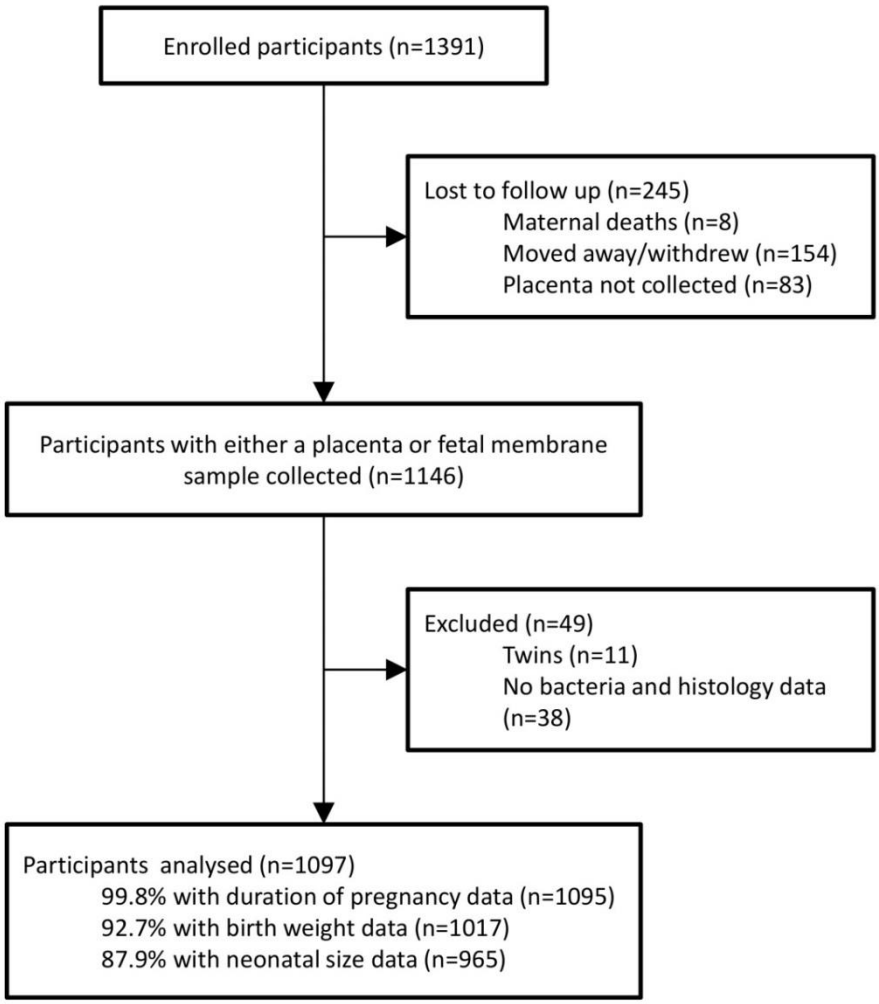

Figure 12-1: Study participant flow diagram.

When the participants that were included in this study were compared to those excluded, those included had a lower BMI (22.1 vs 22.6 kg/m2, p=0.005), were older (25 vs 24 years, p=0.025), had completed less education (3.9 vs 4.7 years, p<0.001), had a lower score for socioeconomic status (-0.07 vs 0.36, p<0.001) and were less likely to be primiparous (20.4% vs 28.1%, p=0.006) (Table 12-1).

Table 12-1: Baseline characteristics of included and excluded participants

Characteristic	Included (n=1097)[1]	Excluded (n=294)	*P* value
Mean (SD) BMI, kg/m^2	22.1 (2.8)	22.6 (3.0)	0.005
Mean (SD) maternal age, years	25.1 (6.1)	24.2 (6.3)	0.025
Mean (SD) maternal education, completed years at school	3.9 (3.4)	4.7 (3.8)	<0.001
Mean (SD) proxy for socioeconomic status	-0.07 (0.9)	0.36 (1.2)	<0.001
Proportion of aenemic women (Hb < 110 g/l)	19.9 %	24.2 %	0.118
Proportion of primiparous women	20.4 %	28.1 %	0.006
Proportion of women with a low BMI (< 18.5 kg/m^2)	5.6 %	4.7 %	0.656
Proportion of women with a positive HIV test	13.3 %	15.0 %	0.469
Number (%) of women with a positive malaria test (RDT)	23.4 %	22.7 %	0.874

[1]Total n=1391
[2] *P* value obtained from t-test (comparison of means) or Fishers exact test (comparison of proportions).

12.3.2 Detection of bacterial DNA

Bacteria were detected in 738 (68·1%) of fetal membranes and in 476 (46·8%) of placental samples. Of those participants that had detectable bacteria the mean (SD) bacterial load was 5.22 (0.84) Log_{10} 16S rDNA copies/µl in the fetal membranes and 4.80 (0.66) Log_{10} 16S rDNA/µl copies in the placenta. The median sequencing depths for placental tissue and fetal membranes were 11,803 and 21,040 reads respectively (Table 12-2).

Table 12-2: Sequencing output.

Sample type	Number of samples	Median number of reads per sample (IQR)	Total number of reads
Placental tissue	476	11,803 (3800,33561)	14,001,032
Fetal membrane	738	21,040 (7340,54473)	30,941,823
Total	1214		44,942,855

12.3.3 The microbiome found in the placental tissues at delivery is distinct

Figure 12-2 shows, across the population analysed, the 25 commonest organisms detected within placental tissues. Bacterial patterns were similar in fetal membranes and placental tissue. Fetal membranes had a higher incidence of *Lactobacillus iners, Gardnerella vaginalis* and *Sneathia sanguinegens* whereas the placental tissues had higher incidences of *Acinetobacter* spp. and *Enterobacteriaceae* spp.

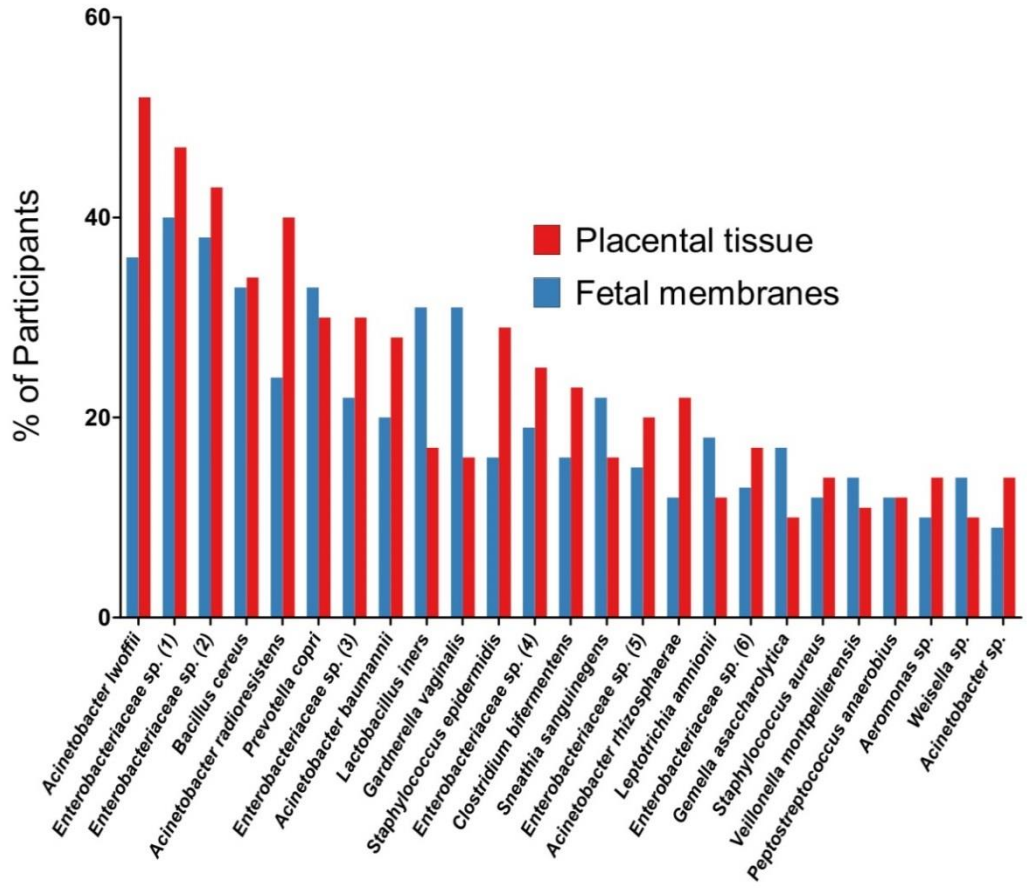

Figure 12-2: Rank abundance curve.
The 25 commonest organisms recovered from participants' placental tissue (n=476) and fetal membranes (n=738).

Comparing individuals, the bacteria detected within the placental and fetal membrane tissues were indistinguishable, suggesting in many cases a large overlap of similar taxa (Figure 12-3). Comparison of bacteria in placental and fetal tissues with bacteria in vaginal and oral samples from the same individuals revealed some overlap with the microbiota within the vagina, but very little overlap with the microbiota from the oral cavity (Figure 12-3). The placental and fetal membrane microbial communities had high variability when compared to the vaginal and oral microbiomes. Intra-individual

unweighted UniFrac distances in the placenta and fetal membranes were higher than the oral cavity and vagina. Placenta and fetal membranes had a mean (SD) UniFrac distance across all individuals of 0.69 (0.06) and 0.71 (0.06) respectively, compared to 0.50 (0.09) in the oral cavity and 0.59 (0.11) in the vagina.

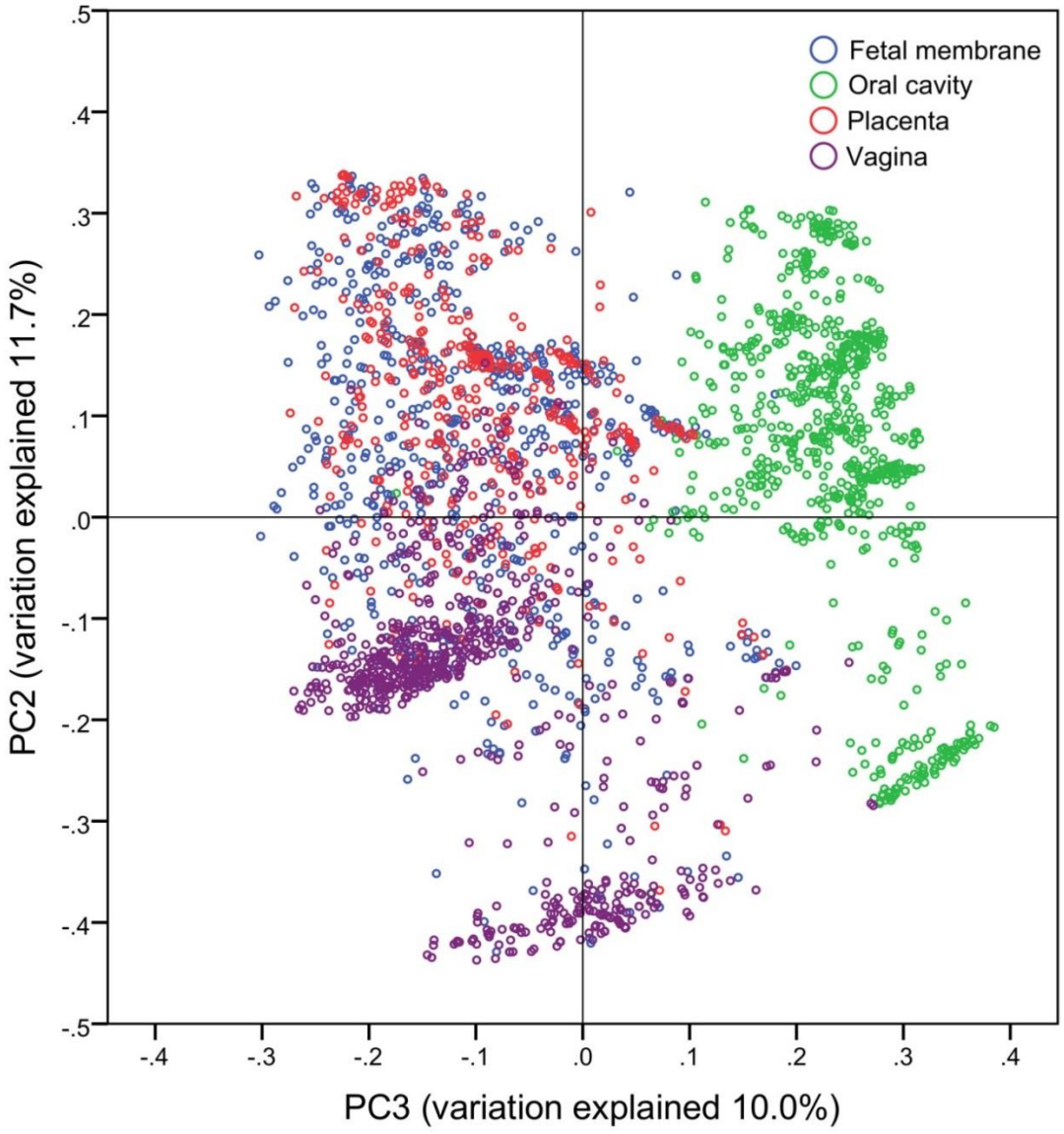

Figure 12-3: Principal component analysis of unweighted UniFrac distances computed for matched individual body sites.
Distances computed from matched placenta (n=445), fetal membrane (n=719), oral (n=725) and vaginal (n=747) samples from 1107 participants.

12.3.4 Variation in the placenta microbiome is determined by abundance of a limited range of species

Figure 12-4 and Figure 12-5 show that higher bacterial loads, as determined by qPCR, were observed when there were fewer OTUs, as determined by 16S amplicon sequencing. In the placental tissue, a drop in observed OTUs of 1 was associated with a mean (95%CI) increase in bacterial load of 0·03 (0·03, 0·04) log 10 16S rDNA copies / µl, p<0·001. A decrease in observed OTUs of 1 in fetal membranes was associated with a mean (95%CI) increase in bacterial load of 0·04 (0·03, 0·05), p<0·001. This indicates that a high bacterial load is linked to the expansion of a limited number of organisms. However, the high median intra-individual unweighted UniFrac distances indicate that the composition of the microbial communities is highly variable between individual women. Figure 3 demonstrates that a decrease in observed OTUs of 1 in placental tissue was associated with a rise of the mean (95%CI) intra-individual unweighted UniFrac distances of 0·02 (0·02, 0·03), p<0·001 (Figure 3a). In fetal membranes a decrease in observed OTUs of 1 was associated with a rise of the mean intra-individual unweighted UniFrac distance of 0·03 (0·03, 0·04), p<0·001 (Figure 3b).

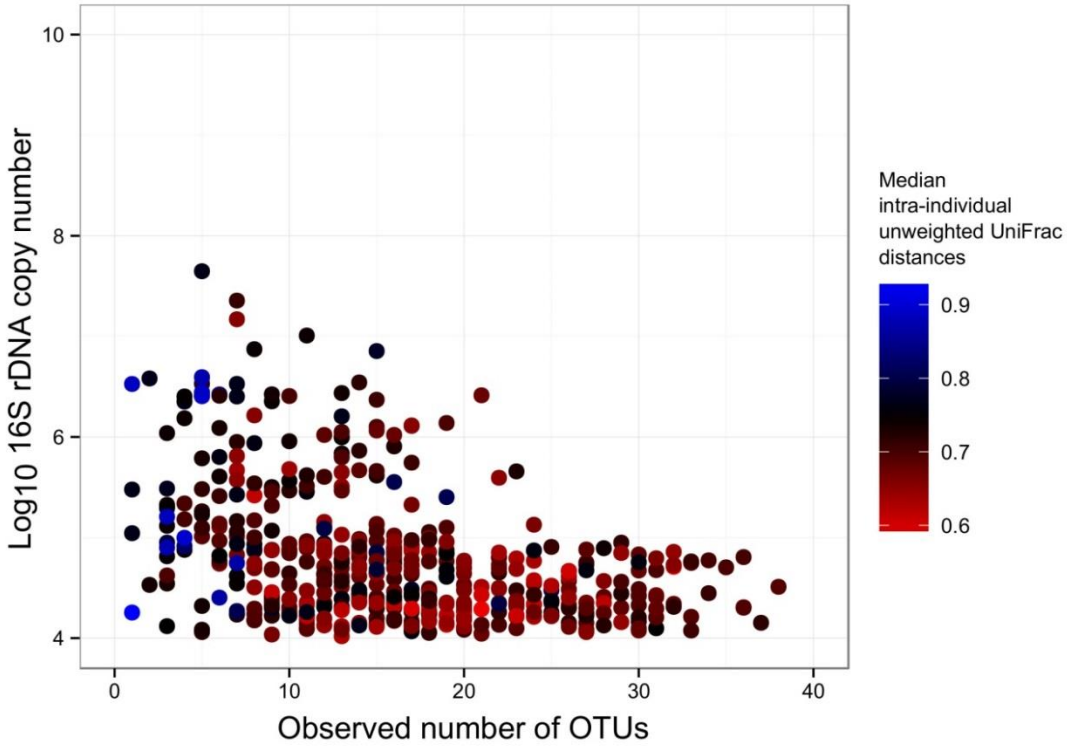

Figure 12-4: Multivariate plot of participant's' placental bacterial load, observed number of OTUs and median intra-individual unweighted UniFrac distances.

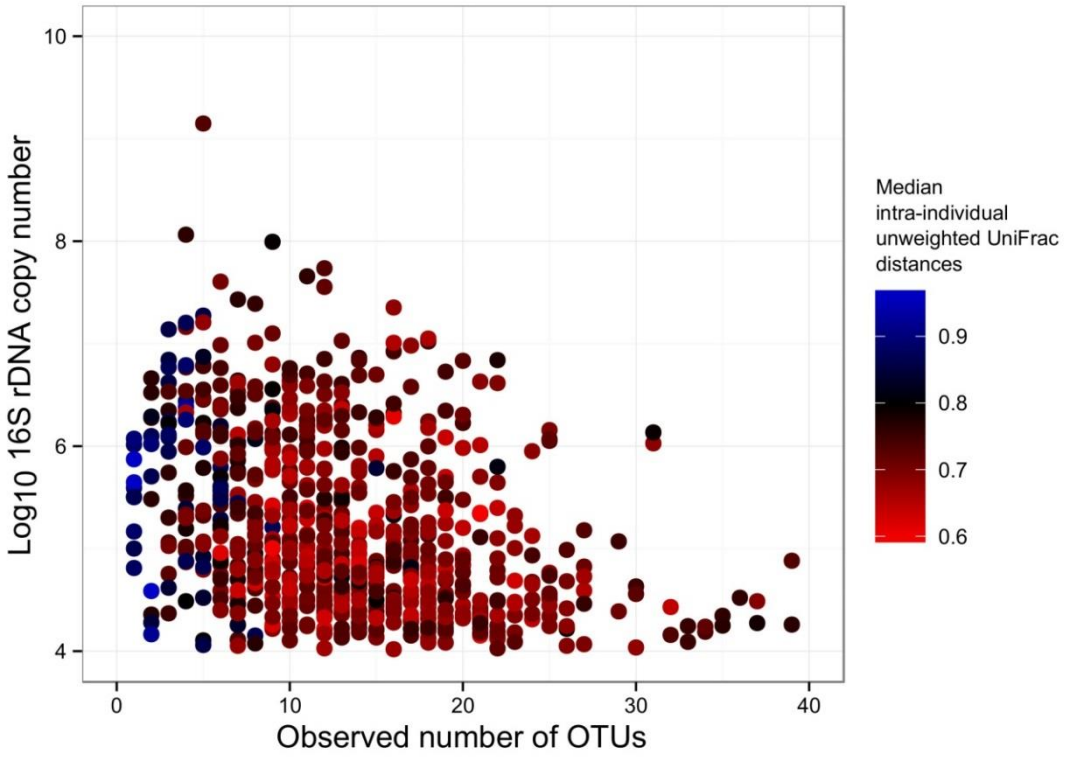

Figure 12-5: Multivariate plot of participant's' fetal membrane bacterial load, observed number of OTUs and median intra-individual unweighted UniFrac distances.

12.3.5 Distinct combinations of bacteria associate with each other

The placental microbial community structure was probed using the associations between the abundances of different OTUs. The abundances of a distinct group of phyla containing Fusobacteria, Tenericutes, Bacteroidetes and Actinobacteria were found to be positively correlated (Figure 12-6). Proteobacteria and Firmucutes were rarely found together and were negatively correlated with all other phyla.

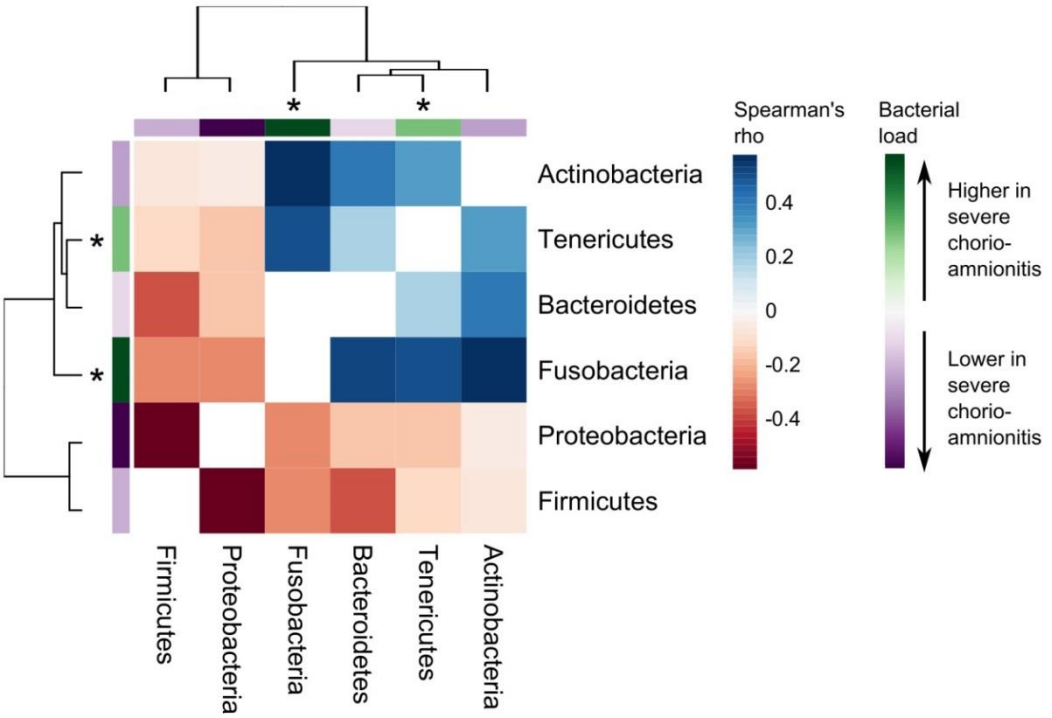

Figure 12-6: Heat map of Spearman's correlations between the 6 most abundant bacterial phyla in fetal membranes.

Hierarchical clustering was computed by complete linkage of Euclidean distances. The heat map is annotated with mean difference in bacterial load between participants with and without severe chorioamnionitis for each bacterial phylum. Asterisks indicate p<0·05 association between higher load of that bacterial phylum and prevalence of severe chorioamnionitis. Adjusted model *P* values were calculated using linear regression adjusting for the nutritional intervention, maternal BMI at enrolment, maternal age, proxy for socioeconomic status, number of previous pregnancies, anaemia, site of enrolment, mode of delivery and time between delivery and placenta sampling.

To explore these interactions in more detail I examined correlations among the 20 most abundant bacterial families and using hierarchical clustering this split the taxa into two major groups (Figure 12-7). One group, in which *Fusobacteriaceae*, *Mycoplasmataceae*, *Bifidobacteriaceae*, *Prevotellaceae* and *Leptotrichiaceae* abundances clustered together, matched the same correlations seen at the phyla level. However, the group also included many families from the Firmicutes phyla such as *Clostridiaceae*, *Lachnospiraceae*,

Lactobacillaceae and *Peptostreptococcaceae*. The second group of organisms that clustered together included a group of families from the phyla Firmicutes that included *Streptococcaceae*, *Staphylococcaceae*, *Leuconostocaceae*, *Enterococcaceae* and *Lactobacillales,* whose abundances were positively correlated with each other. The other organisms in that cluster were overwhelmingly from the phyla Proteobacteria and included *Enterobacteriaceae*, *Aeromonadaceae*, *Pseudomonadaceae* and *Moraxellaceae*. Abundances for all families in the phyla Proteobacteria were positively correlated with each other and with the abundance of *Bacillaceae*. Similar patterns were seen in full-thickness placental tissue samples at both the phylum and family level (Figure 12-8 & Figure 12-9).

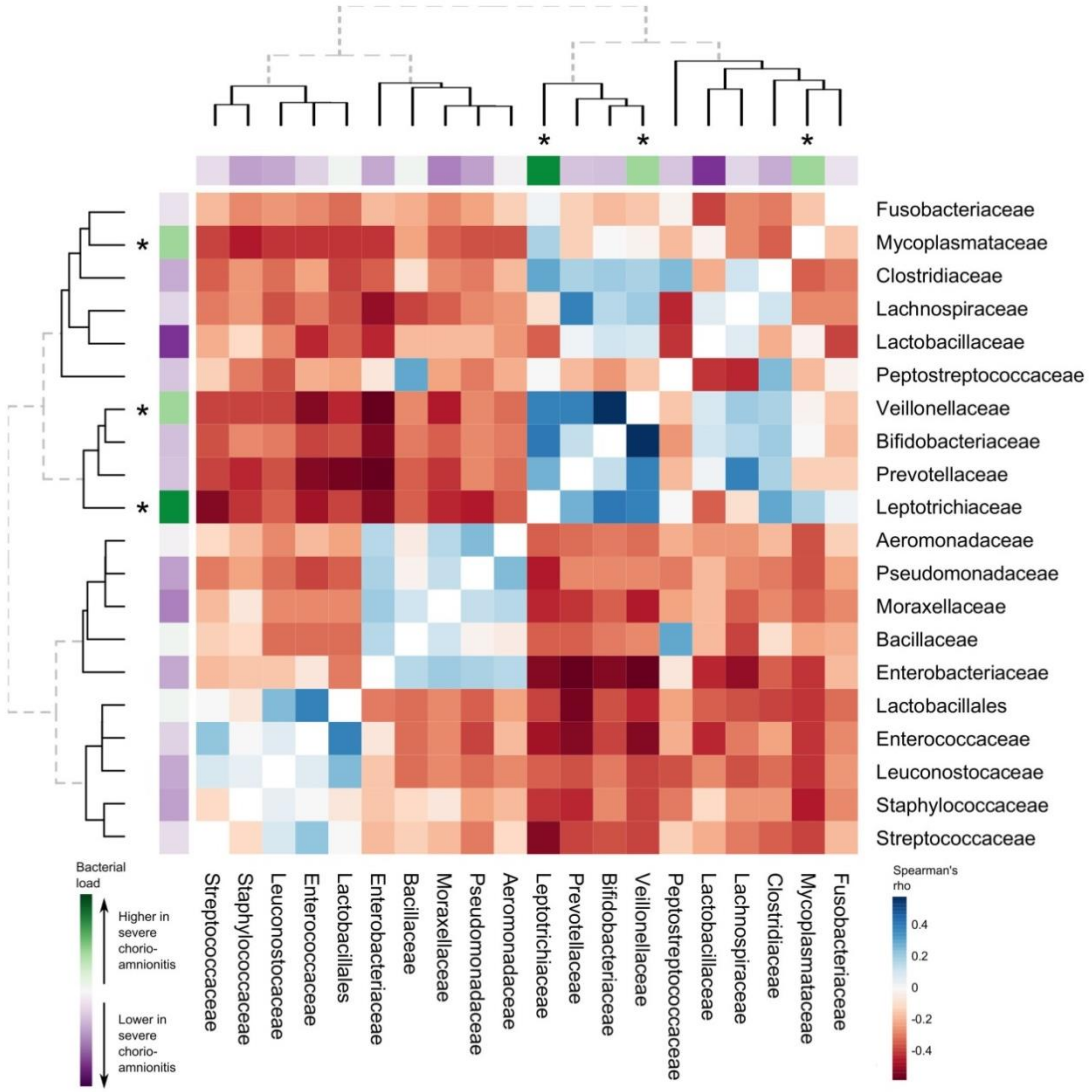

Figure 12-7: Heat map of Spearman's correlations between the 20 most abundant bacterial families in fetal membranes.

Hierarchical clustering was computed by complete linkage of Euclidean distances. The heat map is annotated with mean difference in bacterial load between participants with and without severe chorioamnionitis for each bacterial family. Asterisks indicate p<0·05 association between higher load of that bacterial family and prevalence of severe chorioamnionitis. Adjusted model *P* values were calculated using linear regression adjusting for the nutritional intervention, maternal BMI at enrolment, maternal age, proxy for socioeconomic status, number of previous pregnancies, anaemia, site of enrolment, mode of delivery and time between delivery and placenta sampling.

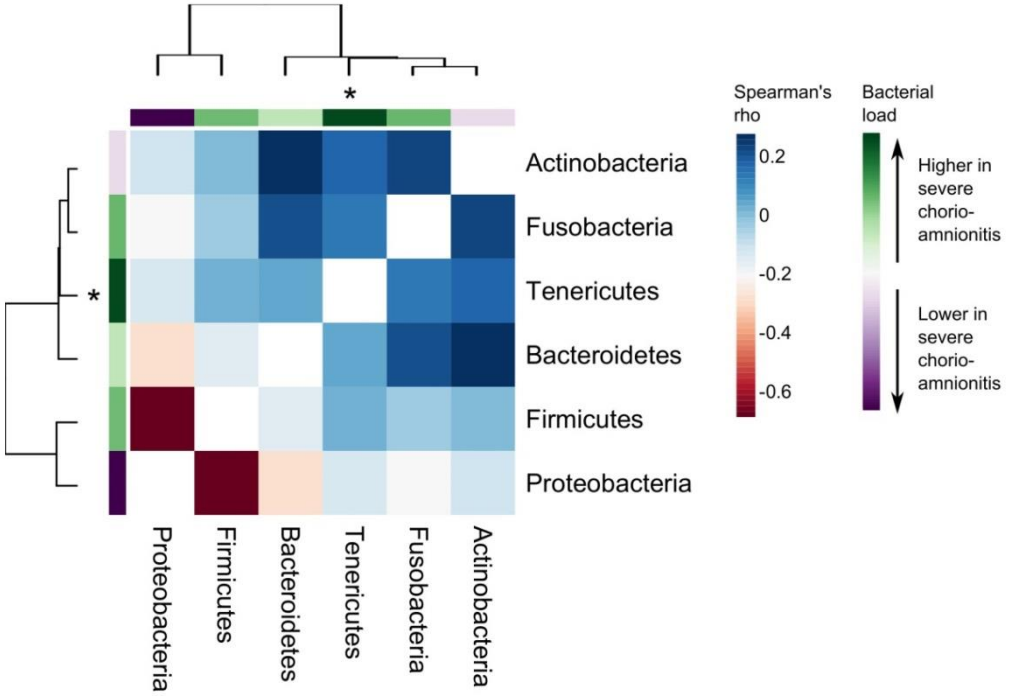

Figure 12-8: Heat map of Spearman's correlations between the 6 most abundant bacterial phyla in placental tissue.

Hierarchical clustering was computed by complete linkage of Euclidean distances. The heat map is annotated with mean difference in bacterial load between participants with and without severe chorioamnionitis for each bacterial phylum. Asterisks indicate p<0·05 association between higher load of that bacterial phylum and prevalence of severe chorioamnionitis. Adjusted model *P* values were calculated using linear regression adjusting for the nutritional intervention, maternal BMI at enrolment, maternal age, proxy for socioeconomic status, number of previous pregnancies, anaemia, site of enrolment, mode of delivery and time between delivery and placenta sampling.

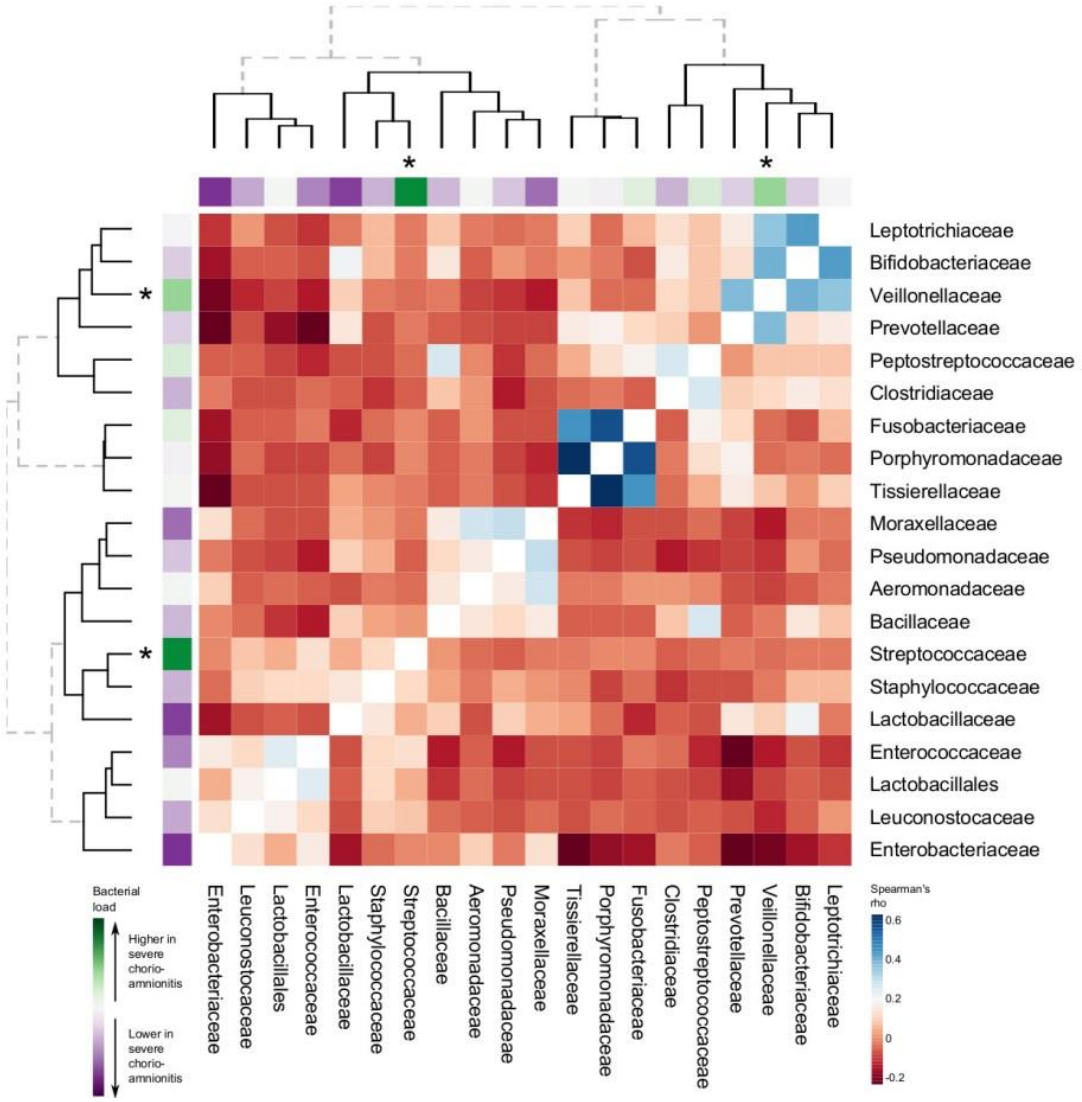

Figure 12-9: Heat map of Spearman's correlations between the 20 most abundant bacterial families in placental tissue.

Hierarchical clustering was computed by complete linkage of Euclidean distances. The heat map is annotated with mean difference in bacterial load between participants with and without severe chorioamnionitis for each bacterial family. Asterisks indicate p<0·05 association between higher load of that bacterial family and prevalence of severe chorioamnionitis. Adjusted model *P* values were calculated using linear regression adjusting for the nutritional intervention, maternal BMI at enrolment, maternal age, proxy for socioeconomic status, number of previous pregnancies, anaemia, site of enrolment, mode of delivery and time between delivery and placenta sampling.

12.3.6 Severe chorioamnionitis

A total of 258 (26·1%) participants had histologically determined chorioamnionitis with 120 (12·1%) participants classified as having severe chorioamnionitis. Presence of chorioamnionitis was not significantly associated with birth outcomes, however, participants with severe chorioamnionitis delivered significantly earlier (-0·4 (-0.8,-0·1) gestational weeks, p=0·019) than those without severe chorioamnionitis. Those with severe chorioamnionitis also had a lower observed number of OTUs in the placenta (p=0·029) and fetal membranes (p=0·025) and increased intra-individual unweighted UniFrac distances in the placenta (p=0·034) and fetal membranes (p=0·003) (Figure 12-10 & Figure 12-11). Severe chorioamnionitis was also associated with distinct phyla.

A higher mean Fusobacteria load of 0·41 (0·17, 0·64) Log_{10} 16S rDNA copies/µl (p=0·001) and higher mean Tenericutes load of 0·18 (0·05, 0·30) Log_{10} 16S rDNA copies/µl (p=0·006) were found in fetal membranes with severe chorioamnionitis (Figure 12-6). This was also reflected at a family level with a significantly higher mean bacterial load of *Mycoplasmataceae* (p=0·010), *Leptotrichiaceae* (p=0·001), and *Veillonaceae* (p=0·001) in placental tissues with severe chorioamnionitis (Figure 12-7). Abundances of *Mycoplasmataceae*, *Leptotrichiaceae,* and *Veillonaceae* were also all positively intercorrelated.

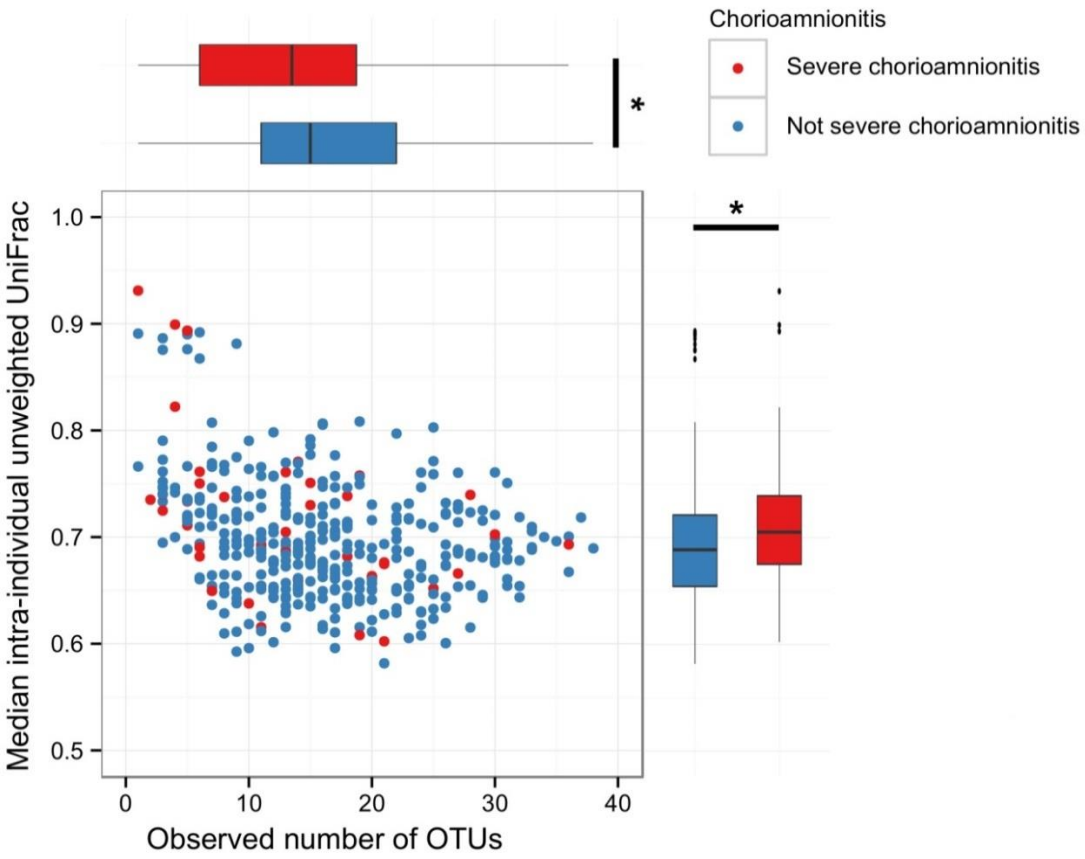

Figure 12-10: Comparison between observed number of OTUs and median intra-individual unweighted UniFrac distances in placental tissue and presence of severe chorioamnionitis.

Adjusted model P values were calculated using linear regression adjusting for the intervention, maternal BMI at enrolment, maternal age, proxy for socioeconomic status, number of previous pregnancies, anaemia, site of enrolment, mode of delivery and time between delivery and placenta sampling (*p<0.05, **p<0.01).

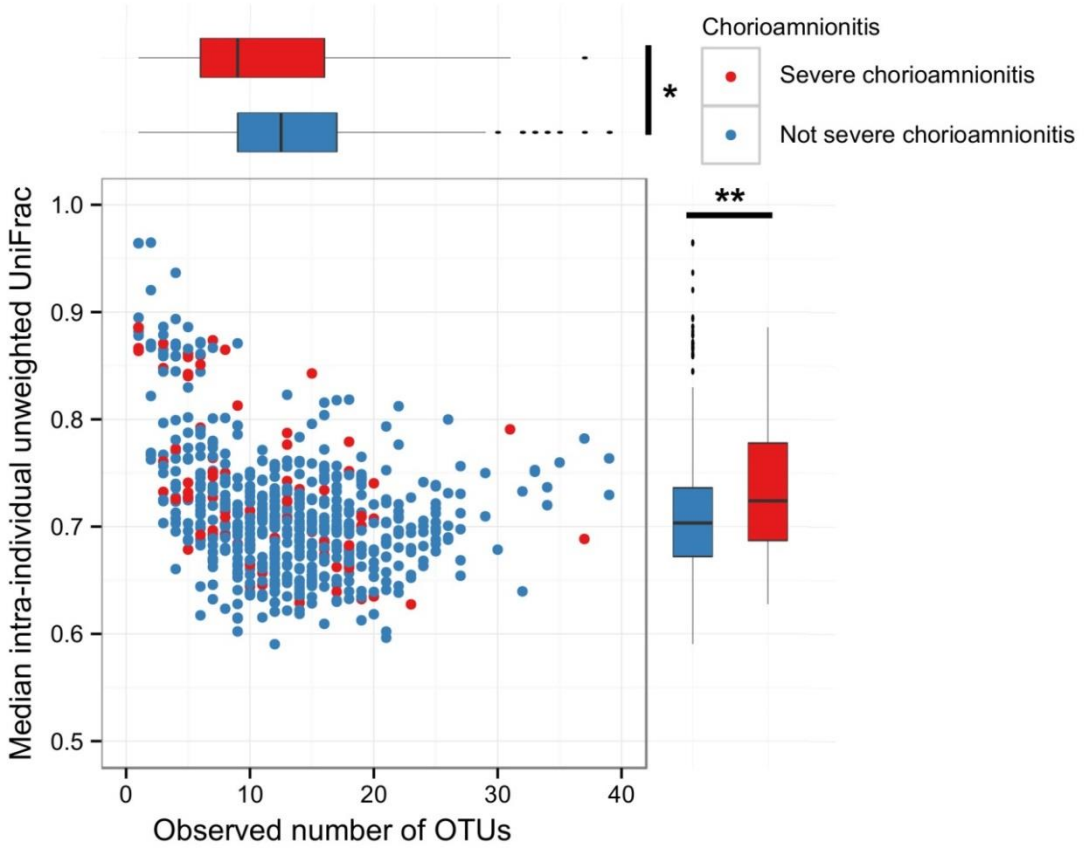

Figure 12-11: Comparison between observed number of OTUs and median intra-individual unweighted UniFrac distances in fetal membranes and presence of severe chorioamnionitis.
Adjusted model P values were calculated using linear regression adjusting for the intervention, maternal BMI at enrolment, maternal age, proxy for socioeconomic status, number of previous pregnancies, anaemia, site of enrolment, mode of delivery and time between delivery and placenta sampling (*p<0.05, **p<0.01).

12.3.7 Microbial composition recovered from fetal membranes differed when the individual delivered early or when the newborn had restricted growth

31 OTUs from fetal membranes were identified whose abundances were positively correlated with or belonging to bacterial families that were significantly associated with severe chorioamnionitis (denoted with an

asterisk in figure Figure 12-7). Generally organisms fell into two groups, those that were associated with a shorter duration of pregnancy and those that were associated with a smaller newborn size (birth weight, length-for-age z-score and head circumference-for-age z-score). Higher bacterial loads of *Fusobacterium nucleatum*, *Ureaplasma* sp. and *Gemella asaccharolytica* were associated with a shorter duration of pregnancy (Table 12-3). Higher loads of *Gardnerella vaginalis*, *Sneathia sanguinegens*, *Prevotella copri*, *Prevotella amnii, Lachnospiraceae* sp., *Blautia* sp. and *Phascolarctobacterium succinatutens* were all significantly associated with smaller newborn size (Table 12-4, Table 12-5 & Table 12-6). After controlling for the false discovery rate using the Benjamini-Hochberg procedure, only a higher mean bacterial load of an unidentified *Lachnospiraceae* sp. was significantly associated with a lower newborn head circumference-for-age z-score (Table 12-6).

Table 12-3: OTUs isolated from fetal membranes significantly associated with differences in duration of delivery.

			Unadjusted analysis			Adjusted analysis		
O.T.U ID (Custom database and greengenes)	Greengenes Taxonomy	Genbank BLASTN result	Correlation coefficient	*P* value[1]	N	Regression coefficient (95%CI)	*P* value[2]	*q* value
270856	g__Ureaplasma; s__	*Ureaplasma sp.*	-0.11	0.003	737	-0.33 (-0.56, -0.09)	0.007	0.217
5119	g__Fusobacterium; s__	*Fusobacterium nucleatum*	-0.11	0.003	737	-0.43 (-0.78, -0.08)	0.015	0.233
558508579	g__Gemella; s__	*Gemella asaccharolytica*	-0.09	0.018	737	-0.12 (-0.01, -0.23)	0.027	0.279
3677	f__Peptostreptococcaceae; g__	*Clostridium sp.*	0.08	0.024	737	0.04 (-0.04, 0.14)	0.349	1.000

[1] P value calculated using Pearson's correlation.
[2] Adjusted p-values were calculated using linear regression models. Regression coefficient shows the change in duration of delivery (weeks) in relation to an increase in bacterial load of 1 Log10 genomes/µl. All models were adjusted for the intervention, maternal BMI at enrolment, maternal age, proxy for socioeconomic status, number of previous pregnancies, anaemia, site of enrolment, mode of delivery and time between delivery and placenta sampling.

Table 12-4: OTUs isolated from fetal membranes significantly associated with differences in birth weight.

O.T.U ID (Custom database and greengenes)	Greengenes Taxonomy	Genbank BLASTN result	Unadjusted analysis			Adjusted analysis		
			Correlation coefficient	*P* value[1]	N	Regression coefficient (95%CI)	*P* value[2]	*q* value
645321357	g__Sneathia; s__	*Sneathia sanguinegens*	-0.10	0.008	681	-25 (-44, -7)	0.008	0.248
288932	Prevotella; s__copri	*Prevotella copri*	-0.10	0.012	681	-51 (-94, -9)	0.018	0.279
10703	f__Bifidobacteriaceae	*Gardnerella vaginalis*	-0.08	0.028	681	-19 (-37, 0)	0.045	0.465
631251895	g__Prevotella; s__	*Prevotella amnii*	-0.08	0.030	681	-24 (-56, 9)	0.158	1.000

[1] P value calculated using Pearson's correlation.
[2] Adjusted p-values were calculated using linear regression models. Regression coefficient shows the change in birth weight (grams) in relation to an increase in bacterial load of 1 Log10 genomes/µl. All models were adjusted for the intervention, maternal BMI at enrolment, maternal age, proxy for socioeconomic status, number of previous pregnancies, anaemia, site of enrolment, mode of delivery and time between delivery and placenta sampling.

Table 12-5: OTUs isolated from fetal membranes significantly associated with differences in length-for-age Z-score.

			Unadjusted analysis			Adjusted analysis		
O.T.U ID (Custom database and greengenes)	Greengenes Taxonomy	Genbank BLASTN result	Correlation coefficient	*P* value[1]	N	Regression coefficient (95%CI)	*P* value[2]	*q* value
2386814	f__Lachnospiraceae; g__; s__	*Lachnospiraceae sp.*	-0.11	0.006	644	-0.3 (-0.5, -0.1)	0.005	0.199
10703	f__Bifidobacteriaceae	*Gardnerella vaginalis*	-0.08	0.044	644	-0.1 (-0.1, -0.0)	0.007	0.678
645321357	g__Sneathia; s__	*Sneathia sanguinegens*	-0.08	0.036	644	-0.1 (-0.1, -0.0	0.020	0.372
302279	g__Phascolarctobacterium; s__	*Phascolarctobacterium succinatutens*	-0.09	0.026	644	-0.1 (-0.3, -0.0)	0.024	0.203
288932	g__Prevotella; s__copri	*Prevotella copri*	-0.09	0.017	644	-0.1 (-0.8, -0.1)	0.025	0.107
2700884	f__Lachnospiraceae	*Blautia sp.*	-0.08	0.048	644	-0.4 (-0.7, -0.0)	0.032	0.248

[1] P value calculated using Pearson's correlation.
[2] Adjusted p-values were calculated using linear regression models. Regression coefficient shows the change in length-for-age Z-score in relation to an increase in bacterial load of 1 Log10 genomes/µl. All models were adjusted for the intervention, maternal BMI at enrolment, maternal age, proxy for socioeconomic status, number of previous pregnancies, anaemia, site of enrolment, mode of delivery and time between delivery and placenta sampling.

Table 12-6: OTUs isolated from fetal membranes significantly associated with differences in head circumference-for-age Z-score.

			Unadjusted analysis			Adjusted analysis		
O.T.U ID (Custom database and greengenes)	Greengenes Taxonomy	Genbank BLASTN result	Correlation coefficient	*P* value[1]	N	Regression coefficient (95%CI)	*P* value[2]	*q* value
2386814	f__Lachnospiraceae; g__; s__	*Lachnospiraceae sp.*	-0.13	0.001	647	-0.4 (-0.6, -0.2)	0.000	0.003
302279	g__Phascolarctobacterium; s__	*Phascolarctobacterium succinatutens*	-0.11	0.007	647	-0.2 (-0.3, -0.1)	0.005	0.078
645321357	g__Sneathia; s__	*Sneathia sanguinegens*	-0.12	0.002	647	-0.1 (-0.1, -0.0	0.007	0.072
288932	g__Prevotella; s__copri	*Prevotella copri*	-0.09	0.023	647	-0.1 (-0.2, 0.0)	0.067	0.519
631251895	g__Prevotella; s__	*Prevotella amnii*	-0.10	0.014	647	-0.1 (-0.2, 0.0)	0.125	0.775

[1] P value calculated using Pearson's correlation.

[2] Adjusted p-values were calculated using linear regression models. Regression coefficient shows the change in head circumference-for-age Z-score in relation to an increase in bacterial load of 1 Log10 genomes/µl. All models were adjusted for the intervention, maternal BMI at enrolment, maternal age, proxy for socioeconomic status, number of previous pregnancies, anaemia, site of enrolment, mode of delivery and time between delivery and placenta sampling.

12.3.8 Source of Placental Bacteria

SourceTracker (92) was used to identify the likely source for the bacteria detected in the placental and fetal membrane tissues. I identified 20 OTUs that were detectable from at least one participant's placental tissue or fetal membrane and were also present in a matched individual sample from either the vagina or oral cavity. I compared their abundances with proportions of bacteria identified from SourceTracker as being from either the vagina or the oral cavity. 16 OTUs were positively associated with only a single source site using SouceTracker's predictions, with 14 OTUs from the vagina and two from the oral cavity (Figure 12-12). Four bacteria could not be positively correlated with either source and so were not taken forward into subsequent analysis.

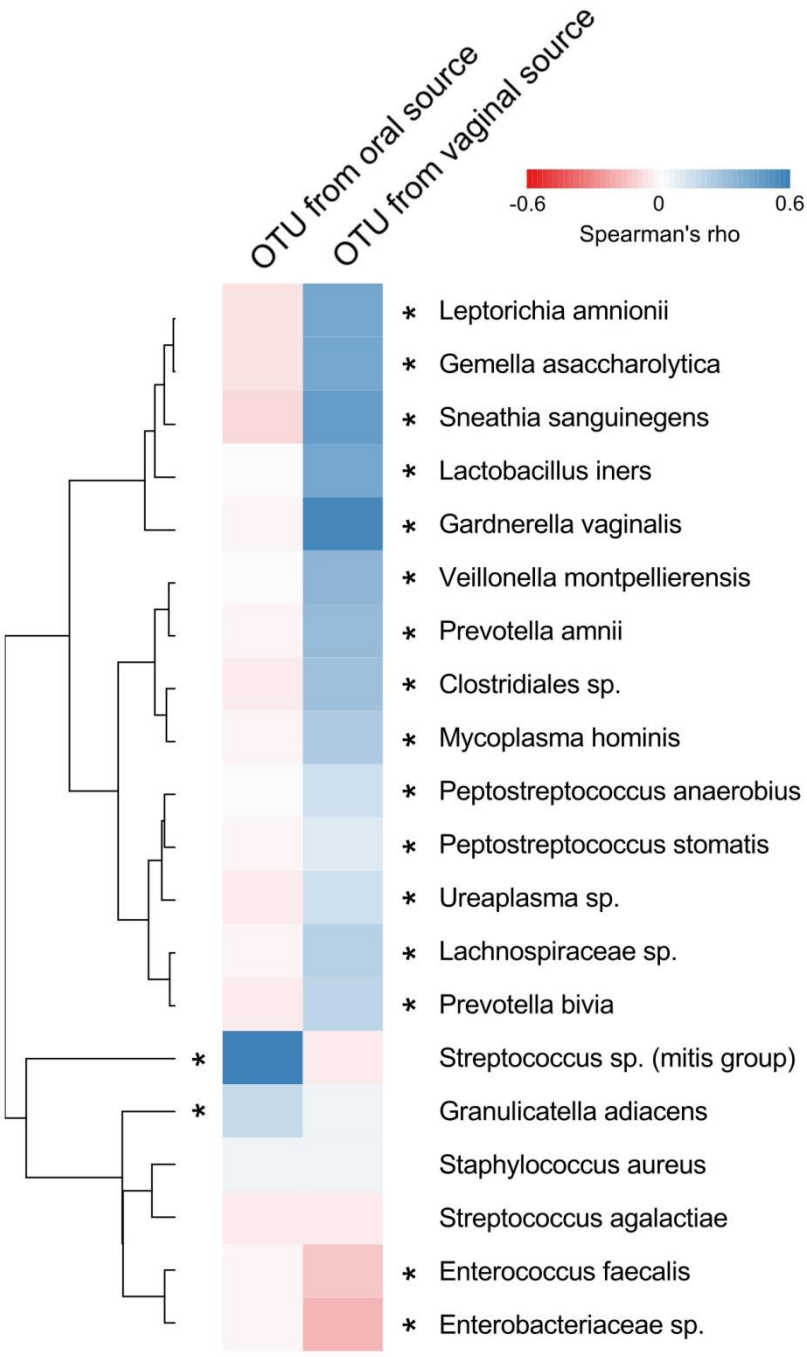

Figure 12-12: Identifying individual OTUs in placental and fetal membrane tissues sourced from either the vagina or oral cavity.

The heat map plots the Spearman's rho correlations between the proportion of OTUs in each individual that was predicted to be sourced from either the vagina or the oral cavity and individual OTU abundances. If both are positively correlated, it is assumed that the OTU is likely to have come from that source. Hierarchical clustering was computed by complete linkage of Euclidean distances.

A larger proportion of OTUs were identified as being sourced from the vagina than from the oral cavity (Figure 12-13). Also, similar proportions of organisms from the vagina were found in vaginally delivered and C-section delivered placenta and fetal membranes (Figure 12-14). A high proportion of organisms found within the placenta and fetal membranes did not have an obvious source (Figure 12-13).

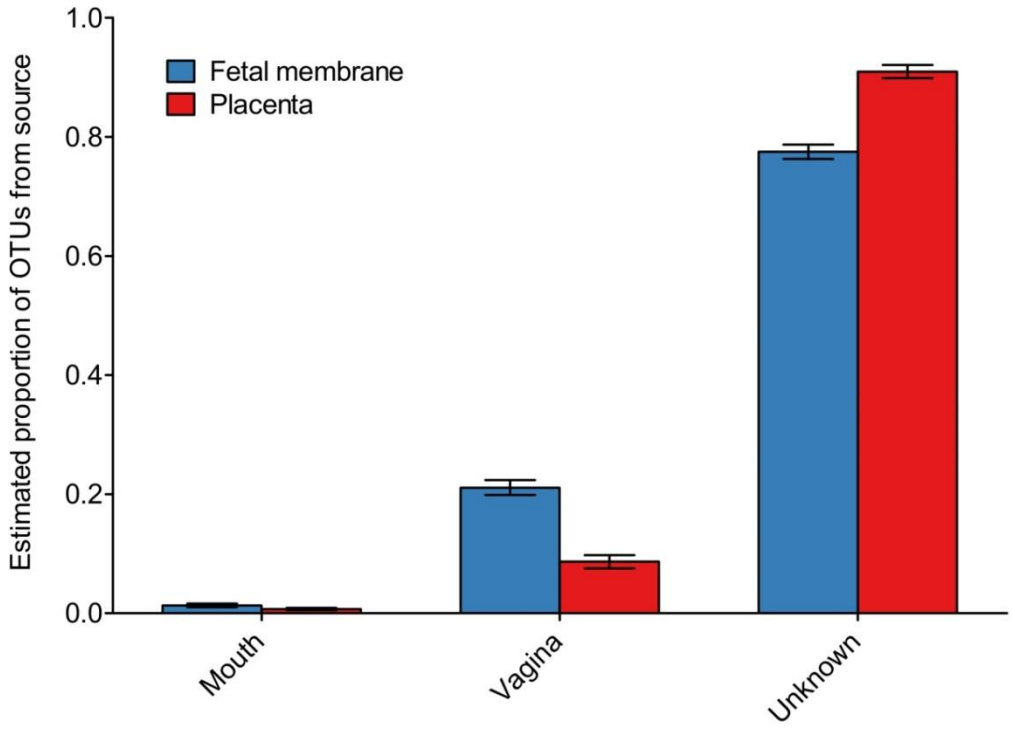

Figure 12-13: The mean ± SEM proportion of OTUs in placenta and fetal membranes predicted to be from either vaginal, oral cavity or an unknown source.

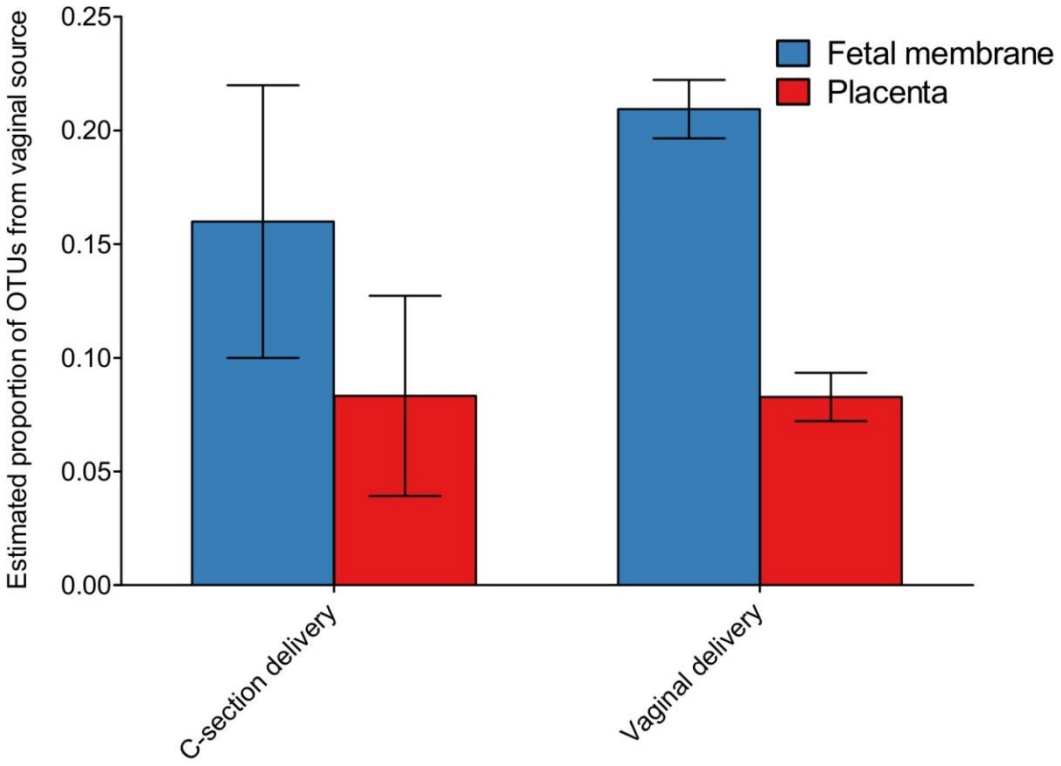

Figure 12-14: The mean ± SEM estimated proportion of OTUs in placenta and fetal membranes sourced from the vagina stratified by whether they were delivered vaginally or by C-section.

Participants were clustered into groups by whether they had these 14 vaginally sourced OTUs within either their placenta or fetal membrane tissues (Figure 12-15). In one cluster, participants had evidence of multiple OTUs present in the vagina and in the placenta. In a second cluster participants had no evidence of these organisms within the vagina or placenta and a third cluster consisted of participants who had a combination of these 14 organisms in the vagina but not in the placenta.

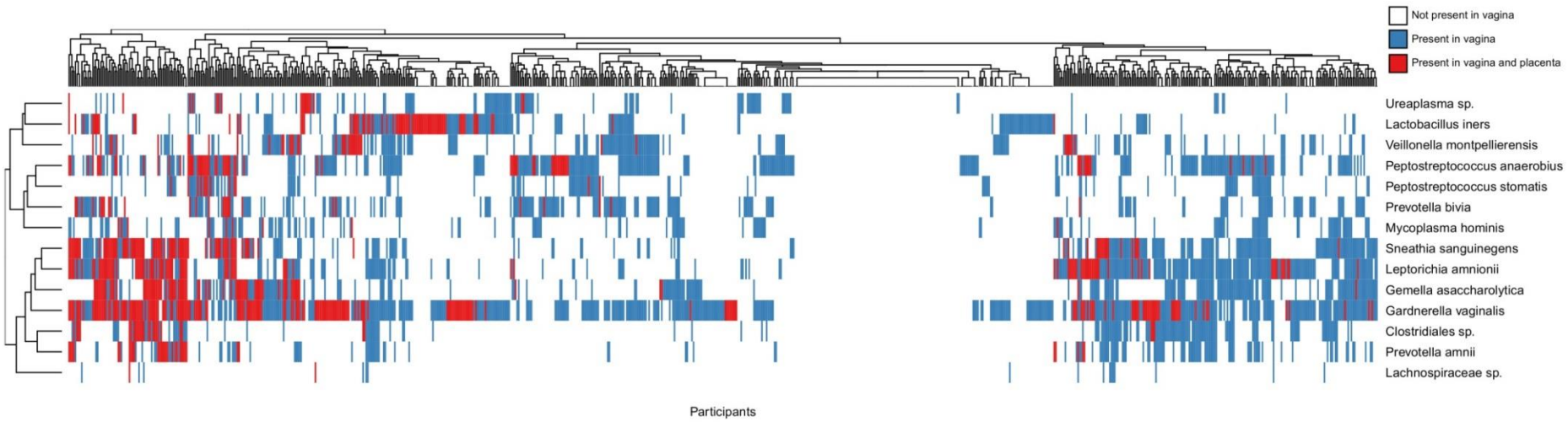

Figure 12-15: Presence of vaginal organisms in the placenta across all participants.
Hierarchical clustering was computed using complete linkage of Euclidean distances. Individuals were ranked by whether an OTU could be identified in both a matched vaginal and placental sample, only in the vaginal and sample or in neither the vagina nor the placenta.

Individuals were grouped by whether they had a detectable presence of these OTUs in their vagina and placenta and the mean LAZ score was taken for each group (Figure 12-16). The mean LAZ score was lowest in individuals when the organisms were present in both the vagina and placenta, with the highest mean LAZ score when the organisms were not present at all. This trend was statistically significant when *Sneathia sanguinengens* and *Peptostreptococcus anaerobius* were detected in the vagina and placental tissues ($q<0.05$). There were no significant associations between the presence of any of the same organisms and duration of pregnancy, birth weight or neonatal head circumference.

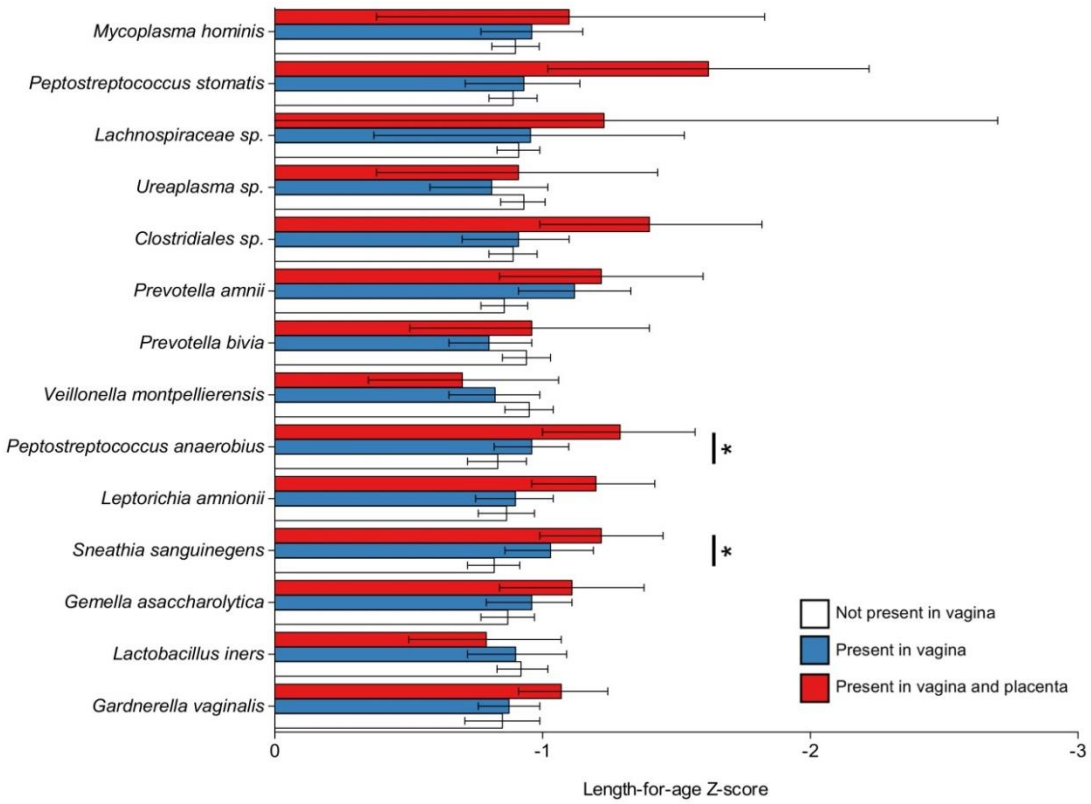

Figure 12-16: LAZ score for individuals grouped by presence of vaginal OTUs.

Mean LAZ scores were plotted with error bars representing 95% confidence intervals. Individuals were group by whether an OTU could be identified in both a matched vaginal and placental sample, only in the vaginal and sample or in neither the vagina nor the placenta. Mean values across the three groups were compared using one-way ANOVA (*q<0.05).

12.4 Discussion

This is the largest study to date of placental microbiota with over 1000 individuals with tissue available. Bacteria were found in more than 50% of placental tissues and the core microbiome was distinct from the oral and vaginal microbiome. This microbial community structure was altered in individuals with severe chorioamnionitis. The abundance of these altered bacteria were associated with birth outcomes and the predominant source for these bacteria was from the vagina.

A study by Aagaard et al (80) of 320 placental samples was recently published and showed that the taxa in the placenta resembled the oral microbiome more than the vagina, gut or skin. They found that certain genera had increased abundance in those who delivered preterm compared to term, including *Burkholderia*, *Anaeromyxobacter*, *Streptosporangium* and *Roseovarius*. These genera have not been previously associated with preterm birth and include environmental organisms or those associated with reagent contamination of sequencing studies (84). Samples with low bacterial biomass, such as from the placenta, are susceptible to reagent, delivery and environmental contamination. Using positive and negative sequencing controls and controlling for the effects of mode of delivery, I have tried to mitigate these effects. The majority of full-thickness placental tissue sections had no detectable bacterial DNA and comparable numbers of bacteria from the vagina in both caesarean sections and vaginally delivered samples. This suggests that contamination of samples during delivery and processing occurred at a relatively low level. The majority of individuals had

an OTU rich placental microbiome, containing OTUs that could not be found in either matched or unmatched vagina or oral cavity samples. A minority of individuals had a relatively small number of phylogenetically diverse OTUs that could also be found in matched vaginal samples and it seems likely that these features represent ‘true’ infection of the placental tissue.

Many of the OTUs I identified as being inversely correlated with duration of delivery have previously been associated with preterm birth. *Fusobacterium nucleatum* and *Ureaplasma spp.* have frequently been found in the amniotic fluid of women who deliver preterm (155,156,58). Although not commonly reported, recent molecular studies have found *Sneathia sanguinegens, Prevotella* spp*., Peptostreptococcus* spp*. and Gardnerella vaginalis* in both amniotic fluid and placental tissues of women who have delivered prematurely (57,63). Interestingly, several OTUs associated with a shorter duration of pregnancy and smaller newborn size in our study have been previously associated with bacterial vaginosis such as *Gardnerella* spp*., Ureaplasma* spp., *Sneathia* spp., *Prevotella spp.* and *Lachnospiraceae* spp*.* (157,158). This includes a recent study that found a higher abundance of *Gardnerella* and *Ureaplasma* in the vaginas of women during pregnancy who subsequently delivered preterm (159). I found that the presence of *Peptostreptococcus anaerobious and Sneathia sanguinegens* in both the vagina and placenta was associated with smaller newborn size, which provides further evidence supporting the view that ascending infection, and hence the vaginal microbiome, may play an important role in birth outcomes. However it is not possible to exclude a role for the oral microbiome in

seeding placental tissue, as has been previously suggested (78), as I have found the same OTUs in the oral cavity and placenta for some individuals. Also, the oral sample collected in this study was a swab of the gingival margins and may not isolate the bacteria found in deep periodontal pockets or periapical tissues that I now know have an association with adverse birth outcomes in this cohort (160).

Severe chorioamnionitis was associated with a shorter duration of pregnancy. Moreover, certain bacterial phyla were associated with chorioamnionitis and also with each other. Based on these distinct microbial profiles, participants enrolled in this study fell broadly into two groups: one group dominated by species in the Proteobacteria or Firmicutes phyla and a second group dominated by a set of phylogenetically different organisms, mainly from the Fusobacteria and Tenericutes phyla. It was the second group who were more likely to have widespread inflammation in their placental tissue, deliver early and deliver a smaller newborn. In particular, higher loads in the load of these two phyla that appeared to be associated with severe chorioamnionitis.

While there are no comparable data from Africa, studies across Europe and North America have shown some similarity between the organisms I have identified and those identified in amniotic fluid (60,161) or placental tissues (64,162). In a different trial in Malawi, prospectively treating women with antibiotics during pregnancy had a protective effect on the incidence of preterm birth and low birth weight (163). This could be explained by the possible clearing of potential pathogens responsible for ascending infection

that could then be leading to preterm birth or growth restriction. The gut microbiome is known to differ significantly between western and African populations (164) and I found a minority of OTUs such as a *Blautia sp., Phascolarctobacterium succinatutens* and a *Lachnospiraceae sp.* that have not previously been associated with adverse birth outcomes. It is unknown at the moment whether these are faecal contaminants or regional differences between this study and previous studies.

In summary, I have identified a distinct microbial community in the placenta and fetal membranes that are associated with severe chorioamnionitis and poor birth outcomes in a large cohort in rural Malawi. Bacteria associated with both severe chorioamnionitis and poor birth outcomes were phylogenetically diverse and not the most abundant taxa recovered in the placenta or fetal membranes. Interestingly, none of the species associated with preterm birth were significantly associated with any of the three measures of growth restriction. Further studies are needed to elucidate mechanisms by which bacteria restrict fetal growth without triggering chorioamnionitis or preterm labour. Previous studies on the use of antibiotics in treating infection as a cause of preterm delivery have had limited success (165,166). A therapeutic approach targeted against the types of bacteria associated with poor birth outcomes may prove more successful, this is supported by trials using clindamycin that have reduced the risk of preterm birth and late miscarriage (167). Strategic control of the microbiome resident in the vagina or oral cavity, using antibiotics or probiotics with proven efficacy

against organisms identified in this study could potentially control potential etiologic agents that spread to the placenta.

13 Chapter 6: The vaginal microbiome in rural Malawi.

13.1 Introduction

Abnormal vaginal microbiota is an alteration in the microbial community that is otherwise predominated by lactobacilli species. Previous studies have found individuals lacking an abundance of lactobacilli are more likely to develop other conditions such as bacterial vaginosis (BV) or aerobic vaginitis (AV) and have an increased risk in transmission of sexually transmitted infections (168). BV is defined by an overabundance of anaerobic organisms such as *Gardnerella vaginalis*, *Prevotella* spp. and *Bacteroides* spp. (169) and has previously been identified by a change in vaginal pH, vaginal discharge and a fishy odour. However symptoms can be rare and many women that have BV are asymptomatic. Another vaginal microbiota disorder is aerobic vaginitis, defined by the overabundance of aerobic bacteria such as *Escherichia coli Staphylococcus* spp. and group B Streptococci (170). It differs from BV not only in the aerobic environment in creates in the vaginal microbiota but also correlates with increased levels of toxic leukocytes and parabasal cells in the epithelium. This strong inflammatory response can lead to itching and burning sensations but has also been linked to preterm birth and miscarriage (171).

Incidence of abnormal vaginal microbiota during pregnancy has been associated with adverse birth outcomes, in particular preterm birth

(66,171,172,65,173) which is the largest cause of neonatal deaths worldwide. Incidence of preterm birth is highest in low-income countries like Malawi (152). Amongst a variety of risk factors, infection has been estimated as the probable cause of preterm birth in 25-40% of cases and ascending infection has long been associated as the potential source of organisms invading maternal and fetal tissues (4). Many of the organisms historically involved in intrauterine infections are commonly found in the genital tract (174)

While previous work in diagnosing BV has used scoring of Gram stains (175), recent studies have been using molecular techniques focusing on the amplification of the 16S rRNA gene to elucidate how structure shifts between Lactobacilli dominated and BV communities (87,176). These studies exclusively acquire data from high-income countries, especially the USA. Ethnicity seems to influence the vaginal microbiome (158) but there have been no studies focusing on low-income populations in Africa. Prevalence of BV in Malawi has been studied and found to be as high as 85.5%,diagnosed using Amsel criteria (vaginal fluid pH >4.5, clue cells on microscopy, nonspecific vaginitis discharge and positive potassium hydroxide odour) (177). No attempt has been made to characterise the bacteria involved.

The aim of this study was to characterise the vaginal microbiome of a large cohort of pregnant women in rural Malawi using 16S rDNA amplicon sequencing. I wanted to expand the current knowledge base for the structure of the vaginal microbiome in a low-income, African setting and investigate possible associations between the abundance of certain organisms and birth outcomes.

13.2 Materials and Methods

13.2.1 Study design and enrolment

See section 9.1.

13.2.2 Study setting

See section 9.2.

13.2.3 Collection of birth outcome and baseline data

See section 9.3.

13.2.4 Sample collection

See section 9.4.

13.2.5 DNA Extraction

See section 9.5.

13.2.6 16S rDNA amplicon high-throughput sequencing

See section 9.7.

13.2.7 Bioinformatics and statistical analysis

See section 9.8.

13.2.8 Ethical approval and consent

See section 9.9.

13.3 Results

13.3.1 Sample collection

Starting in February 2011 a total of 1391 pregnant women were recruited into the iLiNS-DYAD-M trial, with the last delivery taking place in February 2013. A vaginal microbiome was characterised in 1107 (79.6%) participants. Figure 13-1 shows that of those excluded, 222 were lost to follow up, 11 were excluded due to twin deliveries and 51 failed sequencing.

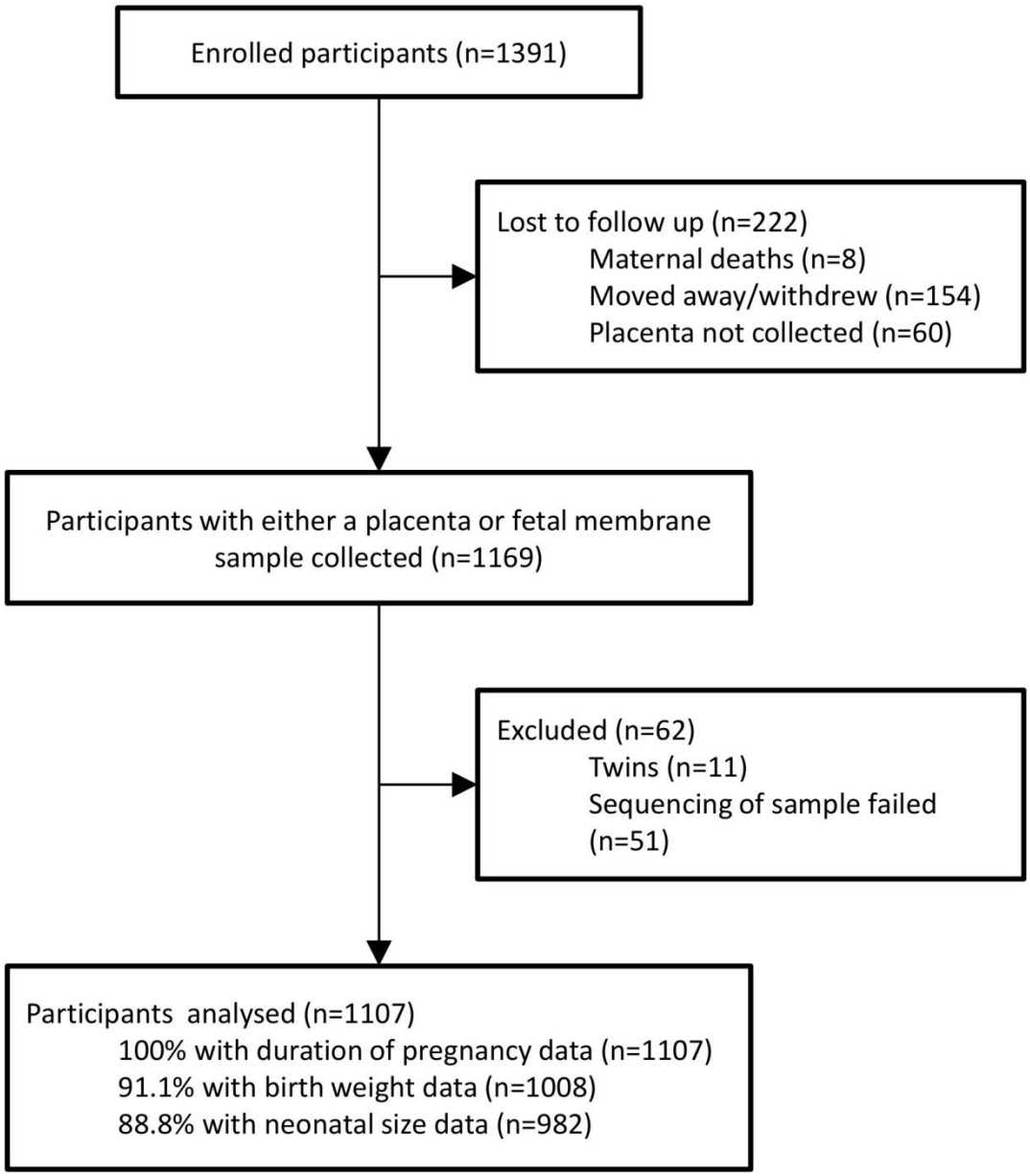

Figure 13-1: Study participant flow diagram.

When the participants that were included in this study were compared to those excluded they tended to be older (25y vs 24y, p=0.025), had completed less time in education (3.9y vs 4.5y, p=0.049), had a lower proxy for socioeconomic status (-0.05 vs 0.30, p<0.001) and were more likely to be primiparous (29.6% vs 20.0%, p=0.001) (Table 13-1). Sequencing 1107

samples from individual participants and filtering for quality and error allowed analysis at a median depth of 14,585 reads per sample (IQR: 7,986 to 21,659) from 18,661,136 reads in total.

Table 13-1: Baseline characteristics of included and excluded participants

Characteristic	Included (n=1107)[1]	Excluded (n=272)[1]	*P* value[2]
Mean (SD) BMI, kg/m^2	22.1 (2.8)	22.4 (2.9)	0.126
Mean (SD) maternal age, years	25.1 (6.10	24.2 (6.6)	0.025
Mean (SD) maternal education, completed years at school	3.9 (3.4)	4.5 (3.7)	0.049
Mean (SD) proxy for socioeconomic status	-0.05 (0.9)	0.30 (1.1)	<0.001
Proportion of aenemic women (Hb < 110 g/l)	29.6 %	20.1 %	0.001
Proportion of primiparous women	4.1 %	5.7 %	0.367
Proportion of women with a low BMI (< 18.5 kg/m^2)	12.2 %	13.9 %	0.591
Proportion of women with a positive HIV test	22.4 %	23.5 %	0.749
Number (%) of women with a positive malaria test (RDT)	23.4 %	22.7 %	0.874

[1]Total n=1391
[2] *P* value obtained from t-test (comparison of means) or Fishers exact test (comparison of proportions)

13.3.2 The vaginal microbiome of rural Malawian women is dominated by *Gardnerella vaginalis* with limited numbers of *Lactobacillus spp.*

Figure 2 shows that *Gardnerella vaginalis* was the most common organism recovered from participant's vaginal swabs. It was prevalent in 75.7% (838/1107) of participant's genital tracts and was the only OTU found in >50% of individuals. In contrast, *Lactobacillus spp.* were found in the vaginal

microbiomes of just 30.4% (337/1107) participants. When *Lactobacillus spp*. were present they were usually *Lactobacillus iners*. A participant positive for the presence of *Lactobacillus iners* was also frequently positive for *Gardnerella vaginalis.* The relative abundances of these two organisms were significantly positively correlated (Spearman's rho = 0.24, $p<0.001$).

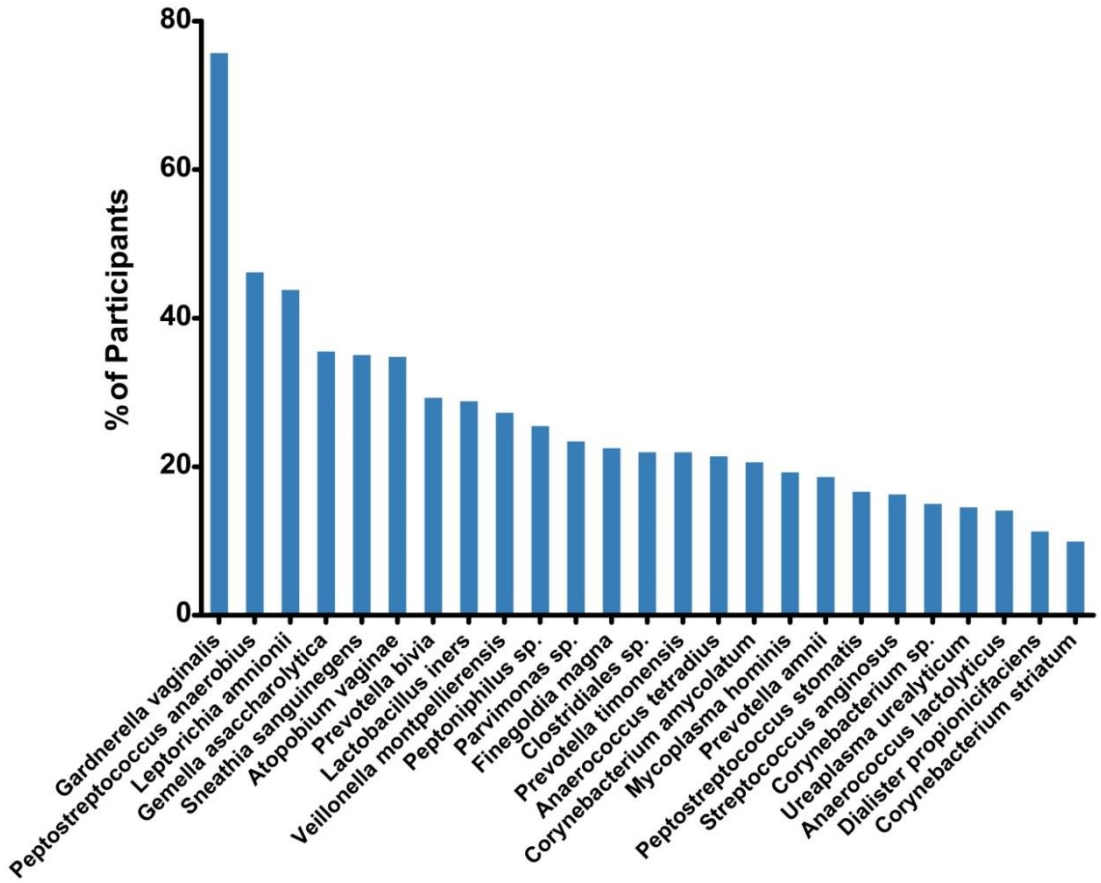

Figure 13-2: Rank abundance curve.
The 25 most prevalent OTUs recovered from participants' vagina (n=1107).

To identify the taxonomic groups that define the vaginal microbiota of this Malawian cohort I compared Spearman's correlations between the 50 most abundant OTUs that accounted for 89.7% of all reads in the dataset. Figure 13-3 shows that these OTUs separated into four taxonomic groups. The first group contained *G. vaginalis, Leptotrichia amnionii, Sneathia sanguinegens Gemella asaccharolytica, Atopobium vaginae, Parvimonas* sp*.* and a *Clostridiales* sp*.* Taxon group one was more likely to be negatively associated with the other three groups, especially taxonomic group two which consisted of *F. magna, S. anginosus, Peptostreptococcus stomatis, Veillonella montpellierensis, Prevotella* spp*., Anaerococcus* spp*. and Corynebacterium spp.. P. anaerobious* did not have a strong positive association with any of the other four taxonomic groups.

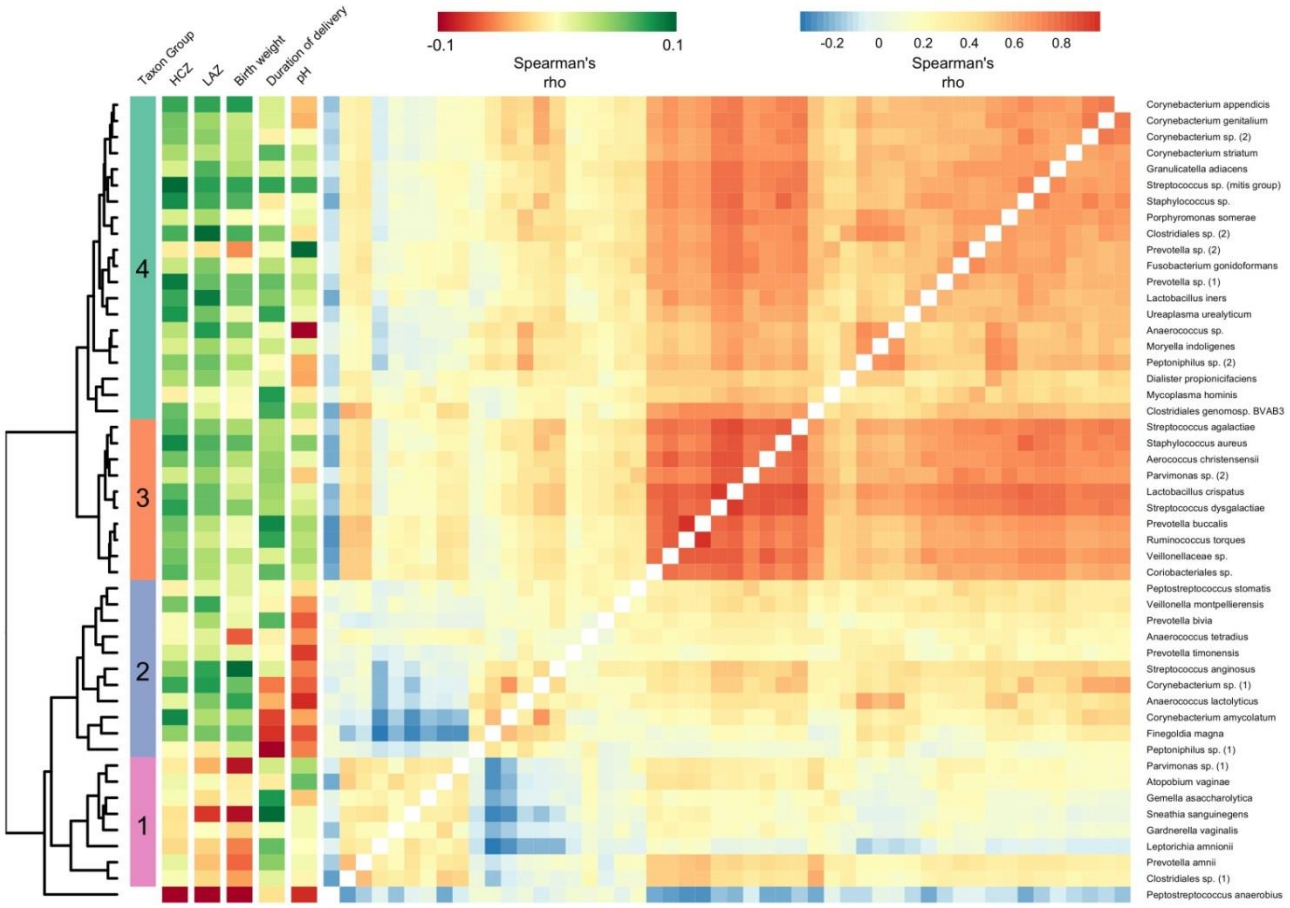

Figure 13-3: Heat map of pairwise Spearman's correlations between the 50 most abundant OTUs.

Hierarchical clustering by complete linkage of Euclidean distances was used to split the OTUs into four taxonomic groups. The heat map is annotated with Spearman's correlation between the abundance of the OTU in that row and either vaginal pH, duration of pregnancy, birth weight, LAZ and HCZ.

To capture the diversity of the vaginal microbiome in each individual and to see if these taxonomic groups also defined groups of individuals, normalised read counts for the top 25 most abundant OTUs were compared and participants were clustered based on their microbiome composition (Figure 13-4). Participants separated into two broad groups which are represented by the separately coloured clades in Figure 13-4. One group was defined by higher abundances of those organisms found in taxon group one and bacterial vaginosis-associated bacterium (BVAB)-3. The second group was more complex, containing a number of different sub-groups defined by higher abundance of *Lactobacillus iners* or clusters of OTUs containing *Corynebacterium* spp*., Ureaplasma urealyticum, Mycoplasma hominis, Streptococcus anginosus, Peptoniphilus* sp*., Finegoldia magna and Anaerococcus* spp*.* Both *Gardnerella vaginalis* and *Peptostreptococcus anaerobious* were present in abundance across a range of individuals and could not be associated with a single participant cluster.

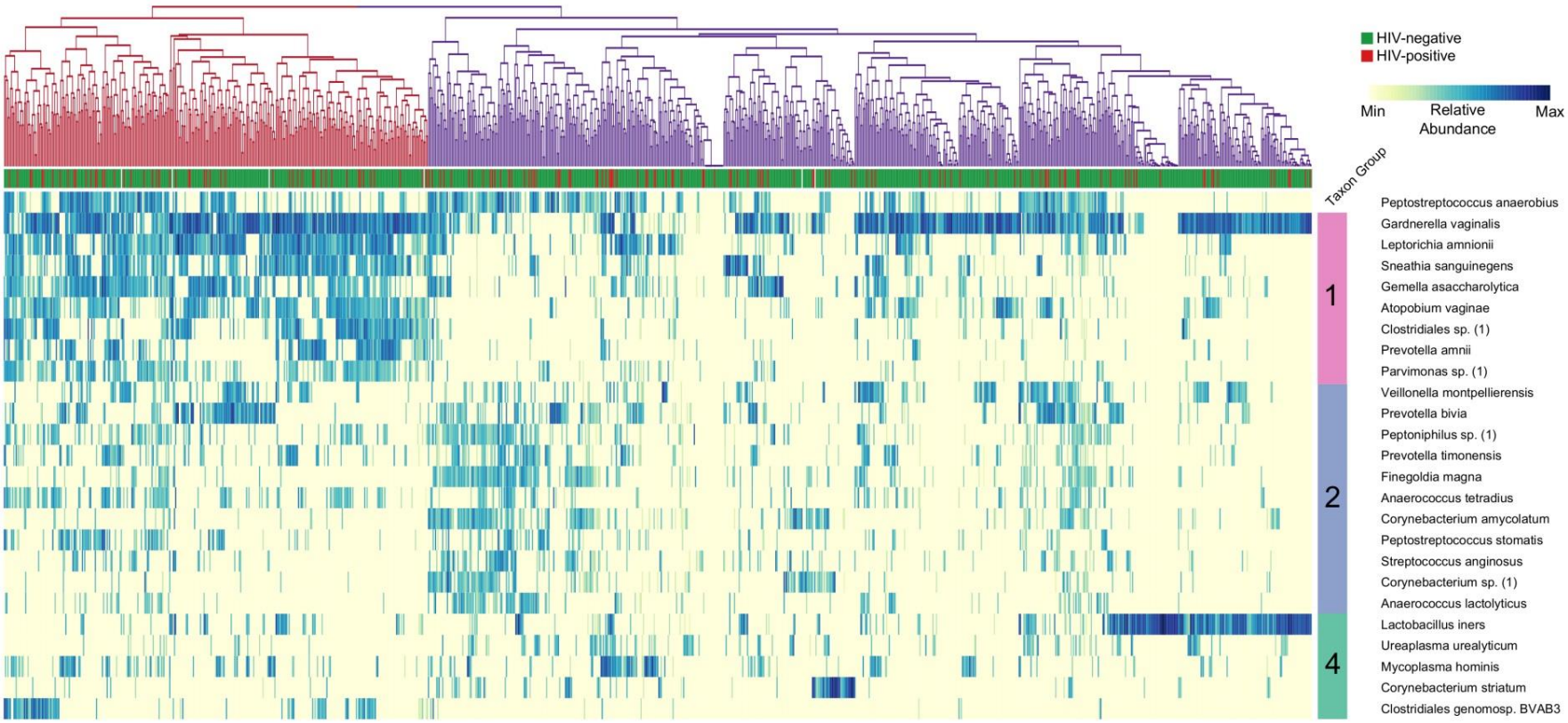

Figure 13-4: Heat map of noramlised read counts from 25 most abundant OTUs in all participants.

Read counts were noramlised using variance stabilizing transformation in DESeq2. Participants are clustered on their microbiota composition using hierarchical clustering by complete linkage of Euclidean distances. OTUs are group by their taxonomic group identified in Figure 13-3. The heat map is annotated by whether that participant was HIV-positive or HIV-negative with the annotation left white if no information was available.

Taxon group one was more likely to dominate the entirety of an individual's vaginal microbiome, unlike the other taxon groups and *P. anaerobius* that often co-existed within the same individual (Figure 13-5). The core microbiome of 46.4% (514/1107) individuals was dominated by taxon group one (>50% relative abundance of taxon group one). In comparison only 16.4% of individuals were dominated by taxon group four, 7.1% by taxon group two and 2.9% by taxon group three.

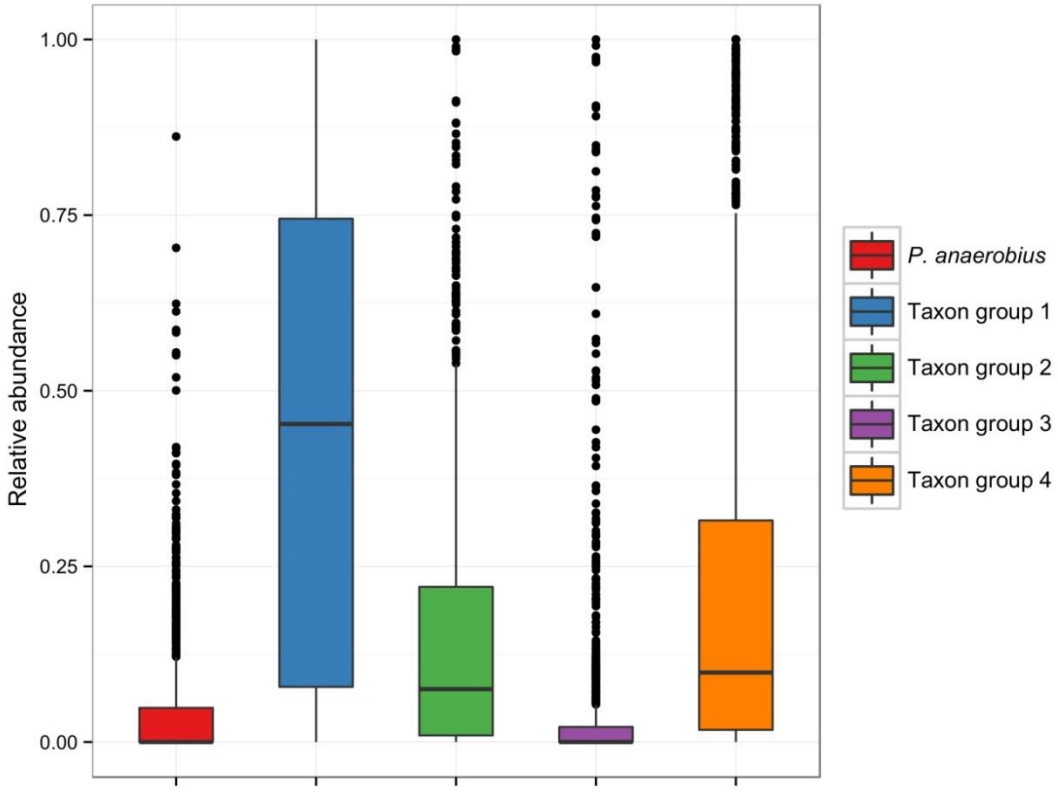

Figure 13-5: Distribution of relative abundances of each taxon group amongst all participants.
Taxon groups are the same as identified in Figure 13-4. Boxes define the taxon group's median relative abundance along with the 25th and 75th percentiles. Error bars plot the 10th and 90th percentiles in the data with all outliers beyond this plotted as individual points.

13.3.3 Factors that may influence or could be influenced by vaginal microbial composition.

13.9 % (154/1107) of individuals in this study were HIV positive. The heat map in Figure 13-4 was annotated with an individual's HIV status to compare its possible association with microbial composition. HIV infection did not explain the observed structure of vaginal OTUs amongst participants. However, I compared differential abundances between OTUs in HIV-negative and HIV-positive participants using log-binomial regression. The two greatest differences in OTU abundance were in *M. hominis* and BVAB3 with a Log2-fold increase in abundance of 0.77 and 0.58 in HIV-positive compared to HIV-negative participants, respectively. Both were not statistically significant changes after adjusting for FDR (q=0.05).

Abundances of each OTU were compared using Spearman's correlation with vaginal pH, gestational age at delivery, birth weight, newborn length-for-age z-score (LAZ) and head circumference-for-age z-score (HCZ) (Figure 13-3). Distinct patterns were seen between groups one and two with pH and birth outcomes. Increased abundances of taxonomic group one and *P. anaerobious* correlated with lower birth weight, lower newborn LAZ and lower HCZ. Higher abundances of taxonomic group 2 OTUs correlated with a lower pH and lower gestational age at birth. Taxonomic groups three and four, as a whole, had no distinct negative association with birth outcomes.

13.3.4 Associations between individual OTUs and adverse birth outcomes

Using the taxonomic groups identified in Figure 13-3, I compared the association of individual OTUs within each group to four birth outcomes. There was a general trend across all individual OTUs from taxonomic group 1 and *P. anaerobius* that an increased abundance corresponded with a lower birthweight, newborn LAZ and newborn HCZ. Increased abundances of *P. anaerobius* and *P. amnii* were both significantly associated with a lower newborn length for age z-score (q=0.023 and q=0.036, respectively) (Table 2). When the association of individual OTUs found in taxonomic group 2 were compared to vaginal pH and duration of pregnancy, although the same trends were seen as for group 2 as a whole, there were no statistically significant associations for individual OTUs seen after adjusting for false discovery rate (supplementary table 1).

Table 13-2: Association between birth weight, newborn LAZ, HCZ and OTUs belonging to taxonomic group one including *P. anaerobius*.

	Outcome					
	Birth weight, grams (n=1008)		Newborn length-for-age Z score (n=979)		Newborn head circumference-for-age Z score (n=982)	
OTU	Adjusted regression coefficient[2] (95% CI)	*q* value[1]	Adjusted regression coefficient[2] (95% CI)	*q* value[1]	Adjusted regression coefficient[2] (95% CI)	*q* value[1]
Peptostreptococcus anaerobius	-7 (-14, 0)	0.306	-0.03 (-0.04, -0.01)	0.023	-0.02 (-0.04, -0.01)	0.023
Parvimonas sp.	-9 (-18, 1)	0.279	-0.02 (-0.04, 0.01)	0.389	-0.02 (-0.04, 0.01)	0.389
Atopobium vaginae	-0 (-8, 7)	0.922	-0.00 (-0.02, 0.02)	1.000	-0.01 (-0.03, 0.01)	1.000
Gemella assacharolytica	3 (-5, 10)	0.722	-0.01 (-0.02, 0.01)	0.831	-0.00 (-0.02, 0.02)	0.831
Sneathia sanguinegens	-6 (-13, 1)	0.250	-0.02 (-0.04, -0.00)	0.057	-0.01 (-0.03, 0.01)	0.057
Gardnerella vaginalis	-2 (-8, 4)	0.735	-0.00 (-0.02, 0.02)	0.989	-0.01 (-0.03, 0.01)	0.989
Leptotrichia amnionii	-1 (-8, 6)	0.849	0.00 (-0.01, 0.02)	0.872	-0.01 (-0.02, 0.01)	0.872
Prevotella amnii	-10 (-19, -1)	0.297	-0.03 (-0.05, -0.01)	0.036	-0.02 (-0.04, 0.01)	0.036
Clostridiales sp.	-4 (-12, 5)	0.679	-0.01 (-0.03, 0.01)	0.526	-0.01 (-0.03, 0.02)	0.526

[1]*P* values calculated by linear regression and *q* values were generated by controlling for the false discovery rate using the Benjamini-Hochberg method. Regression models show the change in birth outcome measured with an increase in OTU's noramlised read count of one. All models were adjusted for nutritional intervention, maternal BMI at enrolment, maternal age, proxy for socioeconomic status, number of previous pregnancies, maternal anaemia at enrolment and site of enrolment.

Table 13-3: Association between vagina pH, duration of pregnancy and OTUs belonging to taxonomic group two.

	Outcome			
	Vaginal pH (n=942)		Duration of pregnancy, weeks (n=1107)	
OTU	Adjusted regression coefficient[2] (95% CI)	*q* value[1]	Adjusted regression coefficient[2] (95% CI)	*q* value[1]
Peptinophilus sp.	-0.00 (-0.03, 0.02)	0.881	-0.02 (-0.08, 0.03)	1.000
Finegoldia magna	-0.01 (-0.04, 0.02)	1.000	-0.02 (-0.08, 0.04)	0.803
Corynebacterium amycolatum	-0.00 (-0.03, 0.03)	0.957	-0.02 (-0.08, 0.04)	0.714
Anaerococcus lactolyticus	-0.02 (-0.05, 0.01)	0.449	0.03 (-0.04, 0.09)	1.000
Corynebacterium sp.	-0.02 (-0.05, -0.01)	0.506	-0.02 (-0.08, 0.04)	0.823
Streptococcus anginosus	-0.02 (-0.05, 0.01)	0.385	0.03 (-0.03, 0.09)	1.000
Prevotella timonensis	-0.03 (-0.05, 0.00)	0.297	0.00 (-0.05, 0.05)	0.956
Anaerococcus tetradius	-0.00 (0.03, 0.02)	0.877	0.02 (-0.03, 0.08)	0.939
Prevotella bivia	-0.02 (-0.04, -0.00)	0.517	0.02 (-0.02, 0.06)	1.000
Veillonella montpellierensis	-0.01 (-0.03, 0.01)	0.587	-0.01 (-0.05, 0.04)	0.854
Peptostreptococcus stomatis	-0.01 (-0.04, 0.01)	0.783	-0.01 (-0.07, 0.05)	0.937

[1]*P* values calculated by linear regression and *q* values were generated by controlling for the false discovery rate using the Benjamini-Hochberg method. Regression models show the change in birth outcome measured with an increase in OTU's noramlised read count of one. All models were adjusted for nutritional intervention, maternal BMI at enrolment, maternal age, proxy for socioeconomic status, number of previous pregnancies, maternal anaemia at enrolment and site of enrolment.

13.4 Discussion

In this study I have characterised the vaginal microbiome of 1,107 women in Southern Africa. This represents the largest cross-sectional study of its kind to date and one of the few studies in an African population (149,178,179). I have shown that a diverse microbial community deficient in *Lactobacillus* spp. dominates the vaginal samples in this cohort of women. A proportion of individuals who have delivered babies with lower LAZ and HCZ tend to have a particular combination of bacterial vaginosis associated bacteria which included higher abundances of *P. anaerobius* and *P. amnii*.

Studies that have analysed the vaginal microbiome using 16S rDNA amplicon sequencing have concluded that the majority of the population has a microbiome dominated by *Lactobacillus* spp. and a less prevalent group of BV-associated organisms (87,180,159). These have taken place in predominantly white Caucasian populations from U.S.A or Europe. Ethnic differences between populations in the U.S.A have been studied and there are show some similarities with our findings. Those with African American ancestry have been found to have a higher prevalence of *Prevotella* spp., *Dialister* spp., *Atopobium* spp., *Gardnerella* spp., *Peptoniphilus* spp., *Sneathia sanguinegens*. *Aerococcus* spp, *Finegoldia magna* and a decreased prevalence of *Lactobacillus* spp. when compared to other groups (87). A study explicitly comparing participants with African American and European ancestry found that those with African American ancestry had a "healthy" microbiome dominated by *M. hominis, Aerococcus* spp*, Anaerococcus* spp*, Dialister* spp*, Peptoniphilus* spp*., Parvimonas* spp*,, S.*

sanguinegens , P. amnii, Atopobium spp*. and G. vaginalis*. When *Lactobacillus* spp. were found in an African American cohort, as noted in this study, it was most likely to be *L. iners* (158). Two studies in East Africa that have used marker gene studies to characterise the vaginal microbiome have focused on either HIV-positive women and those with BV and so the results cannot be extrapolated to see how the population looks as a whole (178,179). As well as ethnic differences, regional microbiome differences have previously been established in gut studies (164), however comparable data for the vagina has not been available until now. Other environmental factors such as vaginal douching are common to this region (181). However there is some disagreement on whether this can (182,183), or cannot (179), alter the vaginal microbiota.

This cross-sectional study sampled the vaginal microbiome at one week post-partum. During pregnancy, longitudinal studies have found that specific community structures are more stable than others (184) and that as pregnancy progressed there was an increase in abundance of *Lactobacillus spp.* (180). However, studies looking at the post-partum vaginal microbiome showed an abrupt drop in abundance of *Lactobacillus* spp and an increase in *Prevotella* spp*.*, *Anaerococcus* spp*., Streptococcus* spp and *Peptoniphilus* spp (159,185). It is known that oestrogen levels post-partum change substantially and fall up to a 1000-fold in some cases, having a significant effect on bacterial community members. A recent study sampled the vaginal microbiome in an African population in Tanzania before and after delivery and found that while during pregnancy the microbiome was generally still

dominated by *Lactobacillus* spp. there was a large shift post-partum (between 3 days and 1 month after birth) (149). Although the numbers of individuals studied were low compared to this study, it showed many similarities with higher abundances of *Prevotella* spp., *Gardnerella* spp. and *Sneathia* spp. However, there was still a relative high prevalence of *Lactobacillus spp*. post-partum in the Tanzanian cohort compared to this Malawian cohort. This could be explained by the lower incidences of adverse birth outcomes such as preterm birth (5%) compared to the population in this study (10%). An extensive longitudinal study is needed to confirm both the stability of these community structures over time and shifts in microbial composition pre- and post-partum.

Although BV has previously been associated with preterm birth, I found the majority of participants in this study were dominated with BV-associated bacteria, and similar to another large study (180), they could not all explain incidences of adverse birth outcome. I found that BVAB-3, previously associated with BV patients, was positively associated with a longer duration of delivery. Interestingly, in an African American population, an increased bacterial load of BVAB3 identified by qPCR was also associated with a decreased risk of preterm birth (186). BV is more prevalent in Malawi than in the U.S.A and Europe, with incidences between 35% and 85.5% previously diagnosed using a mixture of Amsel and Nugent (based on the observation of specific morphotypes under the microscope after Gram stain) methods (177,181,187). Taxonomic groups one and two contain anaerobic bacteria that could represent an altered BV microbiota in this population. I found that

Prevotella amnii and *Peptostreptococcus anaerobius* were both found to be significantly associated with a measure of intrauterine growth restriction. Interestingly, it has been shown that *Prevotella* spp. provide amino acids that increase the growth of *P. anaerobius in vitro* and it seems likely that similar symbiotic relationships exist between the species within the groupings identified in this study (188).

Taxonomic group three could represent a mixed community of both BV and AV flora, with previously identified AV flora such as *Staphylococcus aureus* and *Streptococcus agalactiae* present in high abundance alongside BV-associated genera such as *Prevotella* spp, *Parvimonas* spp and *Coriobacteriales* spp. (170). Although AV has been previously associated with preterm birth (171), I found no associations with birth outcomes measured in this study and the prevalence of these organisms across the entire population was very low.

Administration of antibiotics in pregnancy in Africa have been shown to reduce the risk of preterm birth and low birth weight (163,189,190), presumably by eradicating certain vaginal bacteria and reducing the risk of ascending intrauterine infection. Microbial therapy or administration of probiotics could promote the growth of commensal microbiota but a better understanding of how the vaginal microbial community differs in these populations is needed first to inform any future interventions.

14 Chapter 7: Conclusion

In this thesis I have demonstrated a successful method for identifying the microbial community structure of the placenta and vagina from a rural, low-income setting. Through using this method I have shown distinct community structures in the placenta and vagina are associated with chorioamnionitis and adverse birth outcomes. When interpreting the placental high-throughput sequencing results, one the biggest challenges was differentiating genuine signal from contamination. There were a number of different sources of contamination in this dataset and it was a challenging setting to carry out sensitive bacterial DNA detection. Potential sources of contamination included during deliveries in hospital and health centres that might not have sterile equipment and the threat of constant electricity black-outs during the storage of samples at local sites. It might not be feasible to avoid all possible contamination but if the correct positive and negative controls were used I have shown it is possible to filter potential noise from the dataset post-sequencing. While a study in the UK could be designed to avoid these possible pitfalls, this might be the only workaround in a rural sub-Saharan African setting.

It is still undecided whether the placenta is truly a sterile site or not. Results from sequencing low biomass samples such as these can be heavily effected by contamination after the sample is collected and after the genomic DNA is extracted. High-throughput sequencers are now able to process thousands of samples to a depth spanning millions of reads which gives it unprecedented sensitivity if viable DNA is loaded onto the machine. Previous studies using

molecular techniques to detect bacteria in the placenta have used less sensitive techniques (63). However a recent high-throughput sequencing study has suggested that all placentas have a resident microbiome (80). As mentioned previously, many of the bacteria recovered were environmental organisms, known to be present in a lot of the laboratory reagents used in the preparation and amplification of DNA (84). In this project, I screened samples using a broad-range 16S rDNA qPCR to define samples that would go forward for high-throughput sequencing. Using a less sensitive technique before going ahead with high-throughput sequence would reduce future issues with reagent contamination. This was corroborated by the fact that positive samples with the lowest bacterial loads were both more likely to have a higher relative abundance of environmental organisms that were found in the negative sequencing controls and were more likely to contain higher abundances of the *Sphingomonas* sp. identified in Figure 10-7. The same problem was not encountered when I sequenced samples from established microbial communities such as the vagina and the oral cavity. A screening method such as the broad-range 16S rDNA qPCR might be necessary for low biomass samples such as placental tissue to provide reliable datasets in the future and previous studies not taking this into account need to have their results reassessed in light of this.

Further improvements are needed in reducing PCR error, sequencing error and the reliable assignment of OTUs. At the family level there was a strong association between certain taxa and adverse birth outcomes in this study that was largely lost when comparing individual OTUs. Current OTU picking

strategies that define sequence bins based on global sequence similarity can mistakenly split what should be a single OTU into many spurious OTUs as previously shown in Figure 10-9. Some progress is being made with a move away from global thresholds to a more iterative, *de novo* approach (191). Another largely understudied problem in this field is also the use of large, curated 16S rDNA databases in the assignment of both OTUs and taxonomy to samples. Although proper comparison studies are needed to elucidate the exact effects, it seems that there is a trade off in attempting to represent all bacterial environments and the ability to successfully separate sequences that could originate from biologically different organisms. Computational requirements are drastically reduced when using a restricted database and although there is time saved on using pre-curated databases, a two-step OTU picking approach can be used with a database tailored to the study question. This might be more suited to current OTU and taxonomy assignment algorithms.

Although not included in the scope of this thesis, further work will be needed to elucidate any minor alterations the nutrient interventions might be having on various organism found in the placenta, vagina and oral cavity. All comparisons between bacteria and birth outcomes in this thesis were controlled for the effect of the intervention using relevant regression models. However comparisons between phylogenetic diversity in the various microbiomes between different intervention arms would only reveal very broad shifts in community structure. If lipid-based nutrient supplements are going to be sold locally in Malawi to reduce the prevalence of chronic

undernutrition, then further work is needed to characterise its impact on the microbiome and especially oral health.

Collaborators on this project have published that periapical infections in the same cohort were associated with both shorter duration of pregnancy and IUGR (160). This would provide further evidence for a case of haematogenous spread of bacteria from the mouth to the placenta and perhaps infection of the foetus itself. We only reliably detected two oral organisms that could be traced in both matched oral and placental samples. This included Group B Streptococci which has previously been implicated in preterm birth (192,193) but was not associated with adverse birth outcomes in this population. It is known that many of the more problematic oral pathogens, like *Fusobacterium nucleatum,* are anaerobes found in increasing numbers at greater pocket depth (194). *F. nucleatum* was detected at relatively high prevalence in placental tissue in this study and at the family level. Greater abundances of *Fusobacteriaceae* were also associated with both a higher prevalence of severe chorioamnionitis and a lower duration of delivery. This pathogen has long been associated with adverse birth outcomes (61); however the prevalence in the oral samples was very low. Many participants had a placental tissue sample dominated by *F. nucleatum* that was completely absent from a matched participant oral swab. One explanation would be that the *F. nucleatum* we sampled was being introduced from a different source. However, I think a more likely reason for this discrepancy is the type of sample taken in this study. It is more likely the full diversity of oral bacteria in these participants was not identified. Any

future studies need to sample at multiple sites, including deep periodontal pockets, in order to discover reliable associations between oral pathogens and adverse birth outcomes.

Based on the results of this thesis, a therapeutic intervention (either antibiotics or probiotics) in this population that had proven efficacy against *P. anaerobious, P. amnii* and *S. sanguinegens* early in pregnancy might be an ideal next step to reduce the prevalence of newborn stunting and low birth weight. The ability to screen for these organisms using species-specific PCR might also be useful in targeting at-risk populations. However, I think further studies are needed to confirm these associations. Longitudinal sampling of the vagina is needed at regular intervals throughout pregnancy and after delivery. Unfortunately the only available vaginal microbiome time point in this study was one week post-partum, and as previously mentioned, a wide variety of changes make it difficult to draw conclusions as to microbial community structure pre-partum. The stark contrasts however between the results between different ethnic groups in both prevalence of preterm birth, BV and even healthy vaginal tract microbiota suggest that a greater variation in community structure than perhaps previously thought. Despite this, these results provide further evidence of the important role the vaginal microbiome may play in seeding organisms found on placental tissues in rural Malawi and therapeutic interventions could be designed to impact the microbiome with the goal of reducing the risk of preterm birth or intrauterine growth restriction.

15 References

1. Slattery MM, Morrison JJ. Preterm delivery. The Lancet. 2002 Nov 9;360(9344):1489–97.

2. Lawn JE, Cousens S, Zupan J. 4 million neonatal deaths: when? Where? Why? Lancet. 2005 Mar 5;365(9462):891–900.

3. Lawn JE, Blencowe H, Oza S, You D, Lee ACC, Waiswa P, et al. Every Newborn: progress, priorities, and potential beyond survival. Lancet Lond Engl. 2014 Jul 12;384(9938):189–205.

4. Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. Lancet. 2008 Jan 5;371(9606):75–84.

5. McParland P, Jones G, Taylor D. Preterm labour and prematurity. Curr Obstet Gynaecol. 2004 Oct;14(5):309–19.

6. Ergaz Z, Avgil M, Ornoy A. Intrauterine growth restriction—etiology and consequences: What do we know about the human situation and experimental animal models? Reprod Toxicol. 2005 Sep;20(3):301–22.

7. Williams RL, Creasy RK, Cunningham GC, Hawes WE, Norris FD, Tashiro M. Fetal growth and perinatal viability in California. Obstet Gynecol. 1982 May;59(5):624–32.

8. Institute of Medicine (US) Committee on Understanding Premature Birth and Assuring Healthy Outcomes. Preterm Birth: Causes, Consequences, and Prevention [Internet]. Behrman RE, Butler AS, editors. Washington (DC): National Academies Press (US); 2007 [cited 2015 Aug 31]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK11362/

9. Black RE, Victora CG, Walker SP, Bhutta ZA, Christian P, de Onis M, et al. Maternal and child undernutrition and overweight in low-income and middle-income countries. Lancet Lond Engl. 2013 Aug 3;382(9890):427–51.

10. Sohl B, Moore T. Abnormalities of Fetal Growth. In: Avery's Diseases of the Newborn. 7th Edition. Philadelphia: Elsevier Health Sciences; 2005. p. 90–102.

11. Beck S, Wojdyla D, Say L, Betran AP, Merialdi M, Requejo JH, et al. The worldwide incidence of preterm birth: a systematic review of maternal mortality and morbidity. Bull World Health Organ. 2010 Jan;88(1):31–8.

12. Martin JA, Hamilton BE, Ventura SJ, Osterman MJK, Kirmeyer S, Mathews TJ, et al. Births: final data for 2009. Natl Vital Stat Rep Cent Dis Control Prev Natl Cent Health Stat Natl Vital Stat Syst. 2011 Nov 3;60(1):1–70.

13. Joanne Evans CHAT. Gestation-specific infant mortality in England and Wales, 2010 [Internet]. Office for National Statistics. 2012 [cited 2013 Apr 15]. Available from: http://www.ons.gov.uk/ons/rel/child-health/gestation-specific-infant-mortality-in-england-and-wales/2010/gestation-specific-infant-mortality-in-england-and-wales--2010.html

14. Simmons LE, Rubens CE, Darmstadt GL, Gravett MG. Preventing preterm birth and neonatal mortality: exploring the epidemiology, causes, and interventions. Semin Perinatol. 2010 Dec;34(6):408–15.

15. Luntamo M, Kulmala T, Mbewe B, Cheung YB, Maleta K, Ashorn P. Effect of repeated treatment of pregnant women with sulfadoxine-pyrimethamine and azithromycin on preterm delivery in Malawi: a randomized controlled trial. Am J Trop Med Hyg. 2010 Dec;83(6):1212–20.

16. Rajaratnam JK, Marcus JR, Flaxman AD, Wang H, Levin-Rector A, Dwyer L, et al. Neonatal, postneonatal, childhood, and under-5 mortality for 187 countries, 1970-2010: a systematic analysis of progress towards Millennium Development Goal 4. Lancet. 2010 Jun 5;375(9730):1988–2008.

17. National Statistical Office, Zomba, Malawi, ICF Macro, Calverton, Maryland, USA. Malawi Demographic and Health Survey 2010 [Internet]. 2011 [cited 2013 Apr 15]. Available from: http://www.measuredhs.com/what-we-do/survey/survey-display-333.cfm

18. Gladstone M, White S, Kafulafula G, Neilson JP, van den Broek N. Post-neonatal Mortality, Morbidity, and Developmental Outcome after Ultrasound-Dated Preterm Birth in Rural Malawi: A Community-Based Cohort Study. PLoS Med. 2011 Nov;8(11).

19. Liu L, Johnson HL, Cousens S, Perin J, Scott S, Lawn JE, et al. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. Lancet. 2012 Jun 9;379(9832):2151–61.

20. Romero R, Dey SK, Fisher SJ. Preterm labor: one syndrome, many causes. Science. 2014 Aug 15;345(6198):760–5.

21. Romero R, Espinoza J, Kusanovic JP, Gotsch F, Hassan S, Erez O, et al. The preterm parturition syndrome. BJOG Int J Obstet Gynaecol. 2006 Dec 1;113:17–42.

22. Goldenberg RL, Goepfert AR, Ramsey PS. Biochemical markers for the prediction of preterm birth. Am J Obstet Gynecol. 2005 May;192(5 Suppl):S36–46.

23. Taha TE, Dadabhai SS, Rahman MH, Sun J, Kumwenda J, Kumwenda NI. Trends in birth weight and gestational age for infants born to HIV-infected, antiretroviral treatment-naive women in Malawi. Pediatr Infect Dis J. 2012 May;31(5):481–6.

24. Kulmala T, Vaahtera M, Ndekha M, Koivisto AM, Cullinan T, Salin ML, et al. The importance of preterm births for peri- and neonatal mortality in rural Malawi. Paediatr Perinat Epidemiol. 2000 Jul;14(3):219–26.

25. Sullivan AD, Nyirenda T, Cullinan T, Taylor T, Harlow SD, James SA, et al. Malaria infection during pregnancy: intrauterine growth retardation and preterm delivery in Malawi. J Infect Dis. 1999 Jun;179(6):1580–3.

26. Larsen JW, Goldkrand JW, Hanson TM, Miller CR. Intrauterine infection on an obstetric service. Obstet Gynecol. 1974 Jun;43(6):838–43.

27. Onderdonk AB, Hecht JL, McElrath TF, Delaney ML, Allred EN, Leviton A. Colonization of second-trimester placenta parenchyma. Am J Obstet Gynecol. 2008 Jul;199(1):52.e1–52.e10.

28. Buhimschi CS, Dulay AT, Abdel-Razeq S, Zhao G, Lee S, Hodgson EJ, et al. Fetal inflammatory response in women with proteomic biomarkers characteristic of intra-amniotic inflammation and preterm birth. BJOG Int J Obstet Gynaecol. 2009 Jan;116(2):257–67.

29. Watts DH, Krohn MA, Hillier SL, Eschenbach DA. The association of occult amniotic fluid infection with gestational age and neonatal outcome among women in preterm labor. Obstet Gynecol. 1992 Mar;79(3):351–7.

30. Pereira L, Petitt M, Fong A, Tsuge M, Tabata T, Fang-Hoover J, et al. Intrauterine Growth Restriction Caused by Underlying Congenital Cytomegalovirus Infection. J Infect Dis. 2014 May 15;209(10):1573–84.

31. Elovitz MA, Wang Z, Chien EK, Rychlik DF, Phillippe M. A new model for inflammation-induced preterm birth: the role of platelet-activating factor and Toll-like receptor-4. Am J Pathol. 2003 Nov;163(5):2103–11.

32. Westover AJ, Moss TJM. Effects of intrauterine infection or inflammation on fetal lung development. Clin Exp Pharmacol Physiol. 2012 Sep;39(9):824–30.

33. Viscardi RM, Muhumuza CK, Rodriguez A, Fairchild KD, Sun C-CJ, Gross GW, et al. Inflammatory markers in intrauterine and fetal blood and cerebrospinal fluid compartments are associated with adverse pulmonary and neurologic outcomes in preterm infants. Pediatr Res. 2004 Jun;55(6):1009–17.

34. Holzman C, Lin X, Senagore P, Chung H. Histologic chorioamnionitis and preterm delivery. Am J Epidemiol. 2007 Oct 1;166(7):786–94.

35. Gotsch F, Romero R, Kusanovic JP, Mazaki-Tovi S, Pineles BL, Erez O, et al. The fetal inflammatory response syndrome. Clin Obstet Gynecol. 2007 Sep;50(3):652–83.

36. Romero R, Espinoza J, Gonçalves LF, Gomez R, Medina L, Silva M, et al. Fetal cardiac dysfunction in preterm premature rupture of membranes. J Matern-Fetal Neonatal Med Off J Eur Assoc Perinat Med Fed Asia Ocean Perinat Soc Int Soc Perinat Obstet. 2004 Sep;16(3):146–57.

37. Watterberg KL, Demers LM, Scott SM, Murphy S. Chorioamnionitis and early lung inflammation in infants in whom bronchopulmonary dysplasia develops. Pediatrics. 1996 Feb;97(2):210–5.

38. Leviton A, Paneth N, Reuss ML, Susser M, Allred EN, Dammann O, et al. Maternal infection, fetal inflammatory response, and brain damage in very low birth weight infants. Developmental Epidemiology Network Investigators. Pediatr Res. 1999 Nov;46(5):566–75.

39. Galinsky R, Polglase GR, Hooper SB, Black MJ, Moss TJM. The consequences of chorioamnionitis: preterm birth and effects on development. J Pregnancy. 2013;2013:412831.

40. Dombroski RA, Woodard DS, Harper MJ, Gibbs RS. A rabbit model for bacteria-induced preterm pregnancy loss. Am J Obstet Gynecol. 1990 Dec;163(6 Pt 1):1938–43.

41. Grigsby PL, Hirst JJ, Scheerlinck J-P, Phillips DJ, Jenkin G. Fetal responses to maternal and intra-amniotic lipopolysaccharide administration in sheep. Biol Reprod. 2003 May;68(5):1695–702.

42. Hirsch E, Wang H. The molecular pathophysiology of bacterially induced preterm labor: insights from the murine model. J Soc Gynecol Investig. 2005 Apr;12(3):145–55.

43. Chellam VG, Rushton DI. Chorioamnionitis and funiculitis in the placentas of 200 births weighing less than 2.5 kg. Br J Obstet Gynaecol. 1985 Aug;92(8):808–14.

44. Mueller-Heubach E, Rubinstein DN, Schwarz SS. Histologic chorioamnionitis and preterm delivery in different patient populations. Obstet Gynecol. 1990 Apr;75(4):622–6.

45. Muglia LJ, Katz M. The enigma of spontaneous preterm birth. N Engl J Med. 2010 Feb 11;362(6):529–35.

46. Andrews WW, Hauth JC, Goldenberg RL, Gomez R, Romero R, Cassell GH. Amniotic fluid interleukin-6: correlation with upper genital tract microbial colonization and gestational age in women delivered after spontaneous labor versus indicated delivery. Am J Obstet Gynecol. 1995 Aug;173(2):606–12.

47. Harwick HJ, Iuppa JB, Fekety FR Jr. Microorganisms and amniotic fluid. Obstet Gynecol. 1969 Feb;33(2):256–9.

48. Bobitt JR, Ledger WJ. Unrecognized amnionitis and prematurity: a preliminary report. J Reprod Med. 1977 Jul;19(1):8–12.

49. Cassell GH, Davis RO, Waites KB, Brown MB, Marriott PA, Stagno S, et al. Isolation of Mycoplasma hominis and Ureaplasma urealyticum from amniotic fluid at 16-20 weeks of gestation: potential effect on outcome of pregnancy. Sex Transm Dis. 1983 Dec;10(4 Suppl):294–302.

50. Romero R, Sirtori M, Oyarzun E, Avila C, Mazor M, Callahan R, et al. Infection and labor. V. Prevalence, microbiology, and clinical significance of intraamniotic infection in women with preterm labor and intact membranes. Am J Obstet Gynecol. 1989 Sep;161(3):817–24.

51. Fox GE, Pechman KR, Woese CR. Comparative Cataloging of 16S Ribosomal Ribonucleic Acid: Molecular Approach to Procaryotic Systematics. Int J Syst Bacteriol. 1977 Jan 1;27(1):44–57.

52. Böttger EC. Rapid determination of bacterial ribosomal RNA sequences by direct sequencing of enzymatically amplified DNA. FEMS Microbiol Lett. 1989 Nov;53(1-2):171–6.

53. Kolbert CP, Persing DH. Ribosomal DNA sequencing as a tool for identification of bacterial pathogens. Curr Opin Microbiol. 1999 Jun;2(3):299–305.

54. Harris KA, Hartley JC. Development of broad-range 16S rDNA PCR for use in the routine diagnostic clinical microbiology service. J Med Microbiol. 2003 Aug;52(Pt 8):685–91.

55. Hitti J, Riley DE, Krohn MA, Hillier SL, Agnew KJ, Krieger JN, et al. Broad-spectrum bacterial rDNA polymerase chain reaction assay for detecting amniotic fluid infection among women in premature labor. Clin Infect Dis Off Publ Infect Dis Soc Am. 1997 Jun;24(6):1228–32.

56. Markenson GR, Martin RK, Tillotson-Criss M, Foley KS, Stewart RS Jr, Yancey M. The use of the polymerase chain reaction to detect bacteria in amniotic fluid in pregnancies complicated by preterm labor. Am J Obstet Gynecol. 1997 Dec;177(6):1471–7.

57. DiGiulio DB, Romero R, Amogan HP, Kusanovic JP, Bik EM, Gotsch F, et al. Microbial prevalence, diversity and abundance in amniotic fluid during preterm labor: a molecular and culture-based investigation. PloS One. 2008;3(8):e3056.

58. Han YW, Shen T, Chung P, Buhimschi IA, Buhimschi CS. Uncultivated bacteria as etiologic agents of intra-amniotic inflammation leading to preterm birth. J Clin Microbiol. 2009 Jan;47(1):38–47.

59. Jalava J, Mäntymaa ML, Ekblad U, Toivanen P, Skurnik M, Lassila O, et al. Bacterial 16S rDNA polymerase chain reaction in the detection of intra-amniotic infection. Br J Obstet Gynaecol. 1996 Jul;103(7):664–9.

60. DiGiulio DB, Romero R, Kusanovic JP, Gómez R, Kim CJ, Seok KS, et al. Prevalence and Diversity of Microbes in the Amniotic Fluid, the Fetal Inflammatory Response, and Pregnancy Outcome in Women with Preterm Pre-Labor Rupture of Membranes. Am J Reprod Immunol. 2010 Jul 1;64(1):38–57.

61. DiGiulio DB. Diversity of microbes in amniotic fluid. Semin Fetal Neonatal Med. 2012 Feb;17(1):2–11.

62. Steel JH, Malatos S, Kennea N, Edwards AD, Miles L, Duggan P, et al. Bacteria and inflammatory cells in fetal membranes do not always cause preterm labor. Pediatr Res. 2005 Mar;57(3):404–11.

63. Jones HE, Harris KA, Azizia M, Bank L, Carpenter B, Hartley JC, et al. Differing prevalence and diversity of bacterial species in fetal membranes from very preterm and term labor. PloS One. 2009;4(12):e8205.

64. Doyle RM, Alber DG, Jones HE, Harris K, Fitzgerald F, Peebles D, et al. Term and preterm labour are associated with distinct microbial community structures in placental membranes which are independent of mode of delivery. Placenta. 2014 Dec;35(12):1099–101.

65. Vogel I, Thorsen P, Hogan VK, Schieve LA, Jacobsson B, Ferre CD. The joint effect of vaginal Ureaplasma urealyticum and bacterial vaginosis on adverse pregnancy outcomes. Acta Obstet Gynecol Scand. 2006;85(7):778–85.

66. Hillier SL, Nugent RP, Eschenbach DA, Krohn MA, Gibbs RS, Martin DH, et al. Association between bacterial vaginosis and preterm delivery of a low-birth-weight infant. The Vaginal Infections and Prematurity Study Group. N Engl J Med. 1995 Dec 28;333(26):1737–42.

67. Alanen A. Polymerase chain reaction in the detection of microbes in amniotic fluid. Ann Med. 1998 Jun;30(3):288–95.

68. Fortner KB, Grotegut CA, Ransom CE, Bentley RC, Feng L, Lan L, et al. Bacteria Localization and Chorion Thinning among Preterm Premature Rupture of Membranes. PLoS ONE. 2014 Jan 8;9(1):e83338.

69. Galask RP, Varner MW, Petzold CR, Wilbur SL. Bacterial attachment to the chorioamniotic membranes. Am J Obstet Gynecol. 1984 Apr 1;148(7):915–28.

70. Goldenberg RL, Hauth JC, Andrews WW. Intrauterine infection and preterm delivery. N Engl J Med. 2000 May 18;342(20):1500–7.

71. Han YW, Ikegami A, Bissada NF, Herbst M, Redline RW, Ashmead GG. Transmission of an uncultivated Bergeyella strain from the oral cavity to amniotic fluid in a case of preterm birth. J Clin Microbiol. 2006 Apr;44(4):1475–83.

72. Gonzales-Marin C, Spratt DA, Millar MR, Simmonds M, Kempley ST, Allaker RP. Levels of periodontal pathogens in neonatal gastric aspirates and possible maternal sites of origin. Mol Oral Microbiol. 2011 Oct;26(5):277–90.

73. Collins JG, Windley HW, Arnold RR, Offenbacher S. Effects of a Porphyromonas gingivalis infection on inflammatory mediator response and pregnancy outcome in hamsters. Infect Immun. 1994 Oct;62(10):4356–61.

74. Yeo A, Smith MA, Lin D, Riché EL, Moore A, Elter J, et al. Campylobacter rectus mediates growth restriction in pregnant mice. J Periodontol. 2005 Apr;76(4):551–7.

75. Arce RM, Barros SP, Wacker B, Peters B, Moss K, Offenbacher S. Increased TLR4 expression in murine placentas after oral infection with periodontal pathogens. Placenta. 2009 Feb;30(2):156–62.

76. Arce RM, Caron KM, Barros SP, Offenbacher S. Toll-like receptor 4 mediates intrauterine growth restriction after systemic Campylobacter rectus infection in mice. Mol Oral Microbiol. 2012 Oct;27(5):373–81.

77. Blanc V, O Valle F, Pozo E, Puertas A, León R, Mesa F. Oral bacteria in placental tissues: Increased molecular detection in pregnant periodontitis patients. Oral Dis. 2015 Aug 10;

78. Gonzales-Marin C, Spratt DA, Allaker R. Maternal oral origin of Fusobacterium nucleatum in adverse pregnancy outcomes as determined using the 16S-23S rDNA intergenic transcribed spacer region. J Med Microbiol [Internet]. 2012 Sep 20 [cited 2012 Sep 25]; Available from: http://www.ncbi.nlm.nih.gov/pubmed/23002071

79. Chambrone L, Guglielmetti MR, Pannuti CM, Chambrone LA. Evidence grade associating periodontitis to preterm birth and/or low birth weight: I. A systematic review of prospective cohort studies. J Clin Periodontol. 2011 Sep;38(9):795–808.

80. Aagaard K, Ma J, Antony KM, Ganu R, Petrosino J, Versalovic J. The Placenta Harbors a Unique Microbiome. Sci Transl Med. 2014 May 21;6(237):237ra65–237ra65.

81. Nakamura S, Nakaya T, Iida T. Metagenomic analysis of bacterial infections by means of high-throughput DNA sequencing. Exp Biol Med Maywood NJ. 2011 Aug 1;236(8):968–71.

82. Glenn TC. Field guide to next-generation DNA sequencers. Mol Ecol Resour. 2011 Sep;11(5):759–69.

83. Pinto AJ, Raskin L. PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. PloS One. 2012;7(8):e43093.

84. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 2014 Nov 12;12(1):87.

85. Schloss PD. The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. PLoS Comput Biol. 2010;6(7):e1000844.

86. Chang JY, Antonopoulos DA, Kalra A, Tonelli A, Khalife WT, Schmidt TM, et al. Decreased diversity of the fecal Microbiome in recurrent Clostridium difficile-associated diarrhea. J Infect Dis. 2008 Feb 1;197(3):435–8.

87. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, et al. Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci U S A. 2011 Mar 15;108 Suppl 1:4680–7.

88. Ashorn P, Alho L, Ashorn U, Cheung YB, Dewey KG, Harjunmaa U, et al. The impact of lipid-based nutrient supplement provision to pregnant women on newborn size in rural Malawi: a randomized controlled trial. Am J Clin Nutr. 2015 Feb 1;ajcn.088617.

89. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 2012 Aug;6(8):1621–4.

90. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinforma Oxf Engl. 2011 Nov 1;27(21):2957–63.

91. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010 May;7(5):335–6.

92. Knights D, Kuczynski J, Charlson ES, Zaneveld J, Mozer MC, Collman RG, et al. Bayesian community-wide culture-independent microbial source tracking. Nat Methods. 2011 Sep;8(9):761–3.

93. Cheung YB. Statistical Analysis of Human Growth and Development. 1 edition. Boca Raton: Chapman and Hall/CRC; 2013. 378 p.

94. Fall CHD, Fisher DJ, Osmond C, Margetts BM, Maternal Micronutrient Supplementation Study Group. Multiple micronutrient supplementation during pregnancy in low-income countries: a meta-analysis of effects on birth size and length of gestation. Food Nutr Bull. 2009 Dec;30(4 Suppl):S533–46.

95. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006 Jul;72(7):5069–72.

96. DeSantis TZ Jr, Hugenholtz P, Keller K, Brodie EL, Larsen N, Piceno YM, et al. NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. Nucleic Acids Res. 2006 Jul 1;34(Web Server issue):W394–9.

97. Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R. Short pyrosequencing reads suffice for accurate microbial community analysis. Nucleic Acids Res. 2007;35(18):e120.

98. Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat Meth. 2008 Mar;5(3):235–7.

99. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol. 2013 Jun 21;

100. Feinstein LM, Sul WJ, Blackwood CB. Assessment of bias associated with incomplete extraction of microbial DNA from soil. Appl Environ Microbiol. 2009 Aug;75(16):5428–33.

101. Okubo T, Ikeda S, Yamashita A, Terasawa K, Minamisawa K. Pyrosequence Read Length of 16S rRNA Gene Affects Phylogenetic Assignment of Plant-associated Bacteria. Microbes Environ JSME [Internet]. 2012 Feb 22 [cited 2012 Feb 28]; Available from: http://www.ncbi.nlm.nih.gov/pubmed/22353764

102. Huber JA, Morrison HG, Huse SM, Neal PR, Sogin ML, Mark Welch DB. Effect of PCR amplicon size on assessments of clone library microbial diversity and community structure. Environ Microbiol. 2009 May;11(5):1292–302.

103. Youssef N, Sheik CS, Krumholz LR, Najar FZ, Roe BA, Elshahed MS. Comparison of species richness estimates obtained using nearly complete fragments and simulated pyrosequencing-generated fragments in 16S rRNA gene-based environmental surveys. Appl Environ Microbiol. 2009 Aug;75(16):5227–36.

104. Engelbrektson A, Kunin V, Wrighton KC, Zvenigorodsky N, Chen F, Ochman H, et al. Experimental factors affecting PCR-based estimates of microbial species richness and evenness. ISME J. 2010 May;4(5):642–7.

105. Soergel DAW, Dey N, Knight R, Brenner SE. Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. ISME J. 2012 Jul;6(7):1440–4.

106. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 2013 Jan 7;41(1):e1.

107. Ghyselinck J, Pfeiffer S, Heylen K, Sessitsch A, De Vos P. The Effect of Primer Choice and Short Read Sequences on the Outcome of 16S rRNA Gene Based Diversity Studies. PloS One. 2013 Aug 19;8(8):e71360.

108. Schloss PD, Gevers D, Westcott SL. Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies. PLoS ONE. 2011 Dec 14;6(12):e27310.

109. Bonder MJ, Abeln S, Zaura E, Brandt BW. Comparing clustering and pre-processing in taxonomy analysis. Bioinformatics. 2012 Nov 15;28(22):2891–7.

110. Lee CK, Herbold CW, Polson SW, Wommack KE, Williamson SJ, McDonald IR, et al. Groundtruthing next-gen sequencing for microbial ecology-biases and errors in community structure estimates from PCR amplicon pyrosequencing. PloS One. 2012;7(9):e44224.

111. José C. Clemente, Jesper Jansson, Gabriel Valiente. Accurate taxonomic assignment of short pyrosequencing reads. In: Biocomputing 2010 [Internet]. WORLD SCIENTIFIC; 2009 [cited 2015 Sep 3]. p. 3–9. Available from: http://www.worldscientific.com/doi/abs/10.1142/9789814295291_0002

112. Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ. Removing noise from pyrosequenced amplicons. BMC Bioinformatics. 2011;12:38.

113. Sergeant MJ, Constantinidou C, Cogan T, Penn CW, Pallen MJ. High-throughput sequencing of 16S rRNA gene amplicons: effects of extraction procedure, primer length and annealing temperature. PloS One. 2012;7(5):e38094.

114. Ahn J-H, Kim B-Y, Song J, Weon H-Y. Effects of PCR cycle number and DNA polymerase type on the 16S rRNA gene pyrosequencing analysis of bacterial communities. J Microbiol Seoul Korea. 2012 Dec;50(6):1071–4.

115. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, et al. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Methods. 2013 Jan;10(1):57–9.

116. Werner JJ, Koren O, Hugenholtz P, DeSantis TZ, Walters WA, Caporaso JG, et al. Impact of training sets on classification of high-throughput bacterial 16s rRNA gene surveys. ISME J. 2012 Jan;6(1):94–103.

117. Fettweis JM, Serrano MG, Sheth NU, Mayer CM, Glascock AL, Brooks JP, et al. Species-level classification of the vaginal microbiome. BMC Genomics. 2012 Dec 17;13(Suppl 8):S17.

118. Werner JJ, Zhou D, Caporaso JG, Knight R, Angenent LT. Comparison of Illumina paired-end and single-direction sequencing for microbial 16S rRNA gene amplicon surveys. ISME J. 2012 Jul;6(7):1273–6.

119. Edgar RC, Flyvbjerg H. Error filtering, pair assembly, and error correction for next-generation sequencing reads. Bioinformatics. 2015 Jul 2;btv401.

120. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009 Dec;75(23):7537–41.

121. Huggett JF, Laver T, Tamisak S, Nixon G, O'Sullivan DM, Elaswarapu R, et al. Considerations for the development and application of control materials to improve metagenomic microbial community profiling. Accreditation Qual Assur. 2012 Dec;1–7.

122. Salter S, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent contamination can critically impact sequence-based microbiome analyses. bioRxiv. 2014 Jul 16;007187.

123. Walters WA, Caporaso JG, Lauber CL, Berg-Lyons D, Fierer N, Knight R. PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. Bioinformatics. 2011 Apr 15;27(8):1159–61.

124. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci U S A. 2011 Mar 15;108 Suppl 1:4516–22.

125. Liu Z, DeSantis TZ, Andersen GL, Knight R. Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. Nucleic Acids Res. 2008 Oct;36(18):e120.

126. Weiss S, Amir A, Hyde ER, Metcalf JL, Song SJ, Knight R. Tracking down the sources of experimental contamination in microbiome studies. Genome Biol. 2014 Dec 17;15(12):564.

127. Barton HA, Taylor NM, Lubbers BR, Pemberton AC. DNA extraction from low-biomass carbonate rock: An improved method with reduced contamination and the low-biomass contaminant database. J Microbiol Methods. 2006 Jul;66(1):21–31.

128. Laurence M, Hatzis C, Brash DE. Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. PloS One. 2014 May 16;9(5):e97876.

129. Kulakov LA, McAlister MB, Ogden KL, Larkin MJ, O'Hanlon JF. Analysis of Bacteria Contaminating Ultrapure Water in Industrial Systems. Appl Environ Microbiol. 2002 Apr 1;68(4):1548–55.

130. Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Appl Environ Microbiol. 2005 Dec;71(12):8966–9.

131. Patin NV, Kunin V, Lidström U, Ashby MN. Effects of OTU clustering and PCR artifacts on microbial diversity estimates. Microb Ecol. 2013 Apr;65(3):709–19.

132. Zhou H-W, Li D-F, Tam NF-Y, Jiang X-T, Zhang H, Sheng H-F, et al. BIPES, a cost-effective high-throughput method for assessing microbial diversity. ISME J. 2011 Apr;5(4):741–9.

133. Chen T, Yu W-H, Izard J, Baranova OV, Lakshmanan A, Dewhirst FE. The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. Database J Biol Databases Curation [Internet]. 2010 Jul 6 [cited 2015 Sep 14];2010. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2911848/

134. Burke C, Darling AE. Resolving microbial microdiversity with high accuracy full length 16S rRNA Illumina sequencing. bioRxiv. 2014 Nov 6;010967.

135. Tielsch JM, Rahmathullah L, Katz J, Thulasiraj RD, Coles C, Sheeladevi S, et al. Maternal Night Blindness during Pregnancy Is Associated with Low Birthweight, Morbidity, and Poor Growth in South India. J Nutr. 2008 Apr 1;138(4):787–92.

136. Christian P. Micronutrients, Birth Weight, and Survival. Annu Rev Nutr. 2010;30(1):83–104.

137. Ota E, Mori R, Middleton P, Tobe-Gai R, Mahomed K, Miyazaki C, et al. Zinc supplementation for improving pregnancy and infant outcome. Cochrane Database Syst Rev. 2015;2:CD000230.

138. Bhutta ZA, Das JK, Rizvi A, Gaffey MF, Walker N, Horton S, et al. Evidence-based interventions for improvement of maternal and child nutrition: what can be done and at what cost? The Lancet. 2013 Aug 9;382(9890):452–77.

139. Huybregts L, Roberfroid D, Lanou H, Menten J, Meda N, Van Camp J, et al. Prenatal food supplementation fortified with multiple micronutrients increases birth length: a randomized controlled trial in rural Burkina Faso. Am J Clin Nutr. 2009 Dec;90(6):1593–600.

140. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006 Dec 21;444(7122):1027–31.

141. Hullar MAJ, Lampe JW. The gut microbiome and obesity. Nestlé Nutr Inst Workshop Ser. 2012;73:67–79.

142. Kau AL, Ahern PP, Griffin NW, Goodman AL, Gordon JI. Human nutrition, the gut microbiome, and immune system: envisioning the future. Nature. 2011 Jun 15;474(7351):327–36.

143. Subramanian S, Huq S, Yatsunenko T, Haque R, Mahfuz M, Alam MA, et al. Persistent gut microbiota immaturity in malnourished Bangladeshi children. Nature. 2014 Jun 19;510(7505):417–21.

144. Jaeggi T, Kortman GAM, Moretti D, Chassard C, Holding P, Dostal A, et al. Iron fortification adversely affects the gut microbiome, increases pathogen abundance and induces intestinal inflammation in Kenyan infants. Gut. 2014 Aug 20;

145. Bhaskaram P. Immunobiology of mild micronutrient deficiencies. Br J Nutr. 2001 May;85 Suppl 2:S75–80.

146. Neggers YH, Nansel TR, Andrews WW, Schwebke JR, Yu K, Goldenberg RL, et al. Dietary Intake of Selected Nutrients Affects Bacterial Vaginosis in Women. J Nutr. 2007 Sep;137(9):2128–33.

147. Moynihan P, Petersen PE. Diet, nutrition and the prevention of dental diseases. Public Health Nutr. 2004 Feb;7(1A):201–26.

148. Enwonwu CO, Phillips RS, Falkler WA. Nutrition and oral infectious diseases: state of the science. Compend Contin Educ Dent Jamesburg NJ 1995. 2002 May;23(5):431–4, 436, 438 passim; quiz 448.

149. Bisanz JE, Enos MK, PrayGod G, Seney S, Macklaim JM, Chilton S, et al. Microbiota at Multiple Body Sites during Pregnancy in a Rural Tanzanian Population and Effects of Moringa-Supplemented Probiotic Yogurt. Appl Environ Microbiol. 2015 Aug;81(15):4965–75.

150. Winterbourn CC, Chan T, Buss IH, Inder TE, Mogridge N, Darlow BA. Protein carbonyls and lipid peroxidation products as oxidation markers in preterm infant plasma: associations with chronic lung disease and retinopathy and effects of selenium supplementation. Pediatr Res. 2000 Jul;48(1):84–90.

151. Jarosik GP, Land CB, Duhon P, Chandler R, Mercer T. Acquisition of Iron by Gardnerella vaginalis. Infect Immun. 1998 Oct 1;66(10):5041–7.

152. Blencowe H, Cousens S, Oestergaard MZ, Chou D, Moller A-B, Narwal R, et al. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. Lancet. 2012 Jun 9;379(9832):2162–72.

153. van den Broek NR, Jean-Baptiste R, Neilson JP. Factors associated with preterm, early preterm and late preterm birth in Malawi. PloS One. 2014;9(3):e90128.

154. Sebire NJ, Goldin RD, Regan L. Histological chorioamnionitis in relation to clinical presentation at 14-40 weeks of gestation. J Obstet Gynaecol J Inst Obstet Gynaecol. 2001 May;21(3):242–5.

155. Chaim W, Mazor M. Intraamniotic infection with fusobacteria. Arch Gynecol Obstet. 1992 Feb 1;251(1):1–7.

156. Yoon BH, Romero R, Park JS, Chang JW, Kim YA, Kim JC, et al. Microbial invasion of the amniotic cavity with Ureaplasma urealyticum is associated with a robust host response in fetal, amniotic, and maternal compartments. Am J Obstet Gynecol. 1998 Nov;179(5):1254–60.

157. Shipitsyna E, Roos A, Datcu R, Hallén A, Fredlund H, Jensen JS, et al. Composition of the vaginal microbiota in women of reproductive age - sensitive and specific molecular diagnosis of bacterial vaginosis is possible? PloS One. 2013 Apr 9;8(4):e60670.

158. Fettweis JM, Brooks JP, Serrano MG, Sheth NU, Girerd PH, Edwards DJ, et al. Differences in vaginal microbiome in African American women versus women of European ancestry. Microbiol Read Engl. 2014 Jul 29;

159. DiGiulio DB, Callahan BJ, McMurdie PJ, Costello EK, Lyell DJ, Robaczewska A, et al. Temporal and spatial variation of the human microbiota during pregnancy. Proc Natl Acad Sci. 2015 Aug 17;201502875.

160. Harjunmaa U, Järnstedt J, Alho L, Dewey KG, Cheung YB, Deitchler M, et al. Association between maternal dental periapical infections and pregnancy outcomes: results from a cross-sectional study in Malawi. Trop Med Int Health TM IH. 2015 Jul 30;

161. Wang X, Buhimschi CS, Temoin S, Bhandari V, Han YW, Buhimschi IA. Comparative microbial analysis of paired amniotic fluid and cord blood from pregnancies

complicated by preterm birth and early-onset neonatal sepsis. PloS One. 2013 Feb 20;8(2):e56131.

162. Queiros da Mota V, Prodhom G, Yan P, Hohlfheld P, Greub G, Rouleau C. Correlation between placental bacterial culture results and histological chorioamnionitis: a prospective study on 376 placentas. J Clin Pathol. 2013 Mar;66(3):243–8.

163. Luntamo M, Kulmala T, Mbewe B, Cheung YB, Maleta K, Ashorn P. Effect of repeated treatment of pregnant women with sulfadoxine-pyrimethamine and azithromycin on preterm delivery in Malawi: a randomized controlled trial. Am J Trop Med Hyg. 2010 Dec;83(6):1212–20.

164. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. Nature. 2012 Jun 14;486(7402):222–7.

165. Klebanoff MA, Carey JC, Hauth JC, Hillier SL, Nugent RP, Thom EA, et al. Failure of Metronidazole to Prevent Preterm Delivery among Pregnant Women with Asymptomatic Trichomonas vaginalis Infection. N Engl J Med. 2001 Aug 16;345(7):487–93.

166. Kenyon SL, Taylor DJ, Tarnow-Mordi W, ORACLE Collaborative Group. Broad-spectrum antibiotics for spontaneous preterm labour: the ORACLE II randomised trial. ORACLE Collaborative Group. Lancet. 2001 Mar 31;357(9261):989–94.

167. Lamont RF, Nhan-Chang C-L, Sobel JD, Workowski K, Conde-Agudelo A, Romero R. Treatment of abnormal vaginal flora in early pregnancy with clindamycin for the prevention of spontaneous preterm birth: a systematic review and metaanalysis. Am J Obstet Gynecol. 2011 Sep;205(3):177–90.

168. Donders GGG. Definition and classification of abnormal vaginal flora. Best Pract Res Clin Obstet Gynaecol. 2007 Jun;21(3):355–73.

169. Witkin S. The vaginal microbiome, vaginal anti-microbial defence mechanisms and the clinical challenge of reducing infection-related preterm birth. BJOG Int J Obstet Gynaecol. 2014 Oct 15;

170. Donders G, Bellen G, Rezeberga D. Aerobic vaginitis in pregnancy. BJOG Int J Obstet Gynaecol. 2011 Sep;118(10):1163–70.

171. Donders G, Van Calsteren K, Bellen G, Reybrouck R, Van den Bosch T, Riphagen I, et al. Predictive value for preterm birth of abnormal vaginal flora, bacterial vaginosis and aerobic vaginitis during the first trimester of pregnancy. BJOG Int J Obstet Gynaecol. 2009 Sep 1;116(10):1315–24.

172. Gratacós E, Figueras F, Barranco M, Vila J, Cararach V, Alonso PL, et al. Spontaneous recovery of bacterial vaginosis during pregnancy is not associated with an improved perinatal outcome. Acta Obstet Gynecol Scand. 1998 Jan;77(1):37–40.

173. Kwak D-W, Hwang H-S, Kwon J-Y, Park Y-W, Kim Y-H. Co-infection with vaginal Ureaplasma urealyticum and Mycoplasma hominis increases adverse pregnancy outcomes in patients with preterm labor or preterm premature rupture of membranes. J Matern-Fetal Neonatal Med Off J Eur Assoc Perinat Med Fed Asia Ocean Perinat Soc Int Soc Perinat Obstet. 2013 Jun 24;

174. Oliver RS, Lamont RF. Infection and antibiotics in the aetiology, prediction and prevention of preterm birth. J Obstet Gynaecol J Inst Obstet Gynaecol. 2013 Nov;33(8):768–75.

175. Nugent RP, Krohn MA, Hillier SL. Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation. J Clin Microbiol. 1991 Feb;29(2):297–301.

176. Fredricks DN, Fiedler TL, Marrazzo JM. Molecular identification of bacteria associated with bacterial vaginosis. N Engl J Med. 2005 Nov 3;353(18):1899–911.

177. Chico RM, Mayaud P, Ariti C, Mabey D, Ronsmans C, Chandramohan D. Prevalence of malaria and sexually transmitted and reproductive tract infections in pregnancy in sub-Saharan Africa: a systematic review. JAMA J Am Med Assoc. 2012 May 16;307(19):2079–86.

178. Hummelen R, Fernandes AD, Macklaim JM, Dickson RJ, Changalucha J, Gloor GB, et al. Deep sequencing of the vaginal microbiota of women with HIV. PloS One. 2010;5(8):e12078.

179. Schellenberg JJ, Links MG, Hill JE, Dumonceaux TJ, Kimani J, Jaoko W, et al. Molecular definition of vaginal microbiota in East African commercial sex workers. Appl Environ Microbiol. 2011 Jun;77(12):4066–74.

180. Romero R, Hassan SS, Gajer P, Tarca AL, Fadrosh DW, Bieda J, et al. The vaginal microbiota of pregnant women who subsequently have spontaneous preterm labor and delivery and those with a normal delivery at term. Microbiome. 2014;2:18.

181. Taha TE, Kumwenda NI, Kafulafula G, Makanani B, Nkhoma C, Chen S, et al. Intermittent Intravaginal Antibiotic Treatment of Bacterial Vaginosis in HIV-Uninfected and -Infected Women: A Randomized Clinical Trial. PLoS Clin Trials [Internet]. 2007 Feb 23 [cited 2015 Aug 24];2(2). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1851729/

182. Misra DP, Trabert B. Vaginal douching and risk of preterm birth among African American women. Am J Obstet Gynecol. 2007 Feb;196(2):140.e1–8.

183. Brotman RM, Klebanoff MA, Nansel TR, Andrews WW, Schwebke JR, Zhang J, et al. A longitudinal study of vaginal douching and bacterial vaginosis--a marginal structural modeling analysis. Am J Epidemiol. 2008 Jul 15;168(2):188–96.

184. Gajer P, Brotman RM, Bai G, Sakamoto J, Schütte UME, Zhong X, et al. Temporal dynamics of the human vaginal microbiota. Sci Transl Med. 2012 May 2;4(132):132ra52.

185. MacIntyre DA, Chandiramani M, Lee YS, Kindinger L, Smith A, Angelopoulos N, et al. The vaginal microbiome during pregnancy and the postpartum period in a European population. Sci Rep. 2015;5:8988.

186. Wen A, Srinivasan U, Goldberg D, Owen J, Marrs CF, Misra D, et al. Selected vaginal bacteria and risk of preterm birth: an ecological perspective. J Infect Dis. 2013 Nov 22;

187. Balkus JE, Richardson BA, Rabe LK, Taha TE, Mgodi N, Kasaro MP, et al. Bacterial vaginosis and the risk of trichomonas vaginalis acquisition among HIV-1-negative women. Sex Transm Dis. 2014 Feb;41(2):123–8.

188. Pybus V, Onderdonk AB. A commensal symbiosis between Prevotella bivia and Peptostreptococcus anaerobius involves amino acids: potential significance to the pathogenesis of bacterial vaginosis. FEMS Immunol Med Microbiol. 1998 Dec 1;22(4):317–27.

189. Gray RH, Wabwire-Mangen F, Kigozi G, Sewankambo NK, Serwadda D, Moulton LH, et al. Randomized trial of presumptive sexually transmitted disease therapy during pregnancy in Rakai, Uganda. Am J Obstet Gynecol. 2001 Nov;185(5):1209–17.

190. Gichangi PB, Ndinya-Achola JO, Ombete J, Nagelkerke NJ, Temmerman M. Antimicrobial prophylaxis in pregnancy: A randomized, placebo-controlled trial with cefetamet-pivoxil in pregnant women with a poor obstetric history. Am J Obstet Gynecol. 1997 Sep;177(3):680–4.

191. Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M. Swarm: robust and fast clustering method for amplicon-based studies. PeerJ [Internet]. 2014 Sep 25 [cited 2015 Feb 13];2. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4178461/

192. Adams Waldorf KM, Gravett MG, McAdams RM, Paolella LJ, Gough GM, Carl DJ, et al. Choriodecidual Group B Streptococcal Inoculation Induces Fetal Lung Injury without Intra-Amniotic Infection and Preterm Labor in Macaca nemestrina. PLoS ONE. 2011 Dec 21;6(12):e28972.

193. Randis TM, Gelber SE, Hooven TA, Abellar RG, Akabas LH, Lewis EL, et al. Group B Streptococcus β-hemolysin/cytolysin breaches maternal-fetal barriers to cause preterm birth and intrauterine fetal demise in vivo. J Infect Dis. 2014 Jan 28;

194. Han YW. Fusobacterium nucleatum: a commensal-turned pathogen. Curr Opin Microbiol. 2015 Jan 7;23C:141–7.

16 Appendix

16.1 Qiime script settings

#Pick OTUs

parallel_pick_otus_uclust_ref.py -i /data/seqs.cat.fna -o /data/pick_otus/ -r /data/97_named_trimmed_reference.fasta -O 159 –T

filter_fasta.py -f /data/seqs.cat.fna -o /data/pick_otus/seqs.cat_failures.fasta -s seqs.cat_failures.txt

parallel_pick_otus_uclust_ref.py -i /data/pick_otus/seqs.cat_failures.fasta -o /data/pick_otus/failures/ -r /home/ubuntu/qiime_software/gg_otus-12_10-release/rep_set/97_otus.fasta -O 159 -T

#pick representative sequences for all split otu maps

nohup pick_rep_set.py -i seqs.cat.combined.txt -f /data/seqs.cat.fna -o /data/rep_set/rep_set.fasta &

nohup parallel_assign_taxonomy_uclust.py -i rep_set.fasta -o /data/uclust_assigned_taxonomy/ -O 8 &

cat seqs.cat_otus.txt seqs.cat_otus_failures.txt > seqs.cat_otus_merged.txt

nohup make_otu_table.py -i otu_maps/seqs.cat_otus_merged.txt -o otu_table.biom -t uclust_assigned_taxonomy/rep_set _tax_assignments.txt &

nohup parallel_align_seqs_pynast.py -i rep_set/rep_set_.fasta -o pynast_aligned_seqs/ -T -O 8 &

filter_alignment.py -o pynast_aligned_seqs/ -i pynast_aligned_seqs/rep_set _aligned.fasta

nohup make_phylogeny.py -i pynast_aligned_seqs/rep_set _aligned_pfiltered.fasta -o rep_set.tre &

*** REMOVING NEGATIVE EXTRACTION AND NEGATIVE RUN CONTROL CONTAMINANTS ***

split_otu_table.py -i otu_table.biom -m mapping_combined.txt -f BodySite -o split_by_site/

biom convert -i otu_table_Negative_extraction.biom -o otu_table_Negative_extraction.txt -b --header-key taxonomy

filter_otus_from_otu_table.py -i otu_table.biom -o otu_table_c.biom -e contamination_otus.txt

filter_otus_from_otu_table.py -i otu_table_c.biom -o otu_table_c_mc2.biom -n 2

*** ANALYSING MOCK COMMUNITY DATA ***

biom convert -i otu_table_Mock_community.biom -o otu_table_Mock_community.txt -b --header-key taxonomy

*** FILTERING OTU TABLE ***

python filter_observations_by_sample.py -i combined/output/otu_table_c_mc2.biom -o combined/output/otu_table_c_f.biom -n 0.008 -f

filter_samples_from_otu_table.py -i otu_table_c_f.biom -o otu_table_c_f_lf.biom -m mapping_combined.txt --output_mapping_fp mapping_combined_filtered.txt -n 1000

filter_otus_from_otu_table.py -i otu_table_c_f_lf.biom -o otu_table_c_f_lf_mc2.biom -n 2

biom summarize-table -i otu_table_c_f_lf_mc2.biom -o otu_table_c_f_lf_mc2_summary.txt

single_rarefaction -i otu_table_f_c_mc2_lf_s_mc2.biom -o otu_table_f_c_mc2_lf_s_mc2_1000.biom -d 1000

split_otu_table.py -i otu_table_c_f_lf_mc2.biom -m mapping_combined.txt -f BodySite -o split_by_site/

*** ALPHA DIVERSITY ***

alpha_rarefaction.py -i oral/otu_table_c_f_lf_mc2_Dental_plaque_Malawi.biom -m oral/mapping_Dental_plaque_Malawi.txt -o oral/alpha/ -t /home/ronan/miseq/combined/output/rep_set.tre -a -O 7 -n 1 --min_rare_depth 1000 -e 1000 -p /home/ronan/miseq/combined/alpha_parameter.txt

alpha_rarefaction.py -i vagina/otu_table_c_f_lf_mc2_Vaginal_mucus_Malawi.biom -m vagina/mapping_Vaginal_mucus_Malawi.txt -o vagina/alpha/ -t /home/ronan/miseq/combined/output/rep_set.tre -a -O 7 -n 1 --min_rare_depth 1000 -e 1000 -p /home/ronan/miseq/combined/alpha_parameter.txt

alpha_rarefaction.py -i placenta/otu_table_c_f_lf_mc2_Placenta_Malawi.biom -m placenta/mapping_Placenta_Malawi.txt -o placenta/alpha/ -t /home/ronan/miseq/combined/output/rep_set.tre -a -O 7 -n 1 --min_rare_depth 1000 -e 1000 -p /home/ronan/miseq/combined/alpha_parameter.txt

alpha_rarefaction.py -i membrane/otu_table_c_f_lf_mc2_Fetal_membrane_Malawi.biom -m membrane/mapping_Fetal_membrane_Malawi.txt -o membrane/alpha/ -t /home/ronan/miseq/combined/output/rep_set.tre -a -O 7 -n 1 --min_rare_depth 1000 -e 1000 -p /home/ronan/miseq/combined/alpha_parameter.txt

*** BETA DIVERSITY ***

beta_diversity_through_plots.py -i oral/otu_table_c_f_lf_mc2_Dental_plaque_Malawi.biom -m oral/mapping_Dental_plaque_Malawi.txt -o oral/beta/ -t /home/ronan/miseq/combined/output/rep_set.tre -a -O 7 -p /home/ronan/miseq/combined/beta_parameter.txt -e 1000

beta_diversity_through_plots.py -i vagina/otu_table_c_f_lf_mc2_Vaginal_mucus_Malawi.biom -m vagina/mapping_Vaginal_mucus_Malawi.txt -o vagina/beta/ -t /home/ronan/miseq/combined/output/rep_set.tre -a -O 7 -p /home/ronan/miseq/combined/beta_parameter.txt -e 1000

beta_diversity_through_plots.py -i placenta/otu_table_c_f_lf_mc2_Placenta_Malawi.biom -m placenta/mapping_Placenta_Malawi.txt -o placenta/beta/ -t /home/ronan/miseq/combined/output/rep_set.tre -a -O 7 -p /home/ronan/miseq/combined/beta_parameter.txt -e 1000

beta_diversity_through_plots.py -i membrane/otu_table_c_f_lf_mc2_Fetal_membrane_Malawi.biom -m membrane/mapping_Fetal_membrane_Malawi.txt -o membrane/beta/ -t /home/ronan/miseq/combined/output/rep_set.tre -a -O 7 -p /home/ronan/miseq/combined/beta_parameter.txt -e 1000

split_otu_table.py -i otu_table_f_c_mc2_lf_s_mc2.biom -m mapping/mapping_combined_filtered.txt -f BodySite -o split_by_site_filtered/

filter_otus_from_otu_table.py -i otu_table_f_c_mc2_lf_s_mc2_Dental_plaque_Malawi.biom -o otu_table_f_c_mc2_lf_s_mc2_Dental_plaque_Malawi_mc1.biom -n 1

filter_otus_from_otu_table.py -i otu_table_f_c_mc2_lf_s_mc2_Fetal_membrane_Malawi.biom -o otu_table_f_c_mc2_lf_s_mc2_Fetal_membrane_Malawi_mc1.biom -n 1

filter_otus_from_otu_table.py -i otu_table_f_c_mc2_lf_s_mc2_Placenta_Malawi.biom -o otu_table_f_c_mc2_lf_s_mc2_Placenta_Malawi_mc1.biom -n 1

filter_otus_from_otu_table.py -i otu_table_f_c_mc2_lf_s_mc2_Vaginal_mucus_Malawi.biom -o otu_table_f_c_mc2_lf_s_mc2_Vaginal_mucus_Malawi_mc1.biom -n 1