School of Molecular and Life Sciences
Faculty of Science and Engineering

Characterisation of the microbiome at different life stages using optimised sample collection and data processing methods to assist with long-term health

Jacquelyn Margaret Jones

This thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University

August 2022

– Declaration –

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university

Signed

Jacquelyn M Jones

Date: 15/08/2022

– Acknowledgments –

The journey from undergraduate through to the end of a PhD was a most unexpected, yet deeply rewarding turn of events. I have experienced the excitement of publishing one's own work, the pride in using my skills to navigate extraordinarily large study from processing through to discussion, and the frustration of failed experiments. I would not have had the means to complete this thesis without the support of my supervisors, colleagues, friends, and family.

First and foremost, I would like to express my gratitude to supervisory panel AssProff. Claus Christophersen, thank you for your constant guidance, encouragement, and support and thank you Dr. Debra Palmer, Dr Stacey Reinke, Dr. Seyedehmahsa Mousaviderazmahalleh, Proff. Michael Bunce, and Proff. Morten Allentoft for your invaluable input and assistance. I was incredibly lucky to have such a wealth of knowledge and diversity of support across my supervisory panel.

I would like to extend my deepest gratitude to my dear friends and colleagues at the Trace and Environmental DNA laboratory (TrEnD) "lab rats". Georgia Nester and Mathew Heydenrych, thank you for your significant contribution to sample processing and DNA extraction. The early mornings and late nights spent surrounded by little pots of poo is a testament to the strong friendships and deep love of science I found in my Trend comrades. Mathew Campbell, Mathew Heydenrych, and Mathew Power, thank you all for your kind assistance with everything QIAGEN, ThermoFisher, and MiSeq. Your detailed lessons and troubleshooting expertise were invaluable. A special thank you to Nicole White for your patient assistance during library prep and sequencing, as well as your kind support during difficult times.

Kristen Fernandes, thank you for being my crazy rat friend, empty glove box smuggler, and colleague; and Mieke van der Heyde thank you for being my friend and housemate eh. Thank you both Jason Alexander for being my early morning coffee comrade, and Arne Adam for being the late-night office DJ, you both made the daunting task of writing something to look forward to; and David Juszkiewicz, thank you for being my confidant and thoughtful friend in times of need. Frederik Seersholm, Katrina West, Sophie Preston, and Miwa Takahashi, thank you each for welcoming me into your lives, and sharing in the ups and downs of student life.

To the members of The Western Australian Human Microbiome Collaboration Centre who I have not already mentioned Rose lines, Alicia Tan, and especially Alishum Ali, thank you for showing me the ropes. Ali your continuous help with the daunting task of bioinformatics was an absolute lifeline, and your input and expertise as a coauthor and lab mate was highly valued.

Throughout my studies, pole dancing, and the pole community has kept me grounded, accountable, and provided me with the resilience needed to continue through the more

challenging periods of PhD life. Tammy Baxter, Jacinta Diamonds, Sam Stock, and Amy Breur, thank you for your friendship, support, and encouragement throughout the last four years.

To my Australian family, Gale, Richard, and Ewan George, thank you for taking me in, and helping me find my way in this new country, and to Kelly Shann, thank you for stopping me from blowing forever in the wind, and pushing me towards university. Marty MacCallum, thank you for all you love, patients, kindness, and encouragement, and for allowing me to extend our family with rats Rickey and Boyle, and dog Dusty.

Above all, I would like to thank both my parents Collen and Don Jones. You both gave me every opportunity, even if it meant driving me to the pool at 5am and allowed me the freedom to find my own way in life. Your love and trust provided me with the resilience and persistence needed for this endeavour. Though I do not know whether my father understood I was working on this task before he died, I know he would have been incredibly proud. Thank you to my brother Scott Jones for your love and encouragement, and for keeping me updated on life in Canada.

– Abstract –

The gastrointestinal microbiome is a complex community of trillions of microbes from all three domains of life. Over one thousand different species reside within a healthy gut and provide a metagenome that greatly expands the metabolic activity of the gut well beyond that encoded by the human genome. Like any other body organ, the gut microbiome plays a vital role in our overall health, including maintenance of the local gut wall and providing essential and otherwise unattainable nutrients from our diet including short chain fatty acids (SCFA). The microbiome also has far reached involvement in the development of both the immune system and the brain, meaning it makes a profound contribution to how the human body responds to the environment. The impact of this relationship is however not one directional, and our lifestyle will also affect how the microbiome is structured, and how it functions. In fact, the microbiome may be the most adaptable "organ" we have, which allows for much greater environmental influence, and potential therapeutic application. Understanding how a healthy microbiome might be structured, how it is established in early life and how it changes over time, and how it changes during disease have become important research questions that we have begun to unravel with the development of modern molecular methods.

The research projects detailed in this thesis are focused on the development of improved protocols for faecal sample collection and replicable microbiome data, that are used to reveal the microbial ecosystem during both modern disease; and during interventions designed to modulate the microbiome and elevate disease. I begin by reviewing the literature on this dynamic community and describing its structure from colonization, through adult life, and during pregnancy. I also cover the microbiomes involvement in the development of the immune system and the bidirectional communication pathway of the gut microbiome axis. Examples of methods used to modify the microbiome, and a brief summary of the molecular based methods used to explore the microbiome are also discussed. In chapter two, I compare faecal sample collection, processing, and storage methods for bias that may misrepresent the microbiome and SCFA profiles. Then in chapter three I use these newly developed methods to identify microbial and metabolic signatures that differ between young children with differing levels of neurotypical development. In chapter four, I investigate the effectiveness of a maternal prebiotic supplement on the microbiome composition of mother infant pairs over time, and again highlight the need for appropriate sample processing for longitudinal microbiome analysis.

In chapter 2, I investigate some of the challenges faced when collecting material for microbiome investigation by comparing sample collection tubes, comparing spot sampling to total faecal sample homogenization, and comparing stool samples that are consecutively collected at different times during the day. I also describe how positive and negative controls can be used to better filter and interpret microbiome profiles. For this study, 7 women provided 3 stool samples each, from which bacterial (16S rRNA gene (V4)) and fungal (ITS2) communities, as well as

short chain fatty acid (SCFA) concentrations were investigated. Here I show how spot sampling of stool results in variable detection of some microbial members, and inconsistent levels of SCFA. I also show an inverse relationship between microbial alpha-diversity and metabolite concentrations from stool collected at different consecutive timepoints. The total SCFA concentration (t-test, $p = 0.04$) and acetic acid concentration (t-test, $p = 0.03$) were significantly higher in the second stool sample compared to the first. Commercial stool collection kits are also compared with immediate freezing, and significant differences (PERMANOVA $p < 0.01$) in bacterial beta-diversity resulting from collecting stool samples in these commercial tubes. Therefore, I advocate for the use of raw stool sample freezing, rather than the use of commercial kits, and recommend that stool be collected in bulk prior to processing and subsampling for multiple analysis. I also suggest that participants are provided with guidance on the preferred time of day when samples should be collected.

Chapter 3 details the faecal microbiome and SCFA profiles of very young children who were showing early behavioural signs of autism spectrum disorder (ASD), examined using the methods developed in chapter 2. A total of 30 stool samples were collected from 24 children between 21 and 40 months who were involved in a pre-emptive behavioural intervention RCT and were assessed for neurological and behavioural development using multiple diagnostic tools. The fungal richness and acetic acid concentrations were observed to be higher with increasing autism severity, and the abundance of several bacterial taxa also changed due to the severity of ASD. The microbiome composition was found to change with increasing fungal richness and acetic acid concentrations with increasing autism severity based on neurological and developmental score, indicating changes associated with autistic behaviour. I also explored the microbiome for any association with the pre-emptive treatment, and found evidence of significant community shifts from the genus, to the phyla level (PERMANOVA $p \leq 0.014$). SCFA concentrations were also found to be closely associated with stool form. Together, these findings indicate that subtle changes in bacterial composition may occur in the microbiome of young children with an early diagnosis for autism, and may provide justification for conducting larger pre-emptive or even retrospective studies that assess early changes to the microbiome and stool form of children who go on to develop autism.

In chapter 4, I assess of the temporal changes to the infant and maternal microbiome in response to a maternal prebiotic supplement. The mother infant pairs (74 participants with complete data at all timepoints for 65 participants) were enrolled in a double-blind placebo-controlled trial of the prebiotic supplement. Microbiome and SCFA profiles resulted from mothers at 6 timepoints during (20-, 28-, and 36-weeks) and after pregnancy (2-, 4-, and 6-months), and 4 timepoints (2-, 4-, 6-, and 12-months) during the first year of their infant's life. This large dataset was integrated using replicate sampling and negative controls by the methods which were developed in chapter 2 and improved upon in chapter 3. We found a significant shift in the composition of the microbiome of mothers (PERMANOVA $p < 0.0001$, psudo-F 4.23), and infants (PERMANOVA $p < 0.001$; psudo-F 1.7) between the intervention and placebo groups. Maternal

acetic acid concentrations were also found to increase significantly in the previotic group ($p = 0.008$; Pseudo-F = 7.09) and infant SCFA concentrations differed significantly by time ($p > 0.004$), but not by intervention. I also document the period of diversification and enrichment of the microbiome over the first year of life, that was particularly rapid after the introduction of solid foods (ANOVA, $p>0.001$). Again, I was able to show the value of collecting stool form as all maternal diversity measures dropped significantly in the firm stool group compared to the loose stool group (rm-ANOVA $p<0.03$), while acetate, butyrate, and propionate all significantly increased (rm-ANOVA $p<0.001$).

Overall, the research in this thesis generated microbiome and SCFA profiles for 80 adult women and 104 infants and young children from 820 stool samples. Across 9 sequencing runs over 66 million high quality filtered reads were generated, resulting in 3 studies which contribute to the growing demand for more reproducible microbiome investigation. The work undertaken here addresses the need for better sampling, decontamination, and reporting protocols, and employs this framework in frontier areas of microbiome research. The ability to detect subtle changes in the gut prior to a particular health outcome, and to demonstrate the protective effect of an intervention is among the most important aspects of microbiome research as it can lead to the development of effective treatments for modern non-communicable diseases.

– Table of Contents –

– Thesis Overview –

This thesis contains 5 chapters, and is presented as a hybrid thesis, containing both published manuscripts and a thesis data chapter. The first chapter critically summarises the relevant literature surrounding molecular investigation of the human gut microbiome. Chapter 2 - 4 each present research that were developed as part of this PhD project. The manuscript presented in chapter 2 is publicly available after being published in the peer reviewed journal Scientific Reports. This published work has already made a significant contribution to the microbiome research community, as it received a place among the top 100 microbiology Scientific Reports papers in 2021. The manuscript presented in chapter 3 is also now publicly available after being published in the peer reviewed journal in Frontiers in Microbiology. The layout of both manuscripts has been modified to align with the formatting in this thesis. Chapter 4 contains the largest volume of data in this thesis, and has been formatted accordingly as a thesis data chapter. Chapter 5 synthesises the main findings and significance of this thesis and highlights several questions uncovered that point to future research.

– List of Figures –

– LIST OF TABLES –

– Abbreviations –

16S V4	– Fourth hypervariable region of the 16S rRNA gene
ADOS2	– Autism Diagnostic Observation Schedule, second edition
ACE	– Abundance-based coverage estimator
ANCOM-BC	– Analysis of compositions of microbiomes with bias correction
ANOVA	– Analysis of variance
ASD	– Autism spectrum disorder
ASV	– Amplicon sequence variant
BBB	– Blood brain barrier
BH	– Benjamini-Hochburg adjustment
BLAST	– Basic Local Alignment Search Tool
BSA	– Bovine Serum Albumin
BSFS	– Bristol stool form scale
bp	– Base pairs
CFUs	– Colony forming units
CLR	– Centre log ratio
CNS	– Central nervous system
CSS	– Calibrated severity score
CV	– Coefficient of variation
dbRDA	– Distance-based redundancy analysis
DESeq2	– Differential gene expression analysis for sequence count data
DistLM	– Distance-based liner models
DNA	– Deoxyribonucleic acid
dNTP	– Deoxynucleotide triphosphate
ENS	– Enteric nervous system
FDR	– False discovery rate
FOS	– Fructo-oligosaccharide
GC-MS	– Gas chromatography mass spectrometry
GI	– Gastrointestinal
GPR	– G protein-coupled receivers
GOS	– Galacto-oligosaccharide
GT	– Genome Taxonomy
HMO	– Human milk oligosaccharides
iBASIS-VIPP	– iBASIS–Video Interaction to Promote Positive Parenting
IBD	– Irritable bowel disease
Ig	– immunoglobulin A
IL	– Interleukin
IRT	– inhibiter removal technology
ITS	– Internal transcribed spacer

KEGG	– Kyoto Encyclopedia of Genes and Genomes
KO	– KEGG ortholog
LPS	– Lipo-polysaccharide
MSEL	– Mullens Scale of Early Learning
NAASD	– Non-autism autism spectrum disorder
NDC	– No developmental concern
NGS	– Next generation sequencing
OTU	– Operational taxonomic unit
PD	– Phylogenetic diversity
PERMANOVA	– Permutational multivariate analysis of variance
PERMDISP	– Permutational multivariate analysis of dispersion
PCoA	– Principal coordinates analysis
PCR	– Polymerase chain reaction
PICRUSt	– Phylogenetic investigation of communities by reconstruction of unobserved states
qPCR	– Quantitative PCR
RCT	– Randomised control trial
RDP	– Ribosomal Database Project
rCCA	– Regularised canonical correlation analysis
rRNA	– Ribosomal ribonucleic acid
SACS-R	– Social Attention and Communication Survallance-Revised
SCFA	– Short chain fatty acids
SIMPROF	– Similarity profile test
TGS	– Third generation sequencing
Th(1 or 2)	– Type (1 or 2) helper
UCC	– Usual community care
UNITE	– Unified system for the identification of Eukaryotes

– Chapter 1 –
General Introduction and Literature Review

1.1 The Gastrointestinal Microbiome is a Complex Community

1.1.1 The microbiome structure and symbiosis in the gut

Through coevolution, humans have established a mutualistic relationship with a diverse assortment of microorganisms, the majority of which inhabit the gastrointestinal tract (Pillai, Gouhier and Vollmer, 2014). This complex ecosystem forms one of the densest communities known to science, with some estimates predicting some 100 trillion microbes from all three domains of life coexist in the human gut (Amon and Sanderson, 2017). Mutualistic interactions such as production of public goods through extracellular metabolism (e.g. Bacteroidales use glycoside hydrolases to breakdown polysaccharides) and quorum sensing are exceedingly common between gut microbes, and results in an interdependent and highly cooperative community (Koskella, Hall and Metcalf, 2017). These numerous and diverse microbial inhabitants provide a metagenome predicted to encode 150 times more than the human genome, which greatly expands the metabolic capabilities of the meta-organism (Amon and Sanderson, 2017; Yadav et al., 2018). Bacterial members from 50 different phyla have been found within the gut (Yadav et al., 2018). The most abundant members are Bacteroides and Firmicutes (Eckburg et al., 2005), and some less abundant phyla include Actinobacteria, Fusobacteria, Proteobacteria, and Verrucomicrobia. Fungi from the three phyla Ascomycota, Basidiomycota, and Zygomycota are found in the human gut (Sam, Chang and Chai, 2017). While predominant fungi are recovered less consistently than bacteria, the most commonly recovered fungal members include *Candida*, *Saccharomyces*, and *Penicillium* (Hallen-Adams and Suhr, 2017). This symbiotic community of microorganisms is often called the gastrointestinal micro-flora, or simply gut flora, while the collection of the total genetic output is referred to as the microbiome. The term bacteriome includes only those bacterial members, the most abundant and stable residents, while the mycobiome describes only those fungal members. There are also other less abundant micro-eukaryotes (Hallen-Adams *et al.*, 2015), as well as relatively few Archaea (Nkamga, Henrissat and Drancourt, 2017). Little is known about the non-living phages and viruses which also influence the structure of microbial communities (Columpsi *et al.*, 2016), and microbiome research has primarily focused on bacterial communities, and more recently, interactions between bacteria and fungi (Romani *et al.*, 2015).

The entire intestine averages 7.9 m in length, and is specialized into different regions, including in the large intestine the caecum, and the ascending, transverse, and descending colon. Due to the oxygen, nutrient, and antimicrobial gradients that exist along the digestive tract, as well as host immunity distinct environments are contained within each of these areas, and different microbial traits are required for microbes to successfully colonise (Donaldson, Lee and Mazmanian, 2015). In the small intestine, the oxygen and antimicrobial levels are higher, and the pH shifts from about 6 to 7.4 at the terminal ileum. Here, a low diversity of bile acid tolerant facultative anaerobes including Clostridium predominate. On the other hand, with lower levels of simple sugars, oxygen, and antimicrobials in the large intestine, a more diverse assortment of microbial members are found here, mostly polysaccharide-degrading strict anaerobes including Bacteroidaceae and Ruminococcaceae (Donaldson, Lee and Mazmanian, 2015). Fungi including *Candida* and *Malassezia* are also commonly identified in stool and are thought to be residents of the large intestine as they do not persist outside of the mammal microbiome (Hallen-Adams and Suhr, 2017). Bacteria and fungi can interact antagonistically through the secretion of the penicillin-based antimicrobials by fungi (Arvanitis and Mylonakis, 2015) and the stimulation of colonocytes to produce antimicrobial peptides by bacteria (Fan *et al.*, 2015); and commensally when they form a hypoxic microenvironment through mixed species biofilms (Fox *et al.*, 2014). The further depletion of oxygen and enrichment of immune antibodies also increases as you move deeper into the mucosa from the gut lumen, and only highly specialised microbial members can colonize these areas (Donaldson, Lee and Mazmanian, 2015). The complex activities of the gut (e.g. immunological defence, motility and digestive secretions) are coordinated by the enteric nervous system, which unlike other branches of the peripheral nervous system, can act largely independently of the brain through an integrated neuronal network (Rao and Gershon, 2016). The microbiota therefore interface with a highly sophisticated system that is evolved to receive and manage signals directed from the gut.

As part of the interface of the body, the gut mucosa is always exposed to the external environment, which means it is both influenced by and vulnerable to the surrounding environment and to pathogens (Li *et al.*, 2018). The mucins that are secreted by goblet cells make up the mucus layer, which contain glycoproteins and provide attachment sites and nutrients for particular microorganisms (Cai *et al.*, 2020). Microorganisms with complementary adhesions will have good mucus-binding capacity and therefore, prolonged retention in the gut. These include bacteria such as *Akkermansia muciniphila* and *Bacteroides thetaiotaomicron* (Schroeder, 2019), and while there is less evidence surrounding, mucosa associated fungi, *C. albicans* and *Rhodotorula mucilaginosa* are among those identified in mucosal samples (Zhang *et al.*, 2021). The microbiota present at the mucosa can also promote the expression of genes which affect both the strength and glycan structure of the mucus layer (Schroeder, 2019). Beneath the mucosa, a thin

layer of epithelial cells forms a semipermeable barrier which must allow for both selective absorption and exclusion of compounds in the gut. Consequently, there are a number of ways that the host and the microbial members interact to maintain homeostasis. One way in which host cells recognise bacteria is through physical recognition. Host epithelial cells have pattern recognition receptors that are used to recognize members of the microbiome through conserved regions of the cell (Li *et al.*, 2018). The epithelial cells can also be induced by local microbial metabolites to produce antimicrobial peptides, which can protect the host from microbial pathogens (Venegas *et al.*, 2019).

The innate immune system also takes part in maintaining microbiome homeostasis, aided by the largest collection of immune cells embedded within the epithelial layer of the gut lumen (Rao and Gershon, 2016). The immune effector cells receive signalling molecules produced by bacteria within the gut mucosa, and once these signals are transduced, a response can be elicited that can alter the composition of the microbiome (Li *et al.*, 2018). Interactions between mucosal immune cells and fungi have also been observed in gnotobiotic mice colonised with particular microbes, showing cytokine-producing cells increase with increasing fungal diversity (van Tilburg Bernardes *et al.*, 2020). Microbes that are permitted to inhabit the area in closest proximity to the host play a key role in maintaining the health of the gut by promoting both cell proliferation, and repair of epithelial cells, and by improving gut barrier integrity (Venegas *et al.*, 2019). Likewise, the microbial metabolite butyric acid is the primary energy source used by colonic epithelial cells. In turn, butyrate oxidation by epithelial cells consumes oxygen, which creates a favourable environment for the anaerobes that produce this acid (Ramos, Hernández and Blanca, 2009; Verbeke *et al.*, 2015; Venegas *et al.*, 2019). This intrinsic, tightly regulated relationship between microbiota and host allows for only select microbes to thrive in a healthy gut over the long term.

1.1.2 Human behaviour, diet, and microbial cross-feeding

Of all the external factors that shape the gut microbiome, diet imposes the strongest force on community composition (Yatsunenko *et al.*, 2012; Carmody *et al.*, 2015; Xu and Knight, 2015; Yadav *et al.*, 2018), with a dynamic quantity of microbial metabolites also fluctuating in response to environmental pressures including nutrient availability (Romani *et al.*, 2015). The community composition of each host is unique, and the response to short term dietary changes, including fluctuations in microbial species and the overall metabolic output is highly personal (David *et al.*, 2014; Johnson *et al.*, 2019). Together, this seems to indicate that community composition is largely determined by diet, and when dietary shifts occur, the historical composition of the microbiome restricts and directs the community response both to new nutrients and to new microbial members, and this is highly personal because dietary history is highly personal.

Furthermore, the microbiome of young children is less resistant to change, which may be partly because it has had less time to accrue a historical composition. This idea also in agreement within the fact that the composition of the microbiome is more similar among those sharing the same long-term dietary profile. For example, the western diet provides approximately 12 g of undigested protein to the colon each day, where it is fermented by microbial residents releasing among other compounds, ammonia, phenols and sulphides (Beaumont *et al.*, 2017). If a high-protein diet is consumed, excess protein digestion in the gut can shift the metabolic community and increase the concentration of potentially harmful fermentation products (Beaumont *et al.*, 2017). Alternatively, vegetarian diets with increased carbohydrate consumption have been shown to increase the abundance of both fungi (Hallen-Adams and Suhr, 2017) and diversity of bacteria (Tomova *et al.*, 2019).

Over the last 100 years, there have been dramatic changes to the human diet due to agriculture practices, and overconsumption has become a defining factor in gut microbiome composition. Modern diets range across socio-economic backgrounds, and geographical locations, with meat making up a larger proportion of the diet in industrialised nations and greater fibre consumption in less developed rural areas (Moles and Otaegui, 2020). Plant degrading *Prevotella* species are commonly found in the gut microbiome of individuals living in less developed areas, while *Bacteroides*, *Faecalibatcerium* and Ruminococcus are more common in the gut microbiome of healthy people living in industrialised areas (Cronin *et al.*, 2021). These changes may indicate diet has caused a permanent shift in the microbiota of the human gut; in fact, a longitudinal study found that feeding mice a low fibre diet over 5 generations caused a compounded loss in diversity that could no longer be restored to control levels using diet (Sonnenburg *et al.*, 2016). Across all human diets however, most calories come from carbohydrates, with processed and preserved carbohydrates lacking fibre predominating more western diets (Moles and Otaegui, 2020).

Polysaccharides are the most abundant carbohydrate found in plant and animal food sources, and their structural diversity is exceptionally large. To add to the complexity, a number of different glyosidic linkages are used to build these natural polymers (Koropatkin, Cameron and Martens, 2012). Starch, lactose and glucose are three of the few polysaccharides that the human genome is able to degrade, but the diversity of microorganisms in the gut is able to greatly expand the number of polysaccharides, including fibres, that can be utilised by the host (Yadav *et al.*, 2018). Fibres are plant-based carbohydrates that resist digestion in the small intestine, and are the main source of carbon used by microbial residents as substrates for fermentation; primarily these are resistant oligosaccharides, non-starch polysaccharides, and resistant starch (Brinkworth *et al.*, 2009). Different fibre types have different solubility in aqueous solutions (which can increase faecal bulking), and different chemical properties which determine what

enzymatic toolkit is required to break it down. Thus, the presence or absence of microbial genes will ultimately determine which microbes are able to ferment each fibre (Cronin *et al.*, 2021). Gut microbiota produce over a thousand different carbohydrate-active enzymes from four main groups: carbohydrate esterases, glycoside hydrolases, glycosyltransferases, and polysaccharide lyases (Cai *et al.*, 2020). Bifidobacteria in particular have been shown to prefer short-chain oligosaccharides, and posse an enzyme that enables them to break down beta-(2-1)-fructans (Van De Wiele *et al.*, 2007). On the other hand generalist microorganisms such as *Bacteroides thetaiotaomicron* can degrade more than twelve different polysaccharides (Koropatkin, Cameron and Martens, 2012). However, gel-forming insoluble fibres including gums and pectin are not able to be effectively broken down in the gut by the gut microbiome (Cronin *et al.*, 2021). Primary degraders including many Actinobacteria contribute substantially to fermentation efforts (van der Hee and Wells, 2021), although the extent to which most gut microbes can break down all the carbohydrates of the human diet is largely unknown. In fact, only a few fibre types have well characterised metabolic properties in a limited number of enteric microbes. Therefore, researchers still only have a very limited understanding of the benefit provided by the immense diversity of microbes and metabolic genes contained within the gut microbiome.

A further complexity to the dynamics of the gut microbiome is cross feeding between microbes, which is a mechanism of further breaking down dietary substrates by metabolic exchange (Cronin *et al.*, 2021). This can consist of either metabolic cross-feeding, whereby a microbe might utilise the end products from another, or a microbe might use various breakdown products formed during carbohydrate fermentation (Ríos-Covián *et al.*, 2016). With thousands of microbial taxa inhabiting the gut, it is beyond the limits of our current technology to determining the extent of their interactions. Previously isolation and co-culture experiments have been used to identify the metabolite profile from cross-feeding microbes (Seth and Taga, 2014), although these method are laborious and expensive, they provide sound evidence for simple interactions between microbes. Also batch fermentation and bioreactors with model community inoculum are used to study more complex interactions (Bengtsson-Palme, 2020), however because metabolites can accumulate unlike a human gut (where they would be continually be absorbed), these models have some limitations. Metabolites most commonly exchanged between microbes include acetate, lactate, succinate, and to a lesser extent amino acids and vitamins (Saa *et al.*, 2022). Using isotope labelling, it was observed that acetate was readily converted to butyrate, and less so butyrate was converted to propionate, and almost no conversion happened between acetate and propionate (den Besten *et al.*, 2013). Cross feeding between species of *Bifidobacterium* is well understood in breaking down human milk oligosaccharides, and *Bacteroides* spp also are well known to promote cross feeding with their degradation process (Saa *et al.*, 2022). Cross feeding can also promote an environmental niche for other microbial

species to colonise, for example *Bacteroides thetaiotaomicron* has been shown to facilitate the colonisation of *Faecalibacterium prausnitzii*, and *Salmonella enterica* (Coyte and Rakoff-Nahoum, 2019). Interestingly, *Bacteroides ovatus* engages in costly "public" extracellular inulin degradation, despite being able to directly import and degrade inulin (Rakoff-Nahoum, Foster and Comstock, 2016). This trait is thought to have evolved to allow for cooperation within the diverse gut ecosystem. More recently, direct invitro experiments, using metabolic inference from genome sequences, and even ecological modelling are being used to investigate this area (Saa *et al.*, 2022). However, these methods can only be used to predict the genetic capacity of microbes and likely interactions.

As discussed earlier, diet is the main factor which can drive microbial composition in the gut, although what has become highly relevant more recently is the sensitivity of the microbiome to stress (Bastiaanssen et al., 2021). Chronic periods of psychological stress are becoming more common, and studies involving humans and rats have illustrated the link between an altered gut microbiome composition and stress (Gubert et al., 2020). More directly, the gut microbiome has been shown to contribute to the a number of stress induced consequences including dysregulation of hypothalamic-pituitary-adrenal axis, impaired cognition, inflammation, and damage to gut barrier function (Gubert et al., 2020). The deleterious effects on intestinal permeability can lead to further inflammation, immune activation, and further intestinal permeability (Gilbert et al., 2018). Additionally, the stress response can change the rate of intestinal motility and mucus production which can exacerbate microbiota disruption (Fung, Olson and Hsiao, 2017). The changes observed to the composition of the gut microbiome during stress have been demonstrated in Wistar rats, showing increases in Ruminococcaceae and Lachnospiraceae families. In Rosa26-LSL-Cas9 knockin mice, stress caused more volatility (degree of compositional change over time) to the microbiome composition compared to control mice. In humans, reduced abundance of Bacteroidetes, and increased abundance of Firmicutes has been associated with prolonged stress in soldiers during military training (Karl et al., 2017). Although, in a natural setting, stress can elicit a variety of different behaviours (e.g. insomnia, non-homeostatic eating, sedentary behaviour, and hyperarousal (Bonnet and Arand, 2010; Maniam and Morris, 2012)) that can each impact the microbiome differently. It therefore would be particularly difficult to define a microbiome characterised by stress.

1.1.3 Microbial metabolites as signalling molecules

As mentioned above, one of the most well-studied functions of the gut microbiome is the expansion of the hosts metabolic potential, indeed, the human microbiome project has compiled a catalogue of 3.3 million unique protein encoding microbial genes from human gut microbiota (Qin *et al.*, 2010). While the metabolites produced in the gut are

an important energy source for the host, a number of microbial metabolites contribute additionally to host physiology and maintaining homeostasis in most major systems of the body (Cryan *et al.*, 2020). These include molecules which regulate gut motility (Martin *et al.*, 2018), immune response (Fung, Olson and Hsiao, 2017), and neurological signalling (Cryan *et al.*, 2020). One important group of metabolites produced exclusively by the resident microorganisms through an intricate web of co-metabolism are short chain fatty acids (SCFA). These acids consist of a 1-6 carbon molecule chain, and a carboxylic acid functional group. In the digestive tract, these acids are most concentrated in the colon, and occur at the approximate molar ratio of 60:20:20 for acetate, propionate, and butyrate respectively (Verbeke *et al.*, 2015). Approximately 90-95% of the SCFA produced in the gut are absorbed by the host, and contribute around 10% to the daily energy availability (Ziętek, Celewicz and Szczuko, 2021). Acetate is produced by many enteric bacteria, including most Bacteroidetes (Den Besten *et al.*, 2013). Butyrate is produced by many members of the Clostridium cluster XIVa, and by Lachnospiraceae, as well as some other genera within Firmicutes; and propionate is produced by Bacteroidetes and Negativicutes (Yadav *et al.*, 2018).

Regulation of molecules - including SCFA - from the gut lumen, into the blood stream occurs at the gut wall, which is an important site where microbes and microbial metabolites interact with the host. By induction of genes encoding tight-junctions between the cells, butyrate promotes the functioning, and integrity of the epithelial barrier, and this is especially important during times of inflammation and stress (Venegas *et al.*, 2019). SCFA also contribute to controlling inflammation both by activating G protein-coupled receptors (GPR) in colon epithelial cells, and by inhibition of the inflammatory response by monocytes (Venegas *et al.*, 2019). Butyrate in particular assists in controlling inflammation by inducing T cell differentiation which controls intestinal inflammation (Ríos-Covián *et al.*, 2016). The action of microbial metabolites including SCFA are however, not restricted to the gut lumen, and have far reaching impacts in many other body systems. As leading signalling molecules, acetate, propionate, and butyrate are also involved in hormone production (Ramos, Hernández and Blanca, 2009; Verbeke *et al.*, 2015), including Glucagon-like peptide-1 (Yadav *et al.*, 2013), and other molecules which are involved in energy homeostasis (He *et al.*, 2020). In particular, acetate, butyrate, and propionate interact with GPR, which have downstream effects on suppressing appetite (Martin *et al.*, 2018). These acids also regulate fatty acid metabolism in liver and muscle cells, as well as activate thermogenesis in brown adipose tissue, all of which contribute to controlling body mass (Den Besten *et al.*, 2013). Propionate alone is primarily transported systemically to the liver where it plays a key role in liver mitochondrion function (He *et al.*, 2020). Gut microbes also communicate with gut endocrine cells, which can release more than 20 different signalling molecules into systemic circulation, which can interact with the central nervous system (Martin *et al.*, 2018).

The relationship between brain function and the gut microbiome is described by the gut-brain-microbiota axis (Giri and Sharma, 2022). Bidirectional communication between the central nervous system and gut microbiota is possible through the enteric nervous system, and neuroimmune and neuroendocrine systems using microbial metabolites as mediators (Martin *et al.*, 2018). These signaling molecules are primarily received by enteroendocrine cells, and neuroendocrine cells, but also are passed into the blood where they may pass the blood-brain barrier (BBB) (Martin *et al.*, 2018). Additionally, microbiota can produce or mediate the production of neuroactive compounds that might affect brain chemistry, or the permeability of the BBB including peptides, amino acids, decarboxylated amino acids, SCFA, and phenolic compounds (Giri and Sharma, 2022). Concerning these neuroactive compounds, dopamine and norepinephrine are involved with memory and cognition, and have been identified within the cells of gut commensals *Bacillus mycoides* and *B. subtilis,* while the production of norepinephrine has been observed in *Bacillus* spp *Escherichia*, and *Saccharomyces* spp, with *Bacillus* spp also producing dopamine (Giri and Sharma, 2022). Serotonin is involved in limiting nerve impulses, suppressing hunger, gut motility, and influencing behaviour is produced from the essential amino acid tryptophan, the availability of which is regulated by gut microbiota. The majority of serotonin (95%) is produced and stored in endocrine cells along the gut epithelium, and serotonin is also produced by gut microbes including *Clostridium sporogenes* and *Ruminococcus gnavus* from luminal tryptophan (Giri and Sharma, 2022). Using metagenomic data *Akkermansia*, *Alistipes* and *Roseburia* have also been predicted to produce serotonin in the gut (Valles-Colomer *et al.*, 2019). Lastly, early life gut commensals *Bifidobacterium* and *Lactobacillus* take part in the production of gama-aminobutyric acid, which through the vagus nerve, has positive effects on the brain and depressive behaviour (Giri and Sharma, 2022).

Maintaining the diverse composition of commensal gut microbiota is important for long-term health, as the gut microbiome can otherwise function in a way that is disadvantageous to the host. Microorganisms can induce a pro-inflammatory response, which reduces the integrity of the gut wall and allows bacteria-derived toxins (phenols, and *para*-cresol) and cell wall components (lipopolysaccharides) to enter the bloodstream, which leads to further immune activation (Ding, Taur and Walkup, 2017). The microbiome is most susceptible to perturbation when it is not fully developed, meaning the colonisation process and early life are the most important time periods for determining overall gut microbiome resilience and function.

1.1.4 The colonization process imposes long-lasting effects on community structure

Colonization of the infant gastrointestinal tract by commensal microorganisms may be the most important process for maintaining the long-term stability of the microbiome.

This is because the microbiome is especially vulnerable, during early colonization at infancy due to low diversity and stability (Matamoros *et al.*, 2013). This process is thought to largely begin during and after delivery, but there is also evidence to suggest that some vertical transmission of microbes occurs during pregnancy from mother to infant. The presence of a low abundance of bacteria in the placenta, amniotic fluid and meconium is raising questions around the sterility of the uterine environment (Collado *et al.*, 2016). The first meconium, the infant's first bowel movement, is formed from amniotic fluid that was swallowed throughout pregnancy by the foetus. Live bacteria have been isolated from meconium samples, indicating that live bacteria were swallowed with amniotic fluid (Jiménez *et al.*, 2008). Further work by the same group has shown genetically labelled bacteria can be found in the gut of pups when dams are orally inoculated by those same bacteria, indicating a hematogenous route from the maternal oral cavity or gut to the foetal gut. In another study involving mice, *Fusobacteria* of the mouse oral microbiome were documented binding to endothelial cells and then later colonizing the placenta. Additionally, these bacteria modified the integrity of the endothelial barrier, which the authors predict facilitated the dissemination of other microorganisms (Fardini *et al.*, 2011).

Although, because the biomass of microbial cells recovered from the uterine environment and meconium is vanishingly low, it is difficult to differentiate between contamination and potential true colonisers; indeed recent studies of the meconium using robust negative controls have suggested there is no microbiota present in the gut prior to birth (Dos Santos *et al.*, 2021). Two other studies using sequencing technology, have reported on the microbiome of first meconium without the use of any controls (Nutricionist et al., 2020; Collado et al., 2016), which is not an appropriate methodological choice, and should causes their results to be questioned. On the other hand, Rackaityte et al (2020) described the microbiota in the foetal samll intestine meconium using 87 procedural and technical controls, which accounted for 36% of the total samples sequenced. This group described the molecular signal in their samples as "nearing the limits of detection", and supported their claims with images of bacterial cocci identified in the terminal ileum of human foetal intestines obtained from terminated pregnancy using scanning electron microscopy. With the integral role the microbiome plays in mammalian health it might be expected that microbiota with the fitness to translocate and colonise the nutrient poor foetal gut would be closely retained through generations, rather than strictly passively acquired from the environment; and that potentially their presence in meconium might be below the limits of detection using sequencing and culturing methods. However, until sufficient reproducible evidence is collated the sterile womb hypothesis should not be rejected.

During pregnancy the maternal microbiome changes significantly, mainly due to a reduction in richness and increase in Actinobacteria (Koren *et al.*, 2012). Also a shift in

both immune function and metabolism, which are both influenced by gut microbiota, occurs during pregnancy (Yang *et al.*, 2021). These metabolic shifts involving host and microbiota that may have historically been beneficial in providing nutrients to the developing foetus when resources were scarce, may now be associated with gestational diabetes and excess weight gain (Koren *et al.*, 2012; Delhaes *et al.*, 2018). In particular, consumption of a high fat diet during pregnancy has been associated with changes to gene expression in the foetus of primates (Cox *et al.*, 2010), and impairments in social behaviour in male mice (Buffington *et al.*, 2016).The maternal microbiome is not only an important potential seeding location for microbial colonizers, but microbial activity may also be important as metabolites can be transferred to the foetus, depending on the regulatory action of the placenta (Pessa-Morikawa *et al.*, 2022). The interplay between physiological response to stress and the microbiome (as discussed previously) may also be important to consider during pregnancy, as this can be a particularly stressful period that can cause shifts to gut and vaginal microbial communities (Yang *et al.*, 2021). Maternal stress has been shown to impact the composition of the infant microbiome in the first 110 days after birth (Zijlmans *et al.*, 2015), as well as impact neurodevelopment and behaviour including increased anxiety and lack of attention in offspring (Yadav *et al.*, 2018).

During this critical time in development, birth mode has been shown to have a significant impact in determining the profile of initial colonisers, indicating the fundamental importance of those very first microbial strains (Yang *et al.*, 2021). First exposure to vaginal microbes during a natural birth is proposed as being a protective, compared to first exposure to skin associated microbes during caesarean birth (Dunn *et al.*, 2017). Vaginal birth is associated with enrichment of anaerobic bacteria including *Lactobacillus* which can be acquired when the baby passes through the birth canal (Yang *et al.*, 2021). Infants born via caesarean section will never be exposed directly to the mother's vaginal microbiota, and are instead more often colonised by facultative anaerobes including *Enterobacter*, *Klebsiella*, and *Staphylococcus* (Yang *et al.*, 2021). Vaginal microbiome transfer has been used to attempt to restore the microbiome of caesarean born infants (Dominguez-Bello *et al.*, 2016; Wilson *et al.*, 2021), yet there are a number of complications that severely limit any potential benefit of this practice 1) vaginal microbes likely adhere strongly to the dry gauze used to collect them, and therefore limit any transfer to the infant 2) transfer of the microbiota to the infant does not happen directly after birth, but a number of minutes later. 3) the swab or swab solution must remain at ambient temperatures and is exposed to environmental contamination which may severely alter the composition of the microbes transfer to the infant. Additionally, antibiotics taken during childbirth can reduce the abundance of probiotic species such as *Bifidobacterium* in breastmilk, as well as the abundance of *Bifidobacterium* and *Lactobacillus* in the infant gut microbiome (Dunn *et al.*, 2017). Caesarean birth can also cause delays, and difficulty in breastfeeding, which may help

explain the enrichment of *Bacteroides* spp. in the gut of infants born vaginally (Galazzo *et al.*, 2020). This bacterium is one of the most commonly found bacteria in human breast milk (Zimmermann and Curtis, 2020), and potentially, the impacts of caesarean delivery on breastmilk composition and breastfeeding success impacts the colonisation of *Bacteroides* in the infant gut.

After birth, the mother's skin, and breastmilk also, provide microorganisms that can colonize the infant gut (Yang *et al.*, 2021). Breast feeding provides a rich source of human milk oligosaccharides (HMO), proteins, vitamins, immune cells, and bioactive compounds such as hormones (Carrothers *et al.*, 2015), as well as microbial strains that can colonise the infant gut (Carrothers *et al.*, 2015; Meyer *et al.*, 2016). This is particularly the case with colostrum, which is breastmilk that is produced in the first four days. Bacteria are theorised to translocate to breast milk by the action of Dendritic cells, which select them from the gut and carry them to the mammary glands (Rodríguez, 2014). The composition of breast milk is also affected by maternal diet (Bravi *et al.*, 2016), and the functional capacity of the milk microbiome inferred by metagenomic analysis has also been shown to be impacted by diet (Meyer *et al.*, 2016). Healthy breast-fed infants show lower richness and diversity, with higher abundance of Actinobacteria, and a higher quantity of lactic acid within faeces, compared to formula fed infants (Bazanella *et al.*, 2017). Only a select number of bacteria, most notably *Bifidobacterium infantus*, have the potential to break down HMOs and subsequently change the gut environment by depleting the oxygen concentration and lowering the gut pH during fermentation (Casaburi *et al.*, 2021). Furthermore, the HMOs bind to some pathogenic species, which interferes with their ability of to adhere to the infant gut (Cacho and Lawrence, 2017).

During the colonisation process, a succession of microbiota including *Bifidobacterium*, *Closteridium* and *Lactobacillus* first inhabit the healthy infant gut (Hill *et al.*, 2017). Among fungi, Saccharomycetalean yeasts have been the first members to be detected in the infant gut (Hallen-Adams and Suhr, 2017); however, within the first 30 days of life, *Candida* species were found to be the most abundant (Ward *et al.*, 2018). By 2 to 5 years of age, the microbiome will resemble that of an adult, and while this process is not well understood, a succession of particular bacteria is required if the microbiome is to attain a beneficial composition (Wampach *et al.*, 2017); and the microbes that arrive earlier can put additional constrains on further colonisers through niche modification (Koskella, Hall and Metcalf, 2017). This promotes favourable development of the infant immune system and brain, and can influence the long-term health of the host as discussed in a number of reviews (Matamoros *et al.*, 2013; Lynch, 2016; Ding, Taur and Walkup, 2017; Hirata and Kunisawa, 2017; Cryan *et al.*, 2020). This work has led to the hypothesis that delayed, or disrupted colonization by 'traditional' co-evolved microbes will increase the host's susceptibility to auto-immune disease and neurological

disorders; and therefore, the best time to intervene in an attempt to rectify the microbiome may be before or directly after birth.

1.2 The Microbiome is a Window for Understanding Host Health

1.2.1 Modernization and inflammation in the gut

It is now widely recognized that the gut microbiome plays a crucial role in maintaining not only gut health, but overall wellbeing, from improving susceptibility to pathogens, to coping with stress (Sharma, Neu and Young, 2010; Amon and Sanderson, 2017; Gupta *et al.*, 2019). As human society has modernized, the level and timing of exposure to microbes and microbe carrying agents has been altered among a large fraction of people. In industrialised areas, newborns first enter the world more commonly in hospitals and increasingly often via caesarean birth, potentially increasing exposure to skin and hospital associated microbiota (Chong, Bloomfield and O'Sullivan, 2018). The microbial composition of indoor environments has also been modified due to increased modernised sterilising cleaning practices (Gupta *et al.*, 2019). Furthermore, exposure to antimicrobial compounds has increased due to their inclusion in numerous products including toothpaste and toys (Sanidad, Xiao and Zhang, 2018), as well as the increasing use of emulsifiers, preservatives and other artificial food additives that alter microbiota composition (Cao *et al.*, 2020). Each of these aforementioned changes has happened rapidly over a very short period of human history, which is likely to compound their effects on human health.

The evolutionary capacity of the microbiome is immense. Microbes such as E. coli (wild type) can replicate in under 20 minutes (Ball, 2012), and horizontal gene transfer from the environment or from another cell can facilitate adaptation without replication, and is most common among host associated microbes (Koskella, Hall and Metcalf, 2017). This means that ecological changes in microbial abundance overlap with changes in gene frequency, making de novo mutation – and therefore rapid adaption – highly relevant to the gut microbiome. The human microbiome has been augmented by exposure to antimicrobial agents with significant gains in antibiotic resistance genes (Yassour *et al.*, 2016; Kovtun *et al.*, 2020), and losses in predominant commensal members (Yassour *et al.*, 2016). Gut microbiota have also responded to diets containing a larger proportion of processed and preserved foods by an influx of *Bacteroides* species and a reduction in historically predominant *Prevotella* species (Clemente *et al.*, 2015).

Understanding the gut microbiome, and its response to diet and other external factors is frequently facilitated through studies involving murine models. Unlike human populations, the genetics, and the environment (e.g., sterilised food water and housing)

of animal models can be rigorously controlled, which allows manipulations to the microbiome to be investigated without confounding variables. There are however differences in the structure of the gastrointestinal tract between humans and mice/rats including a less acidic forestomach and large cecum, as well as differences in the presence and abundance of bacterial genera (Hugenholtz, and de Vos, 2018). Most of the murine models used for human gut microbiome research are wild-type, inbred strains, as well those altered through transgenesis with foreign DNA, and thus have an altered phenotype (Hugenholtz, and de Vos, 2018). Using inbred animal models in modern vivarium means research is more easily reproduced, but less likely to be translatable to humans, especially regarding genetic heterogeneity (Porsgaard *et al*., 2016). Additionally, both the strain, husbandry, and the vender source can impact the microbiome of the murine model (Ericsson *et al*., 2015), which can impact the study outcomes if not considered.

Despite these limitations, animal models have been important tools in microbiome research, and have been used to better understand how a modern lifestyle may permanently alter the microbiome through the loss of key species. By reducing the accessibility to fermentable carbohydrates over generations, the microbiota of a rat model increasingly departed from the structure present in control rats, and the losses in diversity could not be recovered with diet switching after 4 generations (Sonnenburg *et al.*, 2016). Although, because germ free rats have suppressed immune function, and were inoculated with a single human faecal microbiome (which is not adapted for the gastrointestinal anatomy of a rat and would afford all rats the same microbiome-associated phenotype), it is not surprising that the rats lost microbial diversity over 4 generations with no opportunity for bacterial exposure. While these results are unlikely to reflect human conditions, they do illustrate the response of the microbiome to hyper-modernization. This is particularly important considering hyper-modernisation has coincided with a rise in non-communicable diseases, and the impact on the gut microbiome may be a contributing factors to the increased prevalence of these conditions (Gray *et al.*, 2017).

Originally, proposed in the late 1980s, the hygiene hypothesis (Gupta *et al.*, 2019) was used to explain the shift away from disease and toward allergies, by a lack of exposure to infection in childhood. The theory has been revised over time, and now extends to microorganisms that are not pathogenic, and is used to understand inflammatory autoimmune diseases as well (Gupta *et al.*, 2019). A loss of exposure to a diversity of microbes overtime is thought to be disadvantageous to the regulation of the immune system; and the protective effect of exposure to endotoxins in reducing the incidence of allergic disease has been demonstrated a number of times in comparing the allergy rates among children in farming and rural settings compared to metropolitan areas (Ege, 2017). What has been observed over the last few decades in western countries, and more

recently in developing areas, is an increase in autoimmune and atopic diseases, especially in young children (Loh and Tang, 2018). The importance of the gut microbiome is continually being realized, and has most recently been regarded as an organ (Moles and Otaegui, 2020). This analogy is fitting as the gut microbiome has evolved to perform an essential function like all other body organs, and removing or destroying it has severe health ramifications.

1.2.2 The microbial role in immune system development and autoimmune disease

Microbes evolve on a timescale that is profoundly different to that of the host, and to interface with a rapidly changing environment the adaptive immune system alone is highly responsive to change. Gastrointestinal bacteria stimulate the development of the gut-associated lymphoid tissue that reside within the intestinal mucosa, which is the largest collection of lymphocytes in the body (Faria, Reis and Mucida, 2017). At birth, the innate immune system has however yet to mature, and therefore, the nonspecific first line of defence that would normally be provided by this system, has not yet been fully developed. Newborns possess monocytes and T cells that produce fewer cytokines than adults, and have lower levels of memory T cells (Holt and Jones, 2000). In early life, the immune system is also shifted towards a Th2 phenotype which can facilitate the colonisation of gut microbiota (Dzidic *et al.*, 2018). Th1-type cytokines induce a proinflammatory response that can respond to infection, and can lead to an autoimmune response, whereas the Th2-type cytokines have more anti-inflammatory activity. Additionally, the physical and chemical barriers of the intestine are incomplete, and there is limited secretory immunoglobulin A (IgA) and M (IgM) production (Holt and Jones, 2000), although IgA is provided through breastmilk. Over time, the tissue-resident collection of immune cells matures in early life through exposure to external agents including toxins, nutrients, as well as commensal and pathogenic microorganisms. Early life is therefore a critical period when the immune system and gut develop in tandem, and immune cells must adapt to the ever-changing environment.

The immune system can respond by participating in the inflammatory process, or by expressing tolerance (Faria, Reis and Mucida, 2017), and the responses elicited at the mucosal surfaces during this process determine in part the allergic response (Holt and Jones, 2000). The gut microbiota produce or mediate metabolites that regulate the host's immune response including shaping Peyer's patches, and inducing the Th1 and Th2 type responses (Dzidic *et al.*, 2018). The regulatory action of these metabolites occurs through a number of different mechanisms including immune cell signal transduction, acting as transcription factors, and regulating gene expression through GPR (Hirata and Kunisawa, 2017). Microbial metabolites with immunomodulatory properties include vitamins such as B9 which maintain Treg cell function and reduce inflammation, and

amino acids including tryptophan which when metabolised to melatonin reduce inflammatory cytokine production (Hirata and Kunisawa, 2017). SCFA are also immunomodulators with butyrate promoting healthy epithelial barrier function, and propionate enhancing the generation of Dendritic cell precursors with reduced capacity to promote Th2 cell function (Hirata and Kunisawa, 2017). Together, top-down and bottom-up signalling between the immune system and gut microbiota helps shape the long-term functioning of the immune system, determining the response to environmental antigens.

Many of the risk factors associated with autoimmune disease are linked to the development of the gut microbiome, including: lack of breast-feeding (Friedman and Zeiger, 2005), maternal and infant antibiotic use (Dzidic *et al.*, 2018), and reduced exposure to microbe carrying agents like pets or livestock (Lynch, 2016). The gut microbiota shows characteristic shifts associated with a range of diseases, as well as autoimmune pathological conditions, and changes to the microbiome, especially in early life are predicted to be contributing to the rise of autoimmune diseases (Gray *et al.*, 2017). While the incidence of newborn autoimmune disease is rare, allergic diseases are increasing in young children and adolescents, and are currently the most common chronic disease in childhood (Asher *et al.*, 2021). There are currently over 80 different autoimmune diseases (Roberts and Erdei, 2020), with Crohn's disease and ulcerative colitis being two inflammatory bowel diseases extensively studied and closely associated with the composition of the gut microbiome (Venegas *et al.*, 2019). In both diseases, the composition of the microbiome is different compared to in healthy controls, and both bacteria and fungi are thought to be involved in the inflammatory process.

Like inflammatory bowel disease, the microbiome of children with allergies has also been described in terms of its differing structure and function; and unfortunately, the morbidity of allergies among children appears to be increasing. For example the likelihood of developing food allergies has been shown to increase in relation to characteristic changes of particular bacterial species during the first few years of life (Tanaka *et al.*, 2017); and pro-inflammatory metabolite profiles along with microbial deficits are able to distinguish groups of infants with different relative risk of developing atopy and asthma (Fujimura *et al.*, 2016). Allergic diseases including atopic dermatitis (Fujimura *et al.*, 2016), food allergy (Prescott *et al.*, 2013), and asthma (Braman, 2006) are most prevalent in children in developed countries and in urban settings. For example, in Australia, the prevalence of food allergy in infancy is estimated at 11%, and the incidence of food anaphylaxis emergency admissions have increased from 2.0 per ten thousand in 1998-1999 to 8.2 per ten thousand in 2011-2012 (Tang and Mullins, 2017). Concerningly, the prevalence rates of food allergy in Australia are the highest in the world (Bellinger *et al.*, 2019), and allergic rhinitis and eczema are also increasing in

prevalence (Dharmage, Perret and Custovic, 2019). The most common allergic disease in children globally however, is asthma (Asher *et al.*, 2021), with boys more affected in childhood, and girls more affected in adolescence (Dharmage, Perret and Custovic, 2019). It is suspected that there is no increase to the incidence of asthma in high-income countries (Dharmage, Perret and Custovic, 2019), however there has not been sufficient global evidence to support this hypothesis. Since 2003, there has only been a single global assessment of asthma prevalence (the Global Asthma Network Phase I) which showed childhood and adolescent asthma increased in prevalence in middle-income countries, decreased in prevalence in low-income countries, and did not change in high-income countries (Asher *et al.*, 2021). Although, in this study Australia, Canada, Germany, Norway, US, and the UK were among the countries not surveyed, which appears to be a sever limitation to the conclusions of this study regarding high income countries.

1.2.3 The gut-brain-microbiome axis as a modulator of brain function

As discussed in section 1.1.3, bi-directional communication between the microbiota and the host is possible using microbial metabolites as mediators, and this link is called the gut-brain-microbiota axis. While this link is bi-directional, the brain is more often on the receiving end of communications, with 90% of vagal fibres conducting signals towards the brain (Rao and Gershon, 2016). Like the immune system, the development of the brain is impacted by the gut microbiota in early life, and is associated with autoimmune encephalomyelitis, ASD, and neurodegenerative diseases including Parkinson disease, multiple sclerosis, and Alzeimer's disease, (Cryan *et al.*, 2020), as well as the stress response, anxiety- and depressive-like behaviours, and food preferences (Martin *et al.*, 2018). The inclusion of the microbiota into the gut-brain axis has been a relatively recent paradigm shift in the way neurological disorders are investigated, and animal studies have provided most of the detailed evidence for specific processes in which the microbiota are involved in this complex system. As previously mentioned in section 1.2.1, the evidence generated using animal models are often not translatable to humans. While these models have been of great importance in uncovering plausible mechanisms for the involvement of the microbiome in neurological functioning - including neural development, myelination, neurogenesis, and microglia activation through immune pathways (Cryan *et al.*, 2020) – they have not been able to demonstrate causality.

The neurological disorders associated with the gut-brain-microbiota axis share several similarities. Among neurodegenerative diseases, Parkinson disease (PD) and Alzheimer disease (AD) involve self-aggregating proteins (alpha-synuclein and amyloid precursor protein respectively) which are used by both the CNS and the ENS, and can cause a similar disease pathology in enteric nerve cells (Rao and Gershon, 2016). PD, AD and ALS are also associated with elevated levels of aluminium in the brain, which is both

neurotoxic and inflammatory (Kawahara, 2005), can increase oxidative stress, and can increase the risk of accumulating self-aggregating proteins which form plaques in the brain (Sanajou *et al*., 2022). Potential involvement of the gut microbiota in these processes include: the assembly of amyloid fibres (which share biochemical and structural properties with disease-linked amyloids) externally on microbial cell surfaces (Pistollato *et al.*, 2016), stimulation of ENS neurogenesis by microbially regulated serotonin (De Vadder *et al.*, 2018), and shifts in microbial metabolism of serotonin and cholesterol after aluminium exposure (Feng *et al*., 2022). In addition, high levels of aluminium have been found in the brains of 5 autistic individuals, indicating a potential link between high aluminium accumulation in the brain and the disease (Mold *et al*., 2018; Exley *et al*., 2020).

Another characteristic that is shared among a number of neurological disorders including multiple sclerosis (Cryan *et al.*, 2020), ASD (Fiorentino *et al.*, 2016), and dementia (Parker, Fonseca and Carding, 2020) is compromised BBB integrity. This barrier is responsible for gatekeeping the passage of molecules in the blood to the brain, and the proteins responsible for maintaining tight junctions between the cells of the BBB have decreased expression in specific-bacteria free mice (Braniste *et al.*, 2014). Also, in mice it has been demonstrated that SCFA improve the number and function of microglia cells which are responsible for the regulation of the BBB, and for controlling neuroinflammation. A final similarity among neurological disorders is impaired gut barrier function, which can allow for improper uptake of luminal contents including bacterial metabolites and toxins. Increased gut permeability may progress through activation of either enteric neural or glial cells (Rao and Gershon, 2016).

In concert with a leaky gut, exposure to microbially produced toxins and metabolites have been linked to the pathology of neurological disorders (Rao and Gershon, 2016), and some bacteria produce neurotoxic compounds. A particular strain of *Bacteroides fragilis* (enterotoxigenic B. fragilis) can produce pro-inflammatory lipopolysaccharide (LPS) as well as a metalloprotease called *fragilysin* which is one of the most potent inflammatory enterotoxins known. Both the LPS and *fragilysin* can leak through a healthy gut barrier and pass the BBB causing inflammation (Lukiw, 2020). Also, *Clostridium perfringens* produce a number of cytotoxins which may be associated with ASD (Góra *et al.*, 2018). Alternatively, gut microbes can also produce metabolites that can have beneficial impacts on neurological disorders. *Anaerococcus*, *Clostridium*, and *Escherichia* produce trimethylamines that can cross the BBB. These amines have both positive and negative impacts on health, but may be beneficial in neurodegenerative disease as they correct misfolded proteins and maintain the proteins original conformation (Parker, Fonseca and Carding, 2020). The microbiota also catabolise amino acids and therefore regulate the balance between inhibitory and excitatory

neurotransmitters, and produce vitamins in particular B vitamins which are vital for maintaining health (Parker, Fonseca and Carding, 2020).

The shifts that are observed by the microbiota in association with neurological disorders are typically subtle, involving the enrichment or depletion of certain key species rather than largescale shifts in diversity, like that seen with the intermittent active periods of irritable bowel disease. However, because these shifts are fine, and the methods used to detect species enrichment varies among studies, there are many taxa that have been described in the literature that are potentially coincidental findings. Although, a recent meta-analyses including 11 studies implicated enrichment of *Bifidobacterium*, Proteobacteria, and *Phascolarctobacterium* while Clostridiaceae, Firmicutes, Lachnospiraceae and Rikenellaceae were reduced in the gut microbiome of those with Alzheimer's disease (Hung *et al.*, 2022). Another meta-analysis of 10 studies showed *Corprococcus*, *Faecalibacterium*, and Prevotellaceae were decreased in those with major depressive disorder (Sanada *et al.*, 2020); and lastly, two recent meta-analyses have implicated enrichment of *Faecalibacterium,* and *Ruminococcus* and the depletion of *Bacteroides*, and *Bifidobacterium* (Xu *et al.*, 2019), as well as depletion of *Streptococcus* and *Bifidobacterium* (Andreo-Martínez *et al.*, 2021) in the gut microbiota of autistic children. Together these meta-analyses indicate that the balance of certain taxa especially *Bifidobacterium* and members of Closteridia may be of particular importance for maintaining both gut and brain health into later life stages. However, to improve the common knowledge regarding potential microbial biomarkers for childhood disease, both individual studies and meta-analyses must consider the age of participants. For example the most recent meta-analysis collating the microbiota associated with autism included 18 studies with children ranging from 2 to 13.4 years of age (Andreo-Martínez *et al.*, 2021). The composition of the microbiome changes significantly from infancy to adolescences as discussed in section 1.1.4, and it follows that the manifestations of gut dysregulation during disease in infancy might look quite different to that in adolescents. In particular, *Faecalibacterium* is one of the most common later colonisers of the healthy human gut, and therefore, enrichment of this genus at less than 2 years may be interpreted differently than enrichment in adolescents, when more adult dietary preferences begin to emerge.

1.3 Modifying the Microbiome

1.3.1 Probiotics and antibiotics use in adults and infants

The community of microorganisms in the gut of healthy adults is not static, but in the long-term seems to demonstrate prolonged stability that is punctuated by periods of disturbance (Voigt *et al.*, 2015; Fu *et al.*, 2019). Although, in recent times it is possible to drastically alter the composition of the established adult microbiome. This can be

achieved by directly delivering beneficial microorganisms to the gut using probiotics, or by eliminating bacterial microbiota through the use of antibiotics. In general, probiotics have positive effects on the gut by providing commensal microbes that often provide anti-inflammatory metabolites to the gut. Conversely, antibiotics can increase gastrointestinal disease and disrupt the microbial community structure, often by eliminating key community members.

Infections cause acute inflammation and immune system activation and pharmaceutical treatment for a number of infections include antibiotics and antifungals however, these treatments can also cause dysbiosis (abnormal microbiota structure) in the gut. In fact, antibiotic use poses a significant risk for developing irritable bowel disease (IBD) (McIlroy et al., 2018), and increased susceptibility to microbial infection (Francino, 2016). However, antibiotics are also used as a non-standard treatment for conditions that are associated with dysbiosis including IBD, despite the risk of side effects (McIlroy *et al.*, 2018).The most commonly noted changes to the microbiome after antibiotic exposure are significant losses in both richness and diversity, as well as changes to the taxonomic community membership (McDonnell *et al.*, 2021). In a large longitudinal study assessing long and short term diet patterns among healthy adults, antibiotic use up to three months prior contributed most strongly to microbiome composition, and significantly reduced alpha-diversity, which is a measure of species richness and evenness (Klimenko et al., 2018).

While these acute effects of antibiotics are relatively understood and accepted, it is less clear if antibiotics impair the microbiome long-term. The extent of microbiome disruption depends largely on the pharmacodynamic properties of the antibiotic, and the dosage/duration of use (Jernberg *et al.*, 2010), which may make it difficult to compare across studies. A recent meta-analysis of commonly prescribed antibiotics found microbiome composition returned to normal within 6-months according to alpha-diversity measures (Elvers *et al.*, 2020). However, earlier studies show that when looking specifically at certain key taxa such as *Bacteroides* affects can persist for years (Jernberg *et al.*, 2010; Francino, 2016). Another well know example of longer-term dysbiosis after treatment with an antibiotic is the development of intractable candidiasis caused by proliferation of fungi (Paterson, Oh and Underhill, 2017). Antibiotic use can also reduce the number of anaerobic bacteria which help maintain the oxygen depleted environment. This can cause an increase in potentially pathogenic aerobic bacteria such as *Salmonella typhimurium* (Venegas *et al.*, 2019), or *Clostridium difficile* (Francino, 2016). A final consideration for antibiotic use is the continual increase in antibiotic resistant genes that are accumulating in enteric microbes. A population-level analysis of the gut microbiome identified resistance against 50 of 68 antibiotics screened for, and each sample had on average resistance to 21 antibiotic types (Francino, 2016).

Exposure to antibiotics during foetal and early life stages are more harmful to the long-term function of the microbiome as it is still involved in colonisation. The use of antibiotics during these early life stages are linked to increase the risk of allergic and inflammatory conditions including asthma and IBD (Chong, Bloomfield and O'Sullivan, 2018), as well as metabolic conditions including increased body mass (Saari *et al.*, 2015) and diabetes (Elvers *et al.*, 2020). During pregnancy, between 20 – 40% of women in developed nations receive antibiotic treatment (Stokholm *et al.*, 2013), and during labour prophylactic antibiotics are also encouraged prior to caesarean delivery and/or to prevent infection complications such as *streptococcus* (Dunn *et al.*, 2017). Antibiotics are known to cross the placenta and enter the foetal bloodstream. This may happen rapidly, and result in an equilibrium concentration between mother and infant, or an incomplete transfer may occur where the concentration of the antibiotic is lower in the infant compared to the mother (Pacifici, 2006). Surprisingly, infants have also been found to host microbiota with antibiotic resistant genes, many of which are shared between mother and infant (Francino, 2016). Therefore, due to the rising concerns of multi-antibiotic resistant bacteria, probiotics are receiving more attention as a potential treatment and preventative therapy to aid in pathogen infection.

Probiotics may be able to support the proper functioning of the microbiome, assist in restoring the microbiome after perturbation, or potentially reduce the risk of disease (Vitetta *et al.*, 2014). This is due to the fact that various different strains of probiotics have demonstrated the ability to mediate immune functions, contribute to beneficial metabolite production, and also interact synergistically with other commensal microorganisms (Sanders *et al.*, 2019). The most frequently used probiotic species to supplement the human gut microbiome are *Lactobacillus* and *Bifidobacterium* (Cunningham *et al.*, 2021). In rats, these bacteria have been used extensively to demonstrate anti-anxiety properties, and have been shown to have positive effects on anxiety in humans by reducing blood cortisol levels (Cryan and O'Mahony, 2011). Together, species of *Lactobacillus*, *Bifidobacterium*, and *Streptococcus* have also been used to reduce immune mediated inflammation in patients with multiple sclerosis (Cryan *et al.*, 2020), as well as improve both gastrointestinal and core behavioural autism symptoms (Santocchi *et al.*, 2020). *Lactobacillus* alone has also been shown to improve cognitive fatigue, demonstrating the ability of these microbes to promote healthy brain function (Cryan *et al.*, 2020). Probiotics have also been used extensively as a potential therapy for allergic disease, and in a recent meta-analysis probiotics (primarily *Lactobacillus*, and *Bifidobacterium*) significantly reduced nasal and eye symptoms and improved quality of life according to questionnaire (Chen *et al.*, 2022). Lastly, *Akkermansia muciniphila* is a recently identified potential probiotic species that only degrades mucin to source carbon, which means it must attach to the mucosal layer which is positioned closed to the gut wall (Zhai *et al.*, 2019). *A. muciniphila* produces both acetate and propionate, and a membrane protein which benefits gut barrier

function, and has been used beneficially in an animal model to alleviate metabolic disease including diabetes and obesity (Zhai *et al.*, 2019). Although, the microbiota response to probiotics is highly variable because each individual hosts a unique microbiome. Accordingly, some individuals may not respond well to conventionally used probiotic strains, or may respond in a way that is not captured within the confines of the microbiome study design. Hence, more microbiome research which includes an in-depth analysis of long-term diet and gut health is needed to better determine which probiotics are suited to different microbiome types.

1.3.2 Prebiotic supplementation in adults and infants

The gut microbiome can also be manipulated using prebiotics, which are defined currently as "substrate[s] that [are] selectively utilized by host microorganisms conferring a health benefit" (Gibson *et al.*, 2017). Because prebiotics promote larger scale changes in multiple microbial groups, prebiotics can induce a community-wide shift in metabolic activity (Sanders *et al.*, 2019). The majority of prebiotics are inulin-type prebiotics including the most commonly used fructo-oligosaccharide (FOS) and galacto-oligosaccharide (GOS), and potential prebiotics also include resistant starch, polyphenols, polyunsaturated fats and vitamins (Cunningham *et al.*, 2021). However, as the prebiotic industry is growing rapidly, it is becoming more common to see commercial prebiotic blends and prebiotics added to human and animal foods. In some cases, there is little scientific evidence of benefit (Jin Song *et al.*, 2019; Cunningham *et al.*, 2021); but because potential prebiotics are generally assumed to be safe, many trademarked products have been developed with little regulation and poor quality research data (Jin Song *et al.*, 2019). For example, the negative effects of stress in healthy adults are claimed to be alleviated by modifying the microbiome using Amare Fundamentals (a particular blend of probiotic/prebiotic/phytobiotic) supplementation (Talbott *et al.*, 2019). The evidence for this claim comes from increased proportions of "good" bacteria using a PCR based detection method called "BiomeTracker" with unspecified gene targets. The Amare Fundamentals products are marketed with ambiguous claims for which the evidence is not robust enough to support.

However, prebiotics including FOS and GOS have a long history of documented beneficial effects, including promoting proliferation of the beneficial bacteria *Bifidobacterium* and *Lactobacillus* which have been documented in a number of reviews (Macfarlane, Steed and Macfarlane, 2008; Nauta and Garssen, 2013; Ashaolu, Ashaolu and Adeyeye, 2021). These oligosaccharides are also well known to increase acetate and butyrate concentrations (Ose *et al.*, 2018), reduce gut pH and reduce inflammation (Ashaolu, Ashaolu and Adeyeye, 2021). In particular, GOS have been shown to reduce inflammatory markers of metabolic syndrome including C-reactive protein and calprotectin in overweight adults (Vulevic *et al.*, 2013), improve anti-social behaviour in

autistic children (Grimaldi *et al.*, 2018), and improve the absorption of Ca, P, and Mg in Osteopenic rats (Seijo *et al.*, 2019). Although, prebiotics will have a disparate impact on the microbial community due to the diversity of metabolic functions presented by the microbiota community (as discussed in section 1.1.2). As a result, the metabolic profile induced by prebiotics can vary widely from person to person (Sanders *et al.*, 2019). However, strong evidence for the benefits of prebiotics in infants comes from interventions of prebiotic infant formula.

In infants, formula supplementation with FOS and GOS is also associated with similar positive changes in the gut including an increase in *Bifidobacterium* (Moro *et al.*, 2006; Scholtens *et al.*, 2008; Grüber *et al.*, 2010; Sierra *et al.*, 2015), reduced faecal pH (Bakker-Zierikzee *et al.*, 2005; Sierra *et al.*, 2015), increased lactate (Bakker-Zierikzee *et al.*, 2005), and increased secretory immunoglobulin A (Scholtens *et al.*, 2008). Formula supplemented with GOS/FOS have also been found to reduce the incidence of atopic dermatitis in infants (Moro *et al.*, 2006; Grüber *et al.*, 2010). However, a similar clinical trial using only GOS supplemented formula compared to non-supplemented formula found no significant reduction in allergy manifestation in infants (Sierra *et al.*, 2015). The benefits of using prebiotics to support infant health may be enhanced if they are also given to the mother prior to birth, because maternal gut health impacts foetal development. Evidence for modulation of the immune system in offspring with maternal supplementation of FOS/GOS/inulin has been shown in mouse models. Changes to the microbiota composition, a reduction in dermatitis like skin lesions, and an increased Th2 response has been observed in pups from dams fed a FOS supplemented diet during pregnancy and lactation (Fujiwara *et al.*, 2010). More recently, maternal GOS/inulin dietary supplementation in pregnant mice increased the concentration of acetate and other metabolite factors in the amniotic fluid, and also increased the frequency of B $CD9^+$ and $CD25^+$ Breg cells that secrete IL-10 in both foetal and 7-week old pups (Brosseau *et al.*, 2021). These studies support the idea that prebiotics support lifelong health, and may provide some protection against the development of allergies and neurological disorders.

1.4 Investigating the Microbiome Using Genetic and Metabolic Profiles

1.4.1 Sample collection and preparation for microbiome research

Investigating the relationship between the microbiome and the host, including its association with disease requires the community to be sampled, and faecal material is commonly used as a proxy for the gut microbiome. Faecal material is comprised primarily of water, which accounts for 63-86% of the wet mass. The dry faecal matter consists primarily of microbial biomass (25-54%), as well as undigested food residue,

macromolecules and small molecules, and shed colonic cells (Rose *et al.*, 2015). Some of these food compounds such as polyphenols and polysaccharides are known to cause inhibition during PCR, and should be removed during DNA extraction (Wagner Mackenzie, Waite and Taylor, 2015). Once faecal material is voided, it will be exposed to a lower room temperature, higher oxygen concentrations, and airborne microbes, each of which can impact the true microbial assembly. Collecting stool samples into DNA preservation tubes can inhibit microbial growth, and allow participants to return their samples without the need to maintain a cold chain, however these storage techniques have been shown to significantly alter the microbial composition compared to freezer storage (Penington *et al.*, 2018; Byrd *et al.*, 2019; Carruthers *et al.*, 2019; Jones *et al.*, 2021). Storing faecal material for a short period at -20°C or -80°C for longer periods, has been determined as the gold standard for gut microbiome research, including both microbial and metabolite profiling (Cunningha *et al.*, 2020; Hosseinkhani *et al.*, 2021). This allows samples to be collected over larger timescales, but may require participants to temporarily store their stool specimens in the household freezer prior to transporting it (potentially causing a freeze-thaw cycle) to the laboratory (Liang *et al.*, 2020). A final consideration prior to DNA extraction from stool is sample homogenisation, which can affectively reduce the impact of subsampling when a sufficient volume of stool is collected (Gorzelak *et al.*, 2015).

Eliminating the bias that is introduced during DNA extraction and metabarcoding library preparation are mostly unavoidable (Allali *et al.*, 2017). However, using the same extraction protocol, and batch of extraction reagents can mitigate the effects of the "microbial kitome" (Wu *et al.*, 2018). Also, performing sample processing under a laboratory hood will reduce environmental contamination (Wang *et al.*, 2018). To capture contamination that cannot be avoided, such as that from reagents, equipment, and sample to sample cross contamination, the use of negative controls has been strongly encouraged for a number of years in the literature (Murray, Coghlan and Bunce, 2015; Kim *et al.*, 2017; Hornung, Zwittink and Kuijper, 2019). Although, a review of studies using high-throughput sequencing methods by Hornung *et al*. (2019) found only 30% report the use of any negative control. Some studies describe the collection of these samples, but then do not mention sequencing them, (Persoon *et al.*, 2017; Carruthers *et al.*, 2019) or if they are sequenced, are not used to reduce contamination (Kim *et al.*, 2020). For example, the first pass meconium contains a low microbial biomass, and to ensure the resulting taxa do not originate from contamination, it is imperative that investigations of this sample type include negative controls. Despite this, a recently published article Nutricionist *et al*. (2020) collected the meconium of 87 preterm infants from a diaper without any control samples. Of the 63 samples sequenced, non were removed bioinformatically. In comparison 16% of the samples sequenced by, Klopp et al (2022) were controls, which when implemented, resulted in 165/330 meconium samples being discarded.

PCR is required for metabarcoding to amplify a short target region of DNA (generally <300 pb). The reagents used for PCR can also introduce a low level of contamination, but the larger impact here is the sporous formation of chimeric sequences (Schnell, Bohmann and Gilbert, 2015), and the preference for shorter target regions resulting in a size selection bias. Reducing the total number of PCR cycles will mitigate their formation (Bakker, 2018), and amplification for less than 30 cycles has been shown to produce more accurate mock community proportions (Gohl *et al*., 2016), and appears to be the accepted standard in human microbiome analysis (Prodan et al., 2020; Brosseau *et al*., 2021; Penington *et al*., 2018; Fu *et al*., 2019), although some groups (Martin., *et al* 2016; Stinson *et al* 2019) report 40 cycles which is likely to result in high numbers of chimeric sequences or poorer data quality. An attempt to remove chimeras can be made bioinformatically, and the most popular method currently is using removeBimeraDenovo in the DADA2 package (Callahan *et al.*, 2016). Many studies advocate for the use of positive controls, including mock communities during PCR amplification, however, their use in the literature is still lacking (Hornung, Zwittink and Kuijper, 2019). Mock communities can indicate differences in primer binding preference between species (Bakker, 2018), more easily illustrate contamination, and evaluate reproducibility across sequencing runs (Claassen-Weitz *et al.*, 2020). The equimolar blending of DNA from each sample and mock community, as well as the total number of samples will each affect the coverage per sample (Pollock *et al.*, 2018). These parameters can fluctuate across sequencing runs, but can be captured and easily visualised within a mock community positive control.

1.4.2 Sequencing microbial genes and genomes

Next generation sequencing (NGS) technology has become one of the most important methods for investigating the human microbiome, and has expedited an explosion of research in this area (Waldor *et al.*, 2015). This has been in part due to a hundred-fold decrease in the cost of sequencing, as well as facilitating the description of microbial inhabitants across all domains of life, including some 90% which cannot be cultured under laboratory conditions (Ranjan *et al.*, 2016). The reduced cost of NGS has also allowed studies to expand by including larger cohorts of participants, spanning a number of years, or by being conducted over a finer scale. The scaling up of microbiome research has also benefitted from advances in high sample throughput, including the application of barcode indexing (Sinclair *et al.*, 2015). Primers that incorporate these unique, short nucleotide sequence barcodes allow hundreds of samples, each containing thousands of amplicons, to be pooled together and sequenced simultaneously. The barcodes can be added during the primary PCR amplification step, i.e. DNA is amplified by primers containing barcodes, or DNA can be amplified and diluted, and then a second round of PCR can be used to add the barcodes (Sinclair *et al.*, 2015).Illumina is the predominate NGS short read sequencer in the industry, and the MiSeq instrument is

the most extensively used to understand microbial communities including those of the gut microbiome (Pollock *et al.*, 2018; Zhang *et al.*, 2021). The sequencing by synthesis approach used by the Illumina instruments provides reads with minimal base-call errors (Allali *et al.*, 2017); and to improve the accuracy even further, amplicons can be read and sequenced in both the forward and reverse direction allowing the paired-end sequences to partially overlap. A recent comparison of the Illumina sequencing platforms showed paired-end sequencing experiments keep error rates between 0.087% – 0.61% across the 7 platforms tested (Stoler and Nekrutenko, 2021). The NGS technology is however not without limitations, which are primarily due to read length (Bukin *et al.*, 2019). Third generation sequencing (TGS) technology does not suffer this same dilemma, and instruments by Pacific Biosciences (PacBio) can sequence reads of over 2500 bp on average (Zhang *et al.*, 2020), but can also yield reads of 10K in length (Au *et al.*, 2012). Although most studies today are still using NGS technology, as the higher cost, and few analysis pipelines keeps TGS technology out of reach for most (Abellan-Schneyder *et al.*, 2021).

Molecular investigations of the gut microbiome using high throughput NGS, rely on the identity of sequenced reads from pooled genetic material being accurately determined. The primers used for metabarcoding will impact what microbes can be identified and to what resolution (Bukin *et al.*, 2019). To target bacterial communities, including those of the gut microbiome using NGS technology, the 16S rRNA gene is the gold standard (Zhang *et al.*, 2020). Both the earth microbiome project (earthmicrobiome.org), and the human microbiome project (hmpdacc.org/hmp), have developed standard operating procedures based on this region. Although, one consideration when targeting the 16S gene, is that different bacterial clads will contain different copy numbers of this gene in their genome, meaning bacteria with high copy numbers will be over represented in community profiling based on read abundance (Louca, Doebeli and Parfrey, 2018). While this genetic feature means the prevalence of a particular amplicon sequence variant (ASV) may be very different from the prevalence of the host bacterium, because the copy number variation is assumed to be consistent among the same ASV, it will not affect comparisons of the same ASV between microbiome samples. The 16S rRNA gene sequence is evolutionarily conserved, which allows for primer binding, and contains hypervariable regions (V1 through to V9) which are used to identify microorganism phylogenetically (Yang, Wang and Qian, 2016). Amplicon length is limited by NGS technology, and because the 16S gene sequence is ~1600bp in length, partial sequences spanning one or two of the nine hyper-variable regions are more generally used (Yang, Wang and Qian, 2016). However, longer reads contain more information compared to shorter reads, and as a result, taxonomic assignments, particularly at the genus and species level are more accurate on longer reads (D'Amore *et al.*, 2016). The rate of evolution is also not consistent between these regions, so the taxonomic diversity can be

significantly impacted by the chosen region. The V3 – V6 regions have been shown to be most reliable, while the V2 and V8 are least reliable (Yang, Wang and Qian, 2016).

Compared to the well-established metabarcoding protocols developed for the bacterial species, the choice of primers for fungi are not as well established (Tiew *et al.*, 2020). The internal transcribed spacer (ITS) region of fungal rRNA operon spans 500-700bp, is commonly targeted for species identification of fungi, and has been proposed as a universal barcode (Tiew *et al.*, 2020). This marker gene has two target regions, and of them, the ITS2 region is preferred over the ITS1 for its less variable length (Ali *et al.*, 2019), and because there are more universal primer binding sites surrounding this region (Nilsson *et al.*, 2019). However, of the primers used for fungal metabarcoding in 2019, all of them will miss >10% of fungal groups (Nilsson *et al.*, 2019). This is because during PCR amplification, no primer can avid preferential binding (where the primer prefers binding to the target region in some species over others), so some microbial species including both bacteria and fungi will inevitably suffer over or under estimation (Pollock *et al.*, 2018). Lastly, a similar dilemma of inaccurate species level identification arises from using one of the short ITS regions as with one of the nine 16S hypervariable regions (Tiew *et al.*, 2020). Taxonomic resolution, and species assignments can also vary based on the reference database used (Ramakodi, 2022).

Some databases that have been developed for fungi include UNITE, INSDC, and Warcup ITS and for bacteria some major databases (that also often contain 18S or 16S sequenced from Eukarya) include SILVA, Ribosomal Database Project (RDP), and the Genome Taxonomy (GT) database. Each of these databases will differ in the total sequence content, and when classifying fungal species, the sparsity of reference sequences contained in databases can be particularly problematic (Tiew *et al.*, 2020). Both alpha and beta diversity may also be impacted by the reference database if negative controls are used to decontaminate sample data (Ramakodi, 2022). While there are currently no strict guidelines on how to select an appropriate database, popularity in the literature might be a deciding aspect, although the SILVA database has been shown to provide more ASVs with taxonomic information compared to RDP, GT database, and Consensus Taxonomy (Ramakodi, 2022). The GT databases however is unique in that its taxonomy is inferred from concatenated protein phylogeny, rather than a phylogenetic marker gene and it only includes full reference genomes and is thus free of chimeric sequences (Parks *et al.*, 2018)

Sequences annotated with taxonomic information can then also be used for predictive functional profiling, as understanding the functional diversity of the microbial community is a key goal of microbiome research. Two approaches to functional prediction are 1) inferring the unknown gene content from a phylogenetic tree of the 16S rRNA gene sequences using an extended ancestral state reconstruction algorithm as with

PICRUSt (Langille *et al.*, 2013), or 2) inferring the unknown gene content using a minimum 16S rRNA sequence similarity to identify the nearest neighbour as in Tax4Fun (Aßhauer *et al.*, 2015). Evolutionary modelling used by PICRUSt requires OTUs to be annotated using the Greengenes database, while Tax4Fun relies on annotations applied by the SILVA database which are transformed to a profile of KEGG organisms. Both methods rely on the relative abundance of taxa in each sample, corrected for multiple gene copy number, and use KEGG Orthologs (KOs) to identify gene families (Langille *et al.*, 2013; Aßhauer *et al.*, 2015). While each of these methods can provide an insight to the functional capabilities of the gut microbiome, the predictions cannot replace the detailed functional potential of the community that can be revealed through metagenomics (Abbasian *et al.*, 2015). Whole, genome sequencing is a more precise, yet more costly alternative to metabarcoding where the methods include fragmenting the total metagenome and then sequencing random pieces (Ranjan *et al.*, 2016). Metagenomics and metabarcoding both provide taxonomic information for the entire microbiome, however the full genome information gained from metagenomics, can improve taxonomic resolution, compared to annotations based solely on the 16S gene region (Ranjan *et al.*, 2016).

1.4.3 Combining metabolite and microbial analysis to improve our understanding of the microbiome

Comparisons of microbial composition and the potential metabolic capabilities identified with metagenomics alone will however miss important functional data that may be equally beneficial for interpreting gut microbiome community structure. Gut microbes carry numerous genes within their genome that can degrade a number of dietary components, however based on the available substrates and the microbial environment, different degradation pathways will be used by the same microbe. Different fermentation pathways will accordingly result in different metabolites being produced in the gut lumen (Verbeke *et al.*, 2015). A complement to sequencing is to also examine the metabolic activity of the microbiome, by either quantifying particular important metabolites, such as SCFA (McOrist *et al.*, 2008), or by profiling nearly all of the low molecular weight metabolites (metabolomics) (Karu *et al.*, 2018).

Methods for profiling faecal metabolites are less refined than those for other biofluids such as blood or urine, but rapid growth in this area has led to more standardised methods in recent times (Hosseinkhani *et al.*, 2021). Another consideration for faecal material is that the resulting profile is a snapshot of a dynamic environment, and will only reflect what is present in the excreted material. For example, 90 - 95% of the SCFA that are produced in the gut are absorbed by the host (Ziętek, Celewicz and Szczuko, 2021).The most commonly used technique for faecal metabolic profiling is gas chromatography mass spectrometry (GC-MS). This separation technique can resolve a

wide range of metabolite classes, and is highly sensitive, particularly for volatile compounds and organic acids (Karu *et al.*, 2018). In metabolomic studies, data processing and metabolite identification can proceed in two different ways. The targeted approach will process a predefined subset of the metabolome using an annotated reference database, and untargeted metabolomics attempts to analyse virtually all of the analytes in a sample (Gorrochategui *et al.*, 2016). Quantifying SCFA concentrations usually consists of the three most prominent acids acetate, butyrate, and propionate, but should also include the more scares iso-butyrate, and iso-valerate (Gray *et al.*, 2022). These methods provide a snapshot of either the primary microbially produced bioactive compounds (SCFA), or a comprehensive measure of the metabolic interactions between the host and the gut microbiota in the case of metabolomics. Combining metabolite data with microbial sequence data will result in a powerful approach that has the ability to not only characterise the perturbed microbiome, but also to reveal its metabolic response. Furthermore, in temporal analyses of the microbiome, combining these complementary methods will allow you to monitor functional changes that occur during disease progression, or in response to therapy such as prebiotics over time.

1.5 Gaps in the Literature

It is widely acknowledged that the health of the gut microbiome is necessary for the health of the individual, and with the ever more rapid changes coming as part of the post-modern lifestyle, finding ways to support and maintain microbiome health are increasingly more important. This is because modern lifestyles do not promote environmental and dietary exposures required for a healthy gut. However, as research linking the human microbiome and human health continues to increase, efforts to mitigate methodological biases lag behind, and additionally, the lack of detail in the methods section of most published work makes study reproducibility challenging (Mirzayi *et al.*, 2021). One difficulty is that study design – in terms of number of participants, and in the geographical area from which they are recruited – varies greatly among microbiome research, making it impossible to implement a single standard protocol for sample collection. However, the impact of collecting subsamples of faecal material, suboptimal sample homogenisation, and storage methods on microbiome and metabolite profiles has not be conclusively determined. Additionally, the amount of intraindividual variation between consecutive bowel movements has never been addressed, which could impact short-term, but more importantly, long-term microbiome studies. A second consideration is the considerable lack of positive and negative controls used in gut microbiome research (Hornung, Zwittink and Kuijper, 2019). Mock communities could be used to identify primer bias, or undetectable bacterial groups, as well as proportional changes to dominant reads across sequencing runs. Negative controls can also be use to assess and mitigate sample decontamination, however neither

of these controls are routinely used. Therefore, it is important that research highlighting their utility is put forward to push for the use of these controls, and so that they may be seen as necessary standard practice.

The microbiome is involved in many aspects of human health, but throughout this review I have highlighted in particular, evidence for the involvement of the gut microbiome in the evolution of both neural and immune system functioning. While it is unfortunate that non-communicable modern diseases and disorders are increasing in prevalence, molecular investigations of the gut microbiome are revealing how these diseases share in the potential for microbial-focused therapeutic interventions.

Treatments for autistic children including antibiotics (Sandler *et al.*, 2000), probiotics (Santocchi *et al.*, 2020) and faecal microbiome transplant (Kang *et al.*, 2019) have demonstrated that shifting the microbiome may lessen the severity of disease. The association between gut, microbiome, and brain health is clear, but it is not known at what stage in the aetiology of ASD the microbiome shifts may become observable. The microbiome goes through a rapid period of change in the first year of life, and during this important time, both the immune system and the brain are developing alongside the gut, with continual bidirectional communication taking place between the gut and distant organs. If assisting the microbiome community structure after disease manifestation reduces the severity of disease, then supporting the microbiome at the earliest stages of disease, may reduce the potential long-term outcomes of the disease. Therefore, it is crucial that we begin to understand when the earliest microbiome-shifts associated with a disease occur.

Identifying disease early allows for the use of early treatment, however there is also considerable research aimed at prevention of disease. Infants have benefited from formula supplemented with prebiotics, with improved gut (Bakker-Zierikzee *et al.*, 2005; Moro *et al.*, 2006; Scholtens *et al.*, 2008; Grüber *et al.*, 2010; Sierra *et al.*, 2015), and immune function (Scholtens *et al.*, 2008), and a reduction in the risk for developing atopic dermatitis (Moro *et al.*, 2006; Grüber *et al.*, 2010). Although, despite the consumer shift towards prebiotic infant formula, increasing numbers of children go on to develop allergies. It is still unclear what impact maternal diet has on the infant gut microbiome, although, the direct link between diet and microbiome composition, and maternal and infant health is well established. Due to the impact the gut microbiome has on overall health, it is not surprising that emerging evidence suggests that maternal gut health and therefore diet might impact the developing foetus and infant health (Gray *et al.*, 2017). The evidence for this potential link comes so far from animal models (Fujiwara *et al.*, 2010; Brosseau *et al.*, 2021). One of the primary advantages of supporting the microbiome with prebiotics is that they are backed by years of documented safety, and supporting the microbiome at conception would likely be the

most opportune time to promote healthy microbiome function. It may therefore be possible to safely assist in preventing the development of disease using prebiotic supplements, but so far this has not been investigated in humans.

1.6 Research Aims

The overarching aim of this project was to develop and utilise a rigorous and transparent method of investigating the faecal microbiome, employing it to investigate fine scale changes that occur early in the development of gut-associated disease.

1.6.1 Fecal sample collection methods and time of day impact microbiome composition and short chain fatty acid concentrations

It was hypothesised that sample collection methods including collecting small unhomogenised subsamples of stool, collecting stool at different times of day, or collecting stool into a DNA stabilisation tube would shift the proportions of the major bacterial phyla, and inflate intraindividual differences in microbiome composition and SCFA concentrations. Therefore, to compare the ability of common faecal sampling techniques to provide robust and stable microbiome and metabolite profiles (chapter 2) four objectives were formulated.

- Use variability in microbial abundance and SCFA concentrations to assess the effectiveness of different sampling techniques and time of day for collection.
- Identify any taxonomic groups that would be impacted by room temperature DNA preservation techniques.
- Use mock communities, negative controls and replicate sampling to capture and compare other levels of bias in the metabarcoding methodology.
- To provide both raw data and detailed methodology to the public to facilitate reproducibility of this study.

1.6.2 Changes to the gut microbiome in young children showing early behavioral signs of autism

It was hypothesised that the microbiome of young children who were showing early behavioural signs of ASD would host microbes that would differ significantly from neurotypically developing children. Furthermore, it was hypothesised that the differences in microbiome composition would be associated with differences in SCFA concentrations. Therefore, to characterise both the microbiome and short chain fatty acid concentrations of faecal samples from young children who were showing early behavioural signs of ASD (chapter 3), the following 3 objectives were formulated.

- Describe the microbiome characteristics of children with differing severity of early autism base on alpha-diversity and SCFA concentration.
- Determine bacterial and fungal taxa with significantly different abundance between children based on autism severity.
- Perform an exploratory analysis of the microbiome associated with pre-emptive behavioural treatment

1.6.3 Assessment of the temporal changes to the maternal and infant microbiome in response to a maternal prebiotic supplement during pregnancy and lactation

It was hypothesised that the maternal microbiome would be supported by prebiotic supplementation during pregnancy, causing significant differences in the composition of the microbiome and SCFA concentrations between those mothers on a prebiotic supplement and those receiving a placebo. As well, it was hypothesised that the healthy composition of the maternal microbiome would benefit the early colonisation and development of the infant microbiome. This would occur by foetal and infant exposure to beneficial bacteria (both during pregnancy and breastfeeding) and the metabolites they produce. These benefits were expected to result in a significant difference in the microbiome of infants whose mothers received the maternal prebiotic supplement and those who did not. To determine the effectiveness of a maternal prebiotic supplement on the maternal and infant microbiome and SCFA concentrations (chapter 4) the following 5 objectives were formulated.

- Identify confounding variables that may impact the microbiome structure during the study period
- Describe the microbiome and SCFA profiles of mothers and infants over the study period.
- Characterise differences in the microbial community and SCFA profiles between placebo and prebiotic group members – including both mothers and infants.
- Use prior knowledge of bacterial metabolism to describe microbial activity (SCFA concentrations) in the context of the microbiome structure and prebiotic supplement.
- Describe shifts of *Bifidobacterium* between the prebiotic and placebo groups

1.7 References

Abbasian, F. *et al.* (2015) 'The integration of sequencing and bioinformatics in metagenomics', *Reviews in Environmental Science and Biotechnology*. Springer Netherlands, 14(3), pp. 357–383. doi: 10.1007/s11157-015-9365-7.

Abellan-Schneyder, I. *et al.* (2021) 'Primer, Pipelines, Parameters: Issues in 16S rRNA Gene Sequencing', *mSphere*. American Society for Microbiology, 6(1). doi: 10.1128/msphere.01202-20.

Ali, N. A. B. M. *et al.* (2019) 'Optimisation and benchmarking of targeted amplicon sequencing for mycobiome analysis of respiratory specimens', *International Journal of Molecular Sciences*. MDPI AG, 20(20). doi: 10.3390/ijms20204991.

Allali, I. *et al.* (2017) 'A comparison of sequencing platforms and bioinformatics pipelines for compositional analysis of the gut microbiome', *BMC Microbiology*. BioMed Central, 17(1), p. 194. doi: 10.1186/s12866-017-1101-8.

Amon, P. and Sanderson, I. (2017) 'What is the microbiome?', *Archives of Disease in Childhood: Education and Practice Edition*, 102(5), pp. 258–261. doi: 10.1136/archdischild-2016-311643.

Andreo-Martínez, P. *et al.* (2021) 'A Meta-analysis of Gut Microbiota in Children with Autism', *Journal of Autism and Developmental Disorders*. Springer, 52(3), pp. 1374–1387. doi: 10.1007/s10803-021-05002-y.

Arvanitis, M. and Mylonakis, E. (2015) 'Fungal-bacterial interactions and their relevance in health', *Cellular Microbiology*. John Wiley & Sons, Ltd (10.1111), pp. 1442–1446. doi: 10.1111/cmi.12493.

Ashaolu, T. J., Ashaolu, J. O. and Adeyeye, S. A. O. (2021) 'Fermentation of prebiotics by human colonic microbiota in vitro and short-chain fatty acids production: a critical review', *Journal of Applied Microbiology*. John Wiley & Sons, Ltd, 130(3), pp. 677–687. doi: 10.1111/JAM.14843.

Asher, M. I. *et al.* (2021) 'Worldwide trends in the burden of asthma symptoms in school-aged children: Global Asthma Network Phase I cross-sectional study', *The Lancet*. Elsevier B.V., 398(10311), pp. 1569–1580. doi: 10.1016/S0140-6736(21)01450-1.

Aßhauer, K. P. *et al.* (2015) 'Tax4Fun: Predicting functional profiles from metagenomic 16S rRNA data', *Bioinformatics*. Oxford University Press, 31(17), pp. 2882–2884. doi: 10.1093/bioinformatics/btv287.

Au, K. F. *et al.* (2012) 'Improving PacBio Long Read Accuracy by Short Read Alignment', *PLOS ONE*. Public Library of Science, 7(10), p. e46679. doi: 10.1371/JOURNAL.PONE.0046679.

Bakker-Zierikzee, A. M. *et al.* (2005) 'Effects of infant formula containing a mixture of galacto- and fructo-oligosaccharides or viable Bifidobacterium animalis on the intestinal microflora during the first 4 months of life', *British Journal of Nutrition*. Cambridge University Press, 94(5), pp. 783–790. doi: 10.1079/BJN20051451.

Bakker, M. G. (2018) 'A fungal mock community control for amplicon sequencing experiments', *Molecular Ecology Resources*. John Wiley & Sons, Ltd (10.1111), 18(3), pp. 541–556. doi: 10.1111/1755-0998.12760.

Ball, P. (2012) 'Bacteria replicate close to the physical limit of efficiency ', *Nature*. Springer Science and Business Media LLC. doi: 10.1038/NATURE.2012.11446.

Bastiaanssen, T. F. S. *et al.* (2021) 'Volatility as a Concept to Understand the Impact of Stress on the Microbiome', *Psychoneuroendocrinology*. Elsevier Ltd, 124. doi: 10.1016/j.psyneuen.2020.105047.

Bazanella, M. *et al.* (2017) 'Randomized controlled trial on the impact of early-life intervention with bifidobacteria on the healthy infant fecal microbiota and metabolome', *American Journal of Clinical Nutrition*. Oxford University Press, 106(5), pp. 1274–1286. doi: 10.3945/ajcn.117.157529.

Beaumont, M. *et al.* (2017) 'Quantity and source of dietary protein influence metabolite production by gut microbiota and rectal mucosa gene expression: A randomized, parallel, double-blind trial in overweight humans', *American Journal of Clinical Nutrition*, 106(4), pp. 1005–1019. doi: 10.3945/ajcn.117.158816.

Bellinger, S. V. *et al.* (2019) 'Evaluating differences in prevalence of food allergies between two geographic regions: Australia and US', *Journal of Allergy and Clinical Immunology*. Elsevier, 143(2), p. AB268. doi: 10.1016/J.JACI.2018.12.818.

Bengtsson-Palme, J. (2020) 'Microbial model communities: To understand complexity, harness the power of simplicity', *Computational and Structural Biotechnology Journal*. Elsevier, pp. 3987–4001. doi: 10.1016/j.csbj.2020.11.043.

den Besten, G. *et al.* (2013) 'Gut-derived short-chain fatty acids are vividly assimilated into host carbohydrates and lipids', *American Journal of Physiology - Gastrointestinal and Liver Physiology*. American Physiological Society Bethesda, MD, 305(12), pp. 900–910. doi: 10.1152/AJPGI.00265.2013/ASSET/IMAGES/LARGE/ZH30241365670006.JPEG.

Den Besten, G. *et al.* (2013) 'The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism', *Journal of Lipid Research*. American Society for Biochemistry and Molecular Biology, pp. 2325–2340. doi: 10.1194/jlr.R036012.

Bonnet, M. H. and Arand, D. L. (2010) 'Hyperarousal and insomnia: State of the science', *Sleep Medicine Reviews*. W.B. Saunders, pp. 9–15. doi: 10.1016/j.smrv.2009.05.002.

Braman, S. S. (2006) 'The Global Burden of Asthma', *Chest*. Elsevier, 130(1), pp. 4S-12S. doi: 10.1378/CHEST.130.1_SUPPL.4S.

Braniste, V. *et al.* (2014) 'The gut microbiota influences blood-brain barrier permeability in mice', *Science Translational Medicine*. Sci Transl Med, 6(263). doi: 10.1126/scitranslmed.3009759.

Bravi, F. *et al.* (2016) 'Impact of maternal nutrition on breast-milk composition: a systematic review', *The American Journal of Clinical Nutrition*. Oxford Academic, 104(3), pp. 646–662. doi: 10.3945/AJCN.115.120881.

Brinkworth, G. D. *et al.* (2009) 'Comparative effects of very low-carbohydrate, high-fat and high-carbohydrate, low-fat weight-loss diets on bowel habit and faecal short-chain fatty acids and bacterial populations', *The British journal of nutrition*. Br J Nutr, 101(10), pp. 1493–1502. doi: 10.1017/S0007114508094658.

Brosseau, C. *et al.* (2021) 'Prebiotic Supplementation During Pregnancy Modifies the Gut Microbiota and Increases Metabolites in Amniotic Fluid, Driving a Tolerogenic Environment In Utero', *Frontiers in Immunology*. Frontiers Media S.A., 12, p. 2857. doi: 10.3389/fimmu.2021.712614.

Buffington, S. A. *et al.* (2016) 'Microbial Reconstitution Reverses Maternal Diet-Induced Social and Synaptic Deficits in Offspring', *Cell*. Elsevier Inc., 165(7), pp. 1762–1775. doi: 10.1016/j.cell.2016.06.001.

Bukin, Y. S. *et al.* (2019) 'The effect of 16s rRNA region choice on bacterial community metabarcoding results', *Scientific Data*. Nature Publishing Groups, 6(1), pp. 1–14. doi: 10.1038/sdata.2019.7.

Byrd, D. A. *et al.* (2019) 'Reproducibility, stability, and accuracy of microbial profiles by fecal sample collection method in three distinct populations', *PLoS ONE*. Public Library of Science, 14(11), p. e0224757. doi: 10.1371/journal.pone.0224757.

Cacho, N. T. and Lawrence, R. M. (2017) 'Innate Immunity and Breast Milk.', *Frontiers in immunology*. Frontiers Media SA, 8, p. 584. doi: 10.3389/fimmu.2017.00584.

Cai, R. *et al.* (2020) 'Interactions of commensal and pathogenic microorganisms with the mucus layer in the colon', *Gut Microbes*. Taylor & Francis, 11(4), p. 680. doi: 10.1080/19490976.2020.1735606.

Callahan, B. J. *et al.* (2016) 'DADA2: High-resolution sample inference from Illumina amplicon data', *Nature Methods*. Nature Publishing Group, 13(7), pp. 581–583. doi: 10.1038/nmeth.3869.

Cao, Y. *et al.* (2020) 'Impact of food additives on the composition and function of gut microbiota: A review', *Trends in Food Science & Technology*. Elsevier, 99, pp. 295–310. doi: 10.1016/J.TIFS.2020.03.006.

Carmody, R. N. *et al.* (2015) 'Diet dominates host genotype in shaping the murine gut microbiota', *Cell Host and Microbe*. Elsevier, 17(1), pp. 72–84. doi: 10.1016/j.chom.2014.11.010.

Carrothers, J. M. *et al.* (2015) 'Fecal Microbial Community Structure Is Stable over Time and Related to Variation in Macronutrient and Micronutrient Intakes in Lactating Women', *Journal of Nutrition*. Oxford University Press, 145(10), pp. 2379–2388. doi: 10.3945/jn.115.211110.

Carruthers, L. V. *et al.* (2019) 'The impact of storage conditions on human stool 16S rRNA microbiome composition and diversity', *PeerJ*. PeerJ Inc., 2019(12). doi: 10.7717/peerj.8133.

Casaburi, G. *et al.* (2021) 'Metagenomic insights of the infant microbiome community structure and function across multiple sites in the United States', *Scientific Reports*. Nature Publishing Group, 11(1), pp. 1–12. doi: 10.1038/s41598-020-80583-9.

Chen, N. *et al.* (2022) 'A Meta-Analysis of Probiotics for the Treatment of Allergic Airway Diseases in Children and Adolescents', *American Journal of Rhinology and Allergy*. SAGE Publications Inc., 36(4), pp. 480–490. doi: 10.1177/19458924221080159.

Chong, C., Bloomfield, F. and O'Sullivan, J. (2018) 'Factors Affecting Gastrointestinal Microbiome Development in Neonates', *Nutrients*. Multidisciplinary Digital Publishing Institute (MDPI), 10(3), p. 274. doi: 10.3390/nu10030274.

Claassen-Weitz, S. *et al.* (2020) 'Optimizing 16S rRNA gene profile analysis from low biomass nasopharyngeal and induced sputum specimens', *BMC Microbiology*. BMC Microbiology, 20(113), pp. 1–26. doi: 10.1186/s12866-020-01795-7.

Clemente, J. C. *et al.* (2015) 'The microbiome of uncontacted Amerindians', *Science Advances*. American Association for the Advancement of Science, 1(3). doi: 10.1126/SCIADV.1500183/SUPPL_FILE/TABLE_S9.XLSX.

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

Columpsi, P. *et al.* (2016) 'Beyond the gut bacterial microbiota: The gut virome', *Journal of Medical Virology*. Wiley-Blackwell, pp. 1467–1472. doi: 10.1002/jmv.24508.

Cox, J. *et al.* (2010) 'A maternal high-fat diet is accompanied by alterations in the fetal primate metabolome', *Am J Obstet Gynecol*, 201(3), pp. 1–15. doi: 10.1016/j.ajog.2009.06.041.A.

Coyte, K. Z. and Rakoff-Nahoum, S. (2019) 'Understanding Competition and Cooperation within the Mammalian Gut Microbiome', *Current Biology*. Cell Press,

29(11), pp. R538–R544. doi: 10.1016/J.CUB.2019.04.017.

Cronin, P. *et al.* (2021) 'Dietary fibre modulates the gut microbiota', *Nutrients*. Multidisciplinary Digital Publishing Institute, p. 1655. doi: 10.3390/nu13051655.

Cryan, J. F. *et al.* (2020) 'The gut microbiome in neurological disorders', *The Lancet Neurology*. Lancet Publishing Group, pp. 179–194. doi: 10.1016/S1474-4422(19)30356-4.

Cryan, J. F. and O'Mahony, S. M. (2011) 'The microbiome-gut-brain axis: from bowel to behavior', *Neurogastroenterology & Motility*. Wiley/Blackwell (10.1111), 23(3), pp. 187–192. doi: 10.1111/j.1365-2982.2010.01664.x.

Cunningha, J. L. *et al.* (2020) 'Impact of time and temperature on gut microbiota and SCFA composition in stool samples', *PLoS ONE*. Edited by J. J. Loor. Public Library of Science, 15(8 August), p. e0236944. doi: 10.1371/journal.pone.0236944.

Cunningham, M. *et al.* (2021) 'Shaping the Future of Probiotics and Prebiotics', *Trends in Microbiology*. Elsevier Current Trends, 29(8), pp. 667–685. doi: 10.1016/J.TIM.2021.01.003.

D'Amore, R. *et al.* (2016) 'A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling', *BMC Genomics*. BioMed Central, 17(1), p. 55. doi: 10.1186/s12864-015-2194-9.

David, L. A. *et al.* (2014) 'Diet rapidly and reproducibly alters the human gut microbiome', *Nature*. Nature Publishing Group, 505(7484), pp. 559–563. doi: 10.1038/nature12820.

Delhaes, F. *et al.* (2018) 'Altered maternal and placental lipid metabolism and fetal fat development in obesity: Current knowledge and advances in non-invasive assessment', *Placenta*. W.B. Saunders, 69, pp. 118–124. doi: 10.1016/J.PLACENTA.2018.05.011.

Dharmage, S. C., Perret, J. L. and Custovic, A. (2019) 'Epidemiology of asthma in children and adults', *Frontiers in Pediatrics*. Frontiers Media S.A., p. 246. doi: 10.3389/fped.2019.00246.

Ding, H. T., Taur, Y. and Walkup, J. T. (2017) 'Gut Microbiota and Autism: Key Concepts and Findings', *Journal of Autism and Developmental Disorders*. Springer US, 47(2), pp. 480–489. doi: 10.1007/s10803-016-2960-9.

Dominguez-Bello, M. G. *et al.* (2016) 'Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer', *Nature Medicine*. NIH Public Access, 22(3), pp. 250–253. doi: 10.1038/nm.4039.

Donaldson, G. P., Lee, S. M. and Mazmanian, S. K. (2015) 'Gut biogeography of the bacterial microbiota', *Nature Reviews Microbiology*. NIH Public Access, pp. 20–32. doi: 10.1038/nrmicro3552.

Dunn, A. B. *et al.* (2017) 'The Maternal Infant Microbiome: Considerations for Labor and Birth', *MCN The American Journal of Maternal/Child Nursing*. Lippincott Williams and Wilkins, 42(6), pp. 318–325. doi: 10.1097/NMC.0000000000000373.

Dzidic, M. *et al.* (2018) 'Gut Microbiota and Mucosal Immunity in the Neonate', *Medical Sciences*. Multidisciplinary Digital Publishing Institute, 6(3), p. 56. doi: 10.3390/medsci6030056.

Eckburg, P. B. *et al.* (2005) 'Microbiology: Diversity of the human intestinal microbial flora', *Science*. NIH Public Access, 308(5728), pp. 1635–1638. doi: 10.1126/science.1110591.

Ege, M. J. (2017) 'The hygiene hypothesis in the age of the microbiome', *Annals of the American Thoracic Society*. American Thoracic Society, 14, pp. S348–S353. doi: 10.1513/AnnalsATS.201702-139AW.

Elvers, K. T. *et al.* (2020) 'Antibiotic-induced changes in the human gut microbiota for the most commonly prescribed antibiotics in primary care in the UK: a systematic review', *BMJ open*. British Medical Journal Publishing Group, 10(9), p. e035677. doi: 10.1136/bmjopen-2019-035677.

Ericsson, A. C., Davis, J. W., Spollen, W., Bivens, N., Givan, S., Hagan, C. E., McIntosh, M., & Franklin, C. L. (2015) 'Effects of Vendor and Genetic Background on the Composition of the Fecal Microbiota of Inbred Mice', PLOS ONE, 10(2), e0116704. https://doi.org/10.1371/JOURNAL.PONE.0116704

Exley, C., & Clarkson, E. (2020) 'Aluminium in human brain tissue from donors without neurodegenerative disease: A comparison with Alzheimer's disease, multiple sclerosis and autism', Scientific Reports, 10(1), pp. 1–7. https://doi.org/10.1038/s41598-020-64734-6

Fan, D. *et al.* (2015) 'Activation of HIF-1α and LL-37 by commensal bacteria inhibits Candida albicans colonization', *Nature Medicine*. Nature Publishing Group, pp. 808–814. doi: 10.1038/nm.3871.

Fardini, Y. *et al.* (2011) 'Fusobacterium nucleatum adhesin FadA binds vascular endothelial cadherin and alters endothelial integrity', *Molecular Microbiology*. John Wiley & Sons, Ltd, 82(6), pp. 1468–1480. doi: 10.1111/J.1365-2958.2011.07905.X.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). 'G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences', *Behavior Research Methods*, *39*, pp. 175-191.

Faria, A. M. C., Reis, B. S. and Mucida, D. (2017) 'Tissue adaptation: Implications for gut immunity and tolerance', *Journal of Experimental Medicine*. The Rockefeller University Press, pp. 1211–1226. doi: 10.1084/jem.20162014.

Feng, R., Fan, Y., Chen, L., Ge, Q., Xu, J., Yang, M., & Chen, K. (2022) 'Based on 16 S rRNA sequencing and metabonomics to reveal the new mechanism of aluminum potassium sulfate induced inflammation and abnormal lipid metabolism in mice', *Ecotoxicology and Environmental Safety*, *247*, 114214. https://doi.org/10.1016/j.ecoenv.2022.114214

Fiorentino, M. *et al.* (2016) 'Blood–brain barrier and intestinal epithelial barrier alterations in autism spectrum disorders', *Molecular Autism 2016 7:1*. BioMed Central, 7(1), pp. 1–17. doi: 10.1186/S13229-016-0110-Z.

Fox, E. P. *et al.* (2014) 'Anaerobic bacteria grow within candida albicans biofilms and induce biofilm formation in suspension cultures', *Current Biology*. Cell Press, 24(20), pp. 2411–2416. doi: 10.1016/j.cub.2014.08.057.

Francino, M. P. (2016) 'Antibiotics and the human gut microbiome: Dysbioses and accumulation of resistances', *Frontiers in Microbiology*. Frontiers Media S.A., 6(JAN), p. 1543. doi: 10.3389/FMICB.2015.01543/BIBTEX.

Friedman, N. J. and Zeiger, R. S. (2005) 'The role of breast-feeding in the development of allergies and asthma', *Journal of Allergy and Clinical Immunology*, 115(6), pp. 1238–1248. doi: 10.1016/j.jaci.2005.01.069.

Fu, B. C. *et al.* (2019) 'Temporal variability and stability of the fecal microbiome: The multiethnic cohort study', *Cancer Epidemiology Biomarkers and Prevention*. American Association for Cancer Research Inc., 28(1), pp. 154–162. doi: 10.1158/1055-9965.EPI-18-0348.

Fujimura, K. E. *et al.* (2016) 'Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation', *Nature Medicine*. Nature Publishing Group, 22(10), pp. 1187–1191. doi: 10.1038/nm.4176.

Fujiwara, R. *et al.* (2010) 'Maternal consumption of fructo-oligosaccharide diminishes the severity of skin inflammation in offspring of NC/Nga mice', *British Journal of Nutrition*. Cambridge University Press, 103(4), pp. 530–538. doi: 10.1017/S000711450999198X.

Fung, T. C., Olson, C. A. and Hsiao, E. Y. (2017) 'Interactions between the microbiota, immune and nervous systems in health and disease', *Nature Neuroscience 2016 20:2*. Nature Publishing Group, 20(2), pp. 145–155. doi: 10.1038/nn.4476.

Galazzo, G. *et al.* (2020) 'Development of the Microbiota and Associations With Birth Mode, Diet, and Atopic Disorders in a Longitudinal Analysis of Stool Samples, Collected From Infancy Through Early Childhood', *Gastroenterology*. W.B. Saunders, 158(6), pp. 1584–1596. doi: 10.1053/J.GASTRO.2020.01.024.

Gibson, G. R. *et al.* (2017) 'Expert consensus document: The International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of prebiotics', *Nature Reviews Gastroenterology and Hepatology*. Nature

Publishing Group, pp. 491–502. doi: 10.1038/nrgastro.2017.75.

Gilbert, J. A. *et al.* (2018) 'Current understanding of the human microbiome', *Nature Medicine*. Nature Publishing Group, 24(4), pp. 392–400. doi: 10.1038/nm.4517.

Giri, R. and Sharma, R. K. (2022) 'Psychobiotics in diet: significance and applications of neuroactive and psychoactive microbial metabolites', *Nutrition Reviews*. Oxford University Press (OUP). doi: 10.1093/NUTRIT/NUAC019.

Gohl, D. M., Vangay, P., Garbe, J., MacLean, A., Hauge, A., Becker, A., Gould, T. J., Clayton, J. B., Johnson, T. J., Hunter, R., Knights, D., & Beckman, K. B. (2016) 'Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies', Nature Biotechnology 34:9, 34(9), pp. 942–949. https://doi.org/10.1038/nbt.3601

Góra, B. *et al.* (2018) 'Toxin profile of fecal Clostridium perfringens strains isolated from children with autism spectrum disorders', *Anaerobe*. Academic Press, 51, pp. 73–77. doi: 10.1016/j.anaerobe.2018.03.005.

Gorrochategui, E. *et al.* (2016) 'Data analysis strategies for targeted and untargeted LC-MS metabolomic studies: Overview and workflow', *TrAC - Trends in Analytical Chemistry*. Elsevier, pp. 425–442. doi: 10.1016/j.trac.2016.07.004.

Gorzelak, M. A. *et al.* (2015) 'Methods for improving human gut microbiome data by reducing variability through sample processing and storage of stool', *PLoS ONE*, 10(8). doi: 10.1371/journal.pone.0134802.

Gray, J. *et al.* (2022) 'A fast, fully validated GC-MS method using a simplified pretreatment for the quantification of short and branched chain fatty acids in human stool', *Journal of Mass Spectrometry*. John Wiley & Sons, Ltd, 57(4), p. e4817. doi: 10.1002/JMS.4817.

Gray, L. E. K. *et al.* (2017) 'The maternal diet, gut bacteria, and bacterial metabolites during pregnancy influence offspring asthma', *Frontiers in Immunology*. Frontiers Media SA. doi: 10.3389/fimmu.2017.00365.

Grimaldi, R. *et al.* (2018) 'A prebiotic intervention study in children with autism spectrum disorders (ASDs)', *Microbiome*. BioMed Central Ltd., 6(133). doi: 10.1186/s40168-018-0523-3.

Grüber, C. *et al.* (2010) 'Reduced occurrence of early atopic dermatitis because of immunoactive prebiotics among low-atopy-risk infants', *Journal of Allergy and Clinical Immunology*, 126(4), pp. 791–797. doi: 10.1016/j.jaci.2010.07.022.

Gubert, C. *et al.* (2020) 'Exercise, diet and stress as modulators of gut microbiota: Implications for neurodegenerative diseases', *Neurobiology of Disease*. Academic Press Inc. doi: 10.1016/j.nbd.2019.104621.

Gupta, V. *et al.* (2019) 'Reconciling Hygiene and Cleanliness: A New Perspective from Human Microbiome', *Indian Journal of Microbiology 2019 60:1*. Springer, 60(1), pp. 37–44. doi: 10.1007/S12088-019-00839-5.

Hallen-Adams, H. E. *et al.* (2015) 'Fungi inhabiting the healthy human gastrointestinal tract: A diverse and dynamic community', *Fungal Ecology*. Elsevier, 15, pp. 9–17. doi: 10.1016/j.funeco.2015.01.006.

Hallen-Adams, H. E. and Suhr, M. J. (2017) 'Fungi in the healthy human gastrointestinal tract', *Virulence*. Taylor & Francis, 8(3), pp. 352–358. doi: 10.1080/21505594.2016.1247140.

He, J. *et al.* (2020) 'Short-chain fatty acids and their association with signalling pathways in inflammation, glucose and lipid metabolism', *International Journal of Molecular Sciences*. Multidisciplinary Digital Publishing Institute, pp. 1–16. doi: 10.3390/ijms21176356.

van der Hee, B. and Wells, J. M. (2021) 'Microbial Regulation of Host Physiology by Short-chain Fatty Acids', *Trends in Microbiology*. Elsevier Current Trends, 29(8), pp. 700–712. doi: 10.1016/J.TIM.2021.02.001.

Heiss, C. N. and Olofsson, L. E. (2019) 'The role of the gut microbiota in development, function and disorders of the central nervous system and the enteric nervous system', *Journal of Neuroendocrinology*. John Wiley & Sons, Ltd, p. e12684. doi: 10.1111/jne.12684.

Hill, C. J. *et al.* (2017) 'Evolution of gut microbiota composition from birth to 24 weeks in the INFANTMET Cohort', *Microbiome*. BioMed Central, 5(4). doi: 10.1186/s40168-016-0213-y.

Hirata, S. ichiro and Kunisawa, J. (2017) 'Gut microbiome, metabolome, and allergic diseases', *Allergology International*. Elsevier, pp. 523–528. doi: 10.1016/j.alit.2017.06.008.

Holt, P. G. and Jones, C. A. (2000) 'The development of the immune system during pregnancy and early life', *Allergy*, 55(8), pp. 688–697. doi: 10.1034/j.1398-9995.2000.00118.x.

Hornung, B. V. H., Zwittink, R. D. and Kuijper, E. J. (2019) 'Issues and current standards of controls in microbiome research', *FEMS Microbiology Ecology*. Oxford University Press. doi: 10.1093/femsec/fiz045.

Hosseinkhani, F. *et al.* (2021) 'Towards standards for human fecal sample preparation in targeted and untargeted lc-hrms studies', *Metabolites*. MDPI AG, 11(6), p. 364. doi: 10.3390/METABO11060364/S1.

Hugenholtz, F., & de Vos, W. M. (2018) 'Mouse models for human intestinal microbiota research: a critical evaluation', Cellular and Molecular Life Sciences, 75(1),

pp. 149–160. https://doi.org/10.1007/S00018-017-2693-8/FIGURES/3

Hung, C. C. *et al.* (2022) 'Gut microbiota in patients with Alzheimer's disease spectrum: a systematic review and meta-analysis', *Aging*. Impact Journals, LLC, 14(1), pp. 477–496. doi: 10.18632/aging.203826.

Iruzubieta, P. *et al.* (2022) 'Supporting Evidence of Human Enteric Nervous System Adult Neurogenesis: Presence of Primary Cilia and Adult Neurogenesis Markers', *Cellular and Molecular Neurobiology*. Springer, 42(2), pp. 473–481. doi: 10.1007/s10571-020-01017-8.

Jernberg, C. *et al.* (2010) 'Long-term impacts of antibiotic exposure on the human intestinal microbiota', *Microbiology*. Microbiology Society, 156(11), pp. 3216–3223. doi: 10.1099/MIC.0.040618-0/CITE/REFWORKS.

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

Jin Song, S. *et al.* (2019) 'Engineering the microbiome for animal health and conservation', *Experimental Biology and Medicine*. SAGE Publications Inc., 244(6), pp. 494–504. doi: 10.1177/1535370219830075.

Johnson, A. J. *et al.* (2019) 'Daily Sampling Reveals Personalized Diet-Microbiome Associations in Humans', *Cell Host and Microbe*. Cell Press, 25(6), pp. 789-802.e5. doi: 10.1016/j.chom.2019.05.005.

Jones, J. *et al.* (2021) 'Fecal sample collection methods and time of day impact microbiome composition and short chain fatty acid concentrations', *Scientific Reports*. Nature Publishing Group, 11(1), pp. 1–13. doi: 10.1038/s41598-021-93031-z.

Kang, D. W. *et al.* (2019) 'Long-term benefit of Microbiota Transfer Therapy on autism symptoms and gut microbiota', *Scientific Reports*. Nature Publishing Group, 9(1), pp. 1–9. doi: 10.1038/s41598-019-42183-0.

Karl, J. P. *et al.* (2017) 'Changes in intestinal microbiota composition and metabolism coincide with increased intestinal permeability in young adults under prolonged physiological stress', *American Journal of Physiology - Gastrointestinal and Liver Physiology*. Am J Physiol Gastrointest Liver Physiol, 312(6), pp. G559–G571. doi: 10.1152/ajpgi.00066.2017.

Karu, N. *et al.* (2018) 'A review on human fecal metabolomics: Methods, applications and the human fecal metabolome database', *Analytica Chimica Acta*. Elsevier, pp. 1–24. doi: 10.1016/j.aca.2018.05.031.

Kawahara, M. (2005) 'Effects of aluminum on the nervous system and its possible link with neurodegenerative diseases', *Journal of Alzheimer's Disease*, *8*(2), pp. 171–182. https://doi.org/10.3233/JAD-2005-8210

Kim, D. *et al.* (2017) 'Optimizing methods and dodging pitfalls in microbiome research', *Microbiome*. doi: 10.1186/s40168-017-0267-5.

Kim, J. Y. *et al.* (2020) 'The human gut archaeome: Identification of diverse haloarchaea in Korean subjects', *Microbiome*. BioMed Central, 8(1), pp. 1–17. doi: 10.1186/s40168-020-00894-x.

Klimenko, N. S. *et al.* (2018) 'Microbiome Responses to an Uncontrolled Short-Term Diet Intervention in the Frame of the Citizen Science Project', *Nutrients*. MDPI AG, 10(5), p. 576. doi: 10.3390/nu10050576.

Klopp, J. *et al*. (2022) 'Meconium Microbiome of Very Preterm Infants across Germany', *MSphere*, *7*(1). https://doi.org/10.1128/MSPHERE.00808-21/SUPPL_FILE/MSPHERE.00808-21-S0005.TIF

Koren, O. *et al.* (2012) 'Host Remodeling of the Gut Microbiome and Metabolic Changes during Pregnancy', *Cell*. Cell Press, 150(3), pp. 470–480. doi: 10.1016/J.CELL.2012.07.008.

Koropatkin, N. M., Cameron, E. A. and Martens, E. C. (2012) 'How glycan metabolism shapes the human gut microbiota', *Nature reviews. Microbiology*. NIH Public Access, 10(5), p. 323. doi: 10.1038/NRMICRO2746.

Koskella, B., Hall, L. J. and Metcalf, C. J. E. (2017) 'The microbiome beyond the horizon of ecological and evolutionary theory', *Nature Ecology and Evolution*. Nature Publishing Group, 1(11), pp. 1606–1615. doi: 10.1038/s41559-017-0340-2.

Kovtun, A. S. *et al.* (2020) 'Antibiotic Resistance Genes in the Gut Microbiota of Children with Autistic Spectrum Disorder as Possible Predictors of the Disease', *Microbial Drug Resistance*. Mary Ann Liebert Inc., 26(11), pp. 1307–1320. doi: 10.1089/mdr.2019.0325.

Langille, M. *et al.* (2013) 'Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences.', *Nature biotechnology*. Nature Publishing Group, 31(9), pp. 814–21. doi: 10.1038/nbt.2676.

Li, B. *et al.* (2018) 'The microbiome and autoimmunity: a paradigm from the gut–liver axis', *Cellular and Molecular Immunology*. Nature Publishing Group, pp. 595–609. doi: 10.1038/cmi.2018.7.

Liang, Y. *et al.* (2020) 'Systematic Analysis of Impact of Sampling Regions and Storage Methods on Fecal Gut Microbiome and Metabolome Profiles', *mSphere*. American Society for Microbiology, 5(1). doi: 10.1128/msphere.00763-19.

Liu, F., Li, P., Chen, M., Luo, Y., Prabhakar, M., Zheng, H., He, Y., Qi, Q., Long, H., Zhang, Y., Sheng, H., & Zhou, H. (2017) 'Fructooligosaccharide (FOS) and Galactooligosaccharide (GOS) Increase Bifidobacterium but Reduce Butyrate Producing Bacteria with Adverse Glycemic Metabolism in healthy young population', *Scientific*

Reports. 7(1), pp. 1–12. https://doi.org/10.1038/s41598-017-10722-2

Loh, W. and Tang, M. L. K. (2018) 'The Epidemiology of Food Allergy in the Global Context', *International Journal of Environmental Research and Public Health 2018, Vol. 15, Page 2043*. Multidisciplinary Digital Publishing Institute, 15(9), p. 2043. doi: 10.3390/IJERPH15092043.

Louca, S., Doebeli, M. and Parfrey, L. W. (2018) 'Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem', *Microbiome*. BioMed Central Ltd., 6(1), pp. 1–12. doi: 10.1186/S40168-018-0420-9/FIGURES/4.

Lukiw, W. J. (2020) 'Gastrointestinal (GI) Tract Microbiome-Derived Neurotoxins—Potent Neuro-Inflammatory Signals From the GI Tract via the Systemic Circulation Into the Brain', *Frontiers in Cellular and Infection Microbiology*. Frontiers Media S.A., 10, p. 22. doi: 10.3389/fcimb.2020.00022.

Lynch, S. V (2016) 'Gut microbiota and allergic disease: New insights', in *Annals of the American Thoracic Society*. American Thoracic Society, pp. S51–S54. doi: 10.1513/AnnalsATS.201507-451MG.

Macfarlane, G. T., Steed, H. and Macfarlane, S. (2008) 'Bacterial metabolism and health-related effects of galacto-oligosaccharides and other prebiotics', *Journal of Applied Microbiology*, 104(2), pp. 305–344. doi: 10.1111/j.1365-2672.2007.03520.x.

Maniam, J. and Morris, M. J. (2012) 'The link between stress and feeding behaviour', *Neuropharmacology*. Pergamon, pp. 97–110. doi: 10.1016/j.neuropharm.2012.04.017.

Martin, C. R. *et al.* (2018) 'The Brain-Gut-Microbiome Axis', *CMGH*. Elsevier, pp. 133–148. doi: 10.1016/j.jcmgh.2018.04.003.

Martin, R. *et al.* (2016) 'Early-Life events, including mode of delivery and type of feeding, siblings and gender, shape the developing gut microbiota', PLoS ONE, 11(6). https://doi.org/10.1371/journal.pone.0158498

Matamoros, S. *et al.* (2013) 'Development of intestinal microbiota in infants and its impact on health', *Trends in Microbiology*, 21(4), pp. 167–173. doi: 10.1016/j.tim.2012.12.001.

McDonnell, L. *et al.* (2021) 'Association between antibiotics and gut microbiome dysbiosis in children: systematic review and meta-analysis', *Gut Microbes*. Gut Microbes, 13(1), pp. 1–18. doi: 10.1080/19490976.2020.1870402.

McIlroy, J. *et al.* (2018) 'Review article: the gut microbiome in inflammatory bowel disease—avenues for microbial management', *Alimentary Pharmacology and Therapeutics*. Wiley/Blackwell (10.1111), pp. 26–42. doi: 10.1111/apt.14384.

McOrist, A. L. *et al.* (2008) 'Bacterial population dynamics and faecal short-chain fatty acid (SCFA) concentrations in healthy humans', *British Journal of Nutrition*, 100(1), pp.

138–146. doi: 10.1017/S0007114507886351.

Meyer, K. *et al.* (2016) 'Maternal diet alters the breast milk microbiome and microbial gene content', *American Journal of Obstetrics and Gynecology*. Mosby, pp. S47–S48. doi: 10.1016/j.ajog.2015.10.084.

Mirzayi, C. *et al.* (2021) 'Reporting guidelines for human microbiome research: the STORMS checklist', *Nature Medicine 2021 27:11*. Nature Publishing Group, 27(11), pp. 1885–1892. doi: 10.1038/s41591-021-01552-x.

Mold, M., Umar, D., King, A., & Exley, C. (2018) 'Aluminium in brain tissue in autism', *Journal of Trace Elements in Medicine and Biology*, *46*, 76–82. https://doi.org/10.1016/J.JTEMB.2017.11.012

Moles, L. and Otaegui, D. (2020) 'The impact of diet on microbiota evolution and human health. Is diet an adequate tool for microbiota modulation?', *Nutrients*. MDPI AG. doi: 10.3390/nu12061654.

Moro, G. *et al.* (2006) 'A mixture of prebiotic oligosaccharides reduces the incidence of atopic dermatitis during the first six months of age.', *Archives of disease in childhood*. BMJ Publishing Group, 91(10), pp. 814–9. doi: 10.1136/adc.2006.098251.

Murray, D. C., Coghlan, M. L. and Bunce, M. (2015) 'From Benchtop to Desktop: Important Considerations when Designing Amplicon Sequencing Workflows', *PLOS ONE*. Edited by T. Gilbert. Public Library of Science, 10(4), p. e0124671. doi: 10.1371/journal.pone.0124671.

Nauta, A. J. and Garssen, J. (2013) 'Evidence-based benefits of specific mixtures of non-digestible oligosaccharides on the immune system', *Carbohydrate Polymers*. Elsevier Ltd., 93(1), pp. 263–265. doi: 10.1016/j.carbpol.2012.02.021.

Nilsson, R. H. *et al.* (2019) 'Mycobiome diversity: high-throughput sequencing and identification of fungi', *Nature Reviews Microbiology*. Nature Publishing Group, pp. 95–109. doi: 10.1038/s41579-018-0116-y.

Nkamga, V. D., Henrissat, B. and Drancourt, M. (2017) 'Archaea: Essential inhabitants of the human digestive microbiota', *Human Microbiome Journal*. Elsevier, 3, pp. 1–8. doi: 10.1016/J.HUMIC.2016.11.005.

Nutricionist, A. C. T., Procianoy, R. S., Roesch, L. F. W., Corso, A. L., Dobbler, P. T., & Silveira, R. C. (2020) 'Meconium microbiome and its relation to neonatal growth and head circumference catch-up in preterm infants', PLOS ONE, 15(9), e0238632. https://doi.org/10.1371/JOURNAL.PONE.0238632

Ose, R. *et al.* (2018) 'The ability of human intestinal anaerobes to metabolize different oligosaccharides: Novel means for microbiota modulation?', *Anaerobe*. Academic Press, 51, pp. 110–119. doi: 10.1016/j.anaerobe.2018.04.018.

Pacifici, G. M. (2006) 'Placental transfer of antibiotics administered to the mother: A review', *International Journal of Clinical Pharmacology and Therapeutics*. Int J Clin Pharmacol Ther, pp. 57–63. doi: 10.5414/CPP44057.

Parker, A., Fonseca, S. and Carding, S. R. (2020) 'Gut microbes and metabolites as modulators of blood-brain barrier integrity and brain health', *Gut Microbes*. Taylor & Francis, pp. 135–157. doi: 10.1080/19490976.2019.1638722.

Parks, D. H. *et al.* (2018) 'A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life', *Nature Biotechnology 2018 36:10*. Nature Publishing Group, 36(10), pp. 996–1004. doi: 10.1038/nbt.4229.

Paterson, M. J., Oh, S. and Underhill, D. M. (2017) 'Host–microbe interactions: commensal fungi in the gut', *Current Opinion in Microbiology*. NIH Public Access, pp. 131–137. doi: 10.1016/j.mib.2017.11.012.

Penington, J. S. *et al.* (2018) 'Influence of fecal collection conditions and 16S rRNA gene sequencing at two centers on human gut microbiota analysis', *Scientific Reports*, 8(1). doi: 10.1038/s41598-018-22491-7.

Persoon, I. F. *et al.* (2017) 'The mycobiome of root canal infections is correlated to the bacteriome', *Clinical Oral Investigations*. Springer Verlag, 21(5), pp. 1871–1881. doi: 10.1007/s00784-016-1980-3.

Pessa-Morikawa, T. *et al.* (2022) 'Maternal microbiota-derived metabolic profile in fetal murine intestine, brain and placenta', *BMC Microbiology*. BioMed Central Ltd, 22(1), pp. 1–16. doi: 10.1186/S12866-022-02457-6/FIGURES/8.

Pillai, P., Gouhier, T. C. and Vollmer, S. V. (2014) 'The cryptic role of biodiversity in the emergence of host-microbial mutualisms', *Ecology Letters*, 17(11), pp. 1437–1446. doi: 10.1111/ele.12349.

Pistollato, F. *et al.* (2016) 'Role of gut microbiota and nutrients in amyloid formation and pathogenesis of Alzheimer disease', *Nutrition Reviews*. Oxford Academic, 74(10), pp. 624–634. doi: 10.1093/nutrit/nuw023.

Pollock, J. *et al.* (2018) 'The madness of microbiome: Attempting to find consensus "best practice" for 16S microbiome studies', *Applied and Environmental Microbiology*. American Society for Microbiology, 84(7), pp. 248–258. doi: 10.1128/AEM.02627-17.

Prescott, S. L. *et al.* (2013) 'A global survey of changing patterns of food allergy burden in children', *World Allergy Organization Journal*. BioMed Central, 6(1), pp. 1–12. doi: 10.1186/1939-4551-6-21.

Prodan, A., Tremaroli, V., Brolin, H., Zwinderman, A. H., Nieuwdorp, M., & Levin, E. (2020) 'Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing', PLOS ONE, 15(1), e0227434. https://doi.org/10.1371/journal.pone.0227434

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

Rackaityte, E. *et al.* (2020) 'Viable bacterial colonization is highly limited in the human intestine in utero', *Nature Medicine*. Nature Publishing Group, 26(4), pp. 599–607. doi: 10.1038/s41591-020-0761-3.

Rakoff-Nahoum, S., Foster, K. R. and Comstock, L. E. (2016) 'The evolution of cooperation within the gut microbiota', *Nature*. Nature Publishing Group, 533(7602), pp. 255–259. doi: 10.1038/nature17626.

Ramakodi, M. P. (2022) 'Influence of 16S rRNA reference databases in amplicon-based environmental microbiome research', *Biotechnology Letters*. Springer Science and Business Media B.V., 44(3), pp. 523–533. doi: 10.1007/S10529-022-03233-2/FIGURES/5.

Ramos, N. S., Hernández, S. M. and Blanca, M. J. (2009) 'Intestinal Short Chain Fatty Acids and their Link with Diet and Human Health', *Ansiedad y Estres*. Frontiers Media SA, 15(2–3), pp. 207–216. doi: 10.3389/fmicb.2016.00185.

Ranjan, R. *et al.* (2016) 'Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing', *Biochemical and Biophysical Research Communications*. Academic Press, 469(4), pp. 967–977. doi: 10.1016/j.bbrc.2015.12.083.

Rao, M. and Gershon, M. D. (2016) 'The bowel and beyond: The enteric nervous system in neurological disorders', *Nature Reviews Gastroenterology and Hepatology*. Nature Publishing Group, pp. 517–528. doi: 10.1038/nrgastro.2016.107.

Ríos-Covián, D. *et al.* (2016) 'Intestinal short chain fatty acids and their link with diet and human health', *Frontiers in Microbiology*. Frontiers Media S.A., 7(185), p. 185. doi: 10.3389/fmicb.2016.00185.

Roberts, M. H. and Erdei, E. (2020) 'Comparative United States autoimmune disease rates for 2010–2016 by sex, geographic region, and race', *Autoimmunity Reviews*. Elsevier, 19(1), p. 102423. doi: 10.1016/J.AUTREV.2019.102423.

Rodríguez, J. M. (2014) 'The Origin of Human Milk Bacteria: Is There a Bacterial Entero-Mammary Pathway during Late Pregnancy and Lactation?', *Advances in Nutrition*. Oxford University Press, 5(6), p. 779. doi: 10.3945/AN.114.007229.

Romani, L. *et al.* (2015) 'The cross-talk between opportunistic fungi and the mammalian host via microbiota's metabolism', *Seminars in Immunopathology*, pp. 163–171. doi: 10.1007/s00281-014-0464-2.

Rose, C. *et al.* (2015) 'The characterization of feces and urine: A review of the literature to inform advanced treatment technology', *Critical Reviews in Environmental Science*

and Technology. Taylor & Francis, 45(17), pp. 1827–1879. doi: 10.1080/10643389.2014.1000761.

Saa, P. *et al.* (2022) 'Modeling approaches for probing cross-feeding interactions in the human gut microbiome', *Computational and Structural Biotechnology Journal*. Elsevier, 20, pp. 79–89. doi: 10.1016/J.CSBJ.2021.12.006.

Saari, A. *et al.* (2015) 'Antibiotic exposure in infancy and risk of being overweight in the first 24 months of life.', *Pediatrics*. American Academy of Pediatrics, 135(4), pp. 617–26. doi: 10.1542/peds.2014-3407.

Sam, Q. H., Chang, M. W. and Chai, L. Y. A. (2017) 'The fungal mycobiome and its interaction with gut bacteria in the host', *International Journal of Molecular Sciences*. MDPI AG. doi: 10.3390/ijms18020330.

Sanada, K. *et al.* (2020) 'Gut microbiota and major depressive disorder: A systematic review and meta-analysis', *Journal of Affective Disorders*. Elsevier, 266, pp. 1–13. doi: 10.1016/J.JAD.2020.01.102.

Sanajou, S., Erkekoğlu, P., Şahin, G., & Baydar, T. (2022) 'Role of aluminum exposure on Alzheimer's disease and related glycogen synthase kinase pathway. In *Drug and Chemical Toxicology*', Taylor & Francis. https://doi.org/10.1080/01480545.2022.2065291

Sanders, M. E. *et al.* (2019) 'Probiotics and prebiotics in intestinal health and disease: from biology to the clinic', *Nature Reviews Gastroenterology and Hepatology*. Nature Publishing Group, pp. 605–616. doi: 10.1038/s41575-019-0173-3.

Sandler, R. H. *et al.* (2000) 'Short-term benefit from oral vancomycin treatment of regressive-onset autism', *Journal of Child Neurology*. Sage PublicationsSage CA: Thousand Oaks, CA, 15(7), pp. 429–435. doi: 10.1177/088307380001500701.

Sanidad, K. Z., Xiao, H. and Zhang, G. (2018) 'Triclosan, a common antimicrobial ingredient, on gut microbiota and gut health', *https://doi.org/10.1080/19490976.2018.1546521*. Taylor & Francis, 10(3), pp. 434–437. doi: 10.1080/19490976.2018.1546521.

Santocchi, E. *et al.* (2020) 'Effects of Probiotic Supplementation on Gastrointestinal, Sensory and Core Symptoms in Autism Spectrum Disorders: A Randomized Controlled Trial', *Frontiers in Psychiatry*. Frontiers Media S.A., 11, p. 944. doi: 10.3389/fpsyt.2020.550593.

Dos Santos, S. J. *et al.* (2021) ' Early Neonatal Meconium Does Not Have a Demonstrable Microbiota Determined through Use of Robust Negative Controls with cpn 60-Based Microbiome Profiling ', *Microbiology Spectrum*. American Society for Microbiology, 9(2). doi: 10.1128/SPECTRUM.00067-21/SUPPL_FILE/REVIEWER-COMMENTS.PDF.

Schnell, I. B., Bohmann, K. and Gilbert, M. T. P. (2015) 'Tag jumps illuminated - reducing sequence-to-sample misidentifications in metabarcoding studies', *Molecular Ecology Resources*. John Wiley & Sons, Ltd (10.1111), 15(6), pp. 1289–1303. doi: 10.1111/1755-0998.12402.

Scholtens, P. A. M. J. *et al.* (2008) 'Fecal Secretory Immunoglobulin A Is Increased in Healthy Infants Who Receive a Formula with Short-Chain Galacto-Oligosaccharides and Long-Chain Fructo-Oligosaccharides', *The Journal of Nutrition*, 138(6), pp. 1141–1147. doi: 10.1093/jn/138.6.1141.

Schroeder, B. O. (2019) 'Fight them or feed them: how the intestinal mucus layer manages the gut microbiota', *Gastroenterology Report*. Oxford Academic, 7(1), pp. 3–12. doi: 10.1093/GASTRO/GOY052.

Seijo, M. *et al.* (2019) 'Effect of Adding a Galacto-Oligosaccharides/Fructo-Oligosaccharides (GOS/FOS®) Mixture to a Normal and Low Calcium Diet, on Calcium Absorption and Bone Health in Ovariectomy-Induced Osteopenic Rats', *Calcified Tissue International*. Springer New York LLC, 104(3), pp. 301–312. doi: 10.1007/S00223-018-0490-5/FIGURES/1.

Seth, E. C. and Taga, M. E. (2014) 'Nutrient cross-feeding in the microbial world', *Frontiers in Microbiology*. Frontiers Research Foundation, 5(JULY), p. 350. doi: 10.3389/FMICB.2014.00350/BIBTEX.

Sharma, R., Neu, J. and Young, C. (2010) 'Molecular modulation of intestinal epithelial barrier: Contribution of microbiota', *Journal of Biomedicine and Biotechnology*, 2010(Figure 1), pp. 11–14. doi: 10.1155/2010/305879.

Sierra, C. *et al.* (2015) 'Prebiotic effect during the first year of life in healthy infants fed formula containing GOS as the only prebiotic: a multicentre, randomised, double-blind and placebo-controlled trial.', *European journal of nutrition*. Springer, 54(1), pp. 89–99. doi: 10.1007/s00394-014-0689-9.

Sinclair, L. *et al.* (2015) 'Microbial community composition and diversity via 16S rRNA gene amplicons: Evaluating the illumina platform', *PLoS ONE*. Public Library of Science, 10(2). doi: 10.1371/journal.pone.0116955.

Sonnenburg, E. D. *et al.* (2016) 'Diet-induced extinctions in the gut microbiota compound over generations', *Nature*. Nature Publishing Group, 529(7585), pp. 212–215. doi: 10.1038/nature16504.

Stokholm, J. *et al.* (2013) 'Prevalence and Predictors of Antibiotic Administration during Pregnancy and Birth', *PLOS ONE*. Public Library of Science, 8(12), p. e82932. doi: 10.1371/JOURNAL.PONE.0082932.

Stoler, N. and Nekrutenko, A. (2021) 'Sequencing error profiles of Illumina sequencing instruments', *NAR Genomics and Bioinformatics*. Oxford Academic, 3(1). doi: 10.1093/NARGAB/LQAB019.

Talbott, S. M. *et al.* (2019) 'Effect of Coordinated Probiotic/Prebiotic/Phytobiotic Supplementation on Microbiome Balance and Psychological Mood State in Healthy Stressed Adults', *Functional Foods in Health and Disease*. Functional Food Institute, 9(4), pp. 265–275. doi: 10.31989/FFHD.V9I4.599.

Tanaka, M. *et al.* (2017) 'Signatures in the gut microbiota of Japanese infants who developed food allergies in early childhood', *FEMS microbiology ecology*. Oxford University Press, 93(8). doi: 10.1093/femsec/fix099.

Tang, M. L. K. and Mullins, R. J. (2017) 'Food allergy: is prevalence increasing?', *Internal Medicine Journal*. Wiley/Blackwell (10.1111), 47(3), pp. 256–261. doi: 10.1111/imj.13362.

Tiew, P. Y. *et al.* (2020) 'The Mycobiome in Health and Disease: Emerging Concepts, Methodologies and Challenges', *Mycopathologia 2020 185:2*. Springer, 185(2), pp. 207–231. doi: 10.1007/S11046-019-00413-Z.

van Tilburg Bernardes, E. *et al.* (2020) 'Intestinal fungi are causally implicated in microbiome assembly and immune development in mice', *Nature Communications*. Nature Publishing Group, 11(1), pp. 1–16. doi: 10.1038/s41467-020-16431-1.

Tomova, A. *et al.* (2019) 'The effects of vegetarian and vegan diets on gut microbiota', *Frontiers in Nutrition*. Frontiers Media S.A., 6, p. 47. doi: 10.3389/FNUT.2019.00047/BIBTEX.

De Vadder, F. *et al.* (2018) 'Gut microbiota regulates maturation of the adult enteric nervous system via enteric serotonin networks', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 115(25), pp. 6458–6463. doi: 10.1073/pnas.1720017115.

Valles-Colomer, M. *et al.* (2019) 'The neuroactive potential of the human gut microbiota in quality of life and depression', *Nature Microbiology*. Nature Publishing Group, 4(4), pp. 623–632. doi: 10.1038/s41564-018-0337-x.

Venegas, D. P. *et al.* (2019) 'Short chain fatty acids (SCFAs)mediated gut epithelial and immune regulation and its relevance for inflammatory bowel diseases', *Frontiers in Immunology*. Frontiers Media S.A., p. 277. doi: 10.3389/fimmu.2019.00277.

Verbeke, K. A. *et al.* (2015) 'Towards microbial fermentation metabolites as markers for health benefits of prebiotics', *Nutrition Research Reviews*, 28(1), pp. 42–66. doi: 10.1017/S0954422415000037.

Vitetta, L. *et al.* (2014) 'Probiotics, prebiotics and the gastrointestinal tract in health and disease', *Inflammopharmacology*, 22(3), pp. 135–154. doi: 10.1007/s10787-014-0201-4.

Voigt, A. Y. *et al.* (2015) 'Temporal and technical variability of human gut metagenomes', *Genome Biology*. BioMed Central Ltd., 16(1), p. 73. doi: 10.1186/s13059-015-0639-8.

Vulevic, J. *et al.* (2013) 'A Mixture of trans-Galactooligosaccharides Reduces Markers of Metabolic Syndrome and Modulates the Fecal Microbiota and Immune Function of Overweight Adults', *The Journal of Nutrition*. Oxford Academic, 143(3), pp. 324–331. doi: 10.3945/JN.112.166132.

Wagner Mackenzie, B., Waite, D. W. and Taylor, M. W. (2015) 'Evaluating variation in human gut microbiota profiles due to DNA extraction method and inter-subject differences.', *Frontiers in microbiology*. Frontiers Media SA, 6, p. 130. doi: 10.3389/fmicb.2015.00130.

Waldor, M. K. *et al.* (2015) 'Where Next for Microbiome Research?', *PLoS Biology*. Public Library of Science, 13(1). doi: 10.1371/journal.pbio.1002050.

Wampach, L. *et al.* (2017) 'Colonization and succession within the human gut microbiome by archaea, bacteria, and microeukaryotes during the first year of life', *Frontiers in Microbiology*. Frontiers, 8(MAY), p. 738. doi: 10.3389/fmicb.2017.00738.

Wang, Z. *et al.* (2018) 'Comparison of Fecal Collection Methods for Microbiome and Metabolomics Studies.', *Frontiers in cellular and infection microbiology*. Frontiers Media SA, 8, p. 301. doi: 10.3389/fcimb.2018.00301.

Ward, T. L. *et al.* (2018) 'Development of the Human Mycobiome over the First Month of Life and across Body Sites', *mSystems*. mSystems, 3(3). doi: 10.1128/msystems.00140-17.

Van De Wiele, T. *et al.* (2007) 'Inulin-type fructans of longer degree of polymerization exert more pronounced in vitro prebiotic effects', *Journal of Applied Microbiology*. John Wiley & Sons, Ltd, 102(2), pp. 452–460. doi: 10.1111/J.1365-2672.2006.03084.X.

Wilson, B. C. *et al.* (2021) 'Oral administration of maternal vaginal microbes at birth to restore gut microbiome development in infants born by caesarean section: A pilot randomised placebo-controlled trial', *EBioMedicine*. Elsevier, 69, p. 103443. doi: 10.1016/j.ebiom.2021.103443.

Wu, W. K. *et al.* (2018) 'Optimization of fecal sample processing for microbiome study - The journey from bathroom to bench', *Journal of the Formosan Medical Association*. doi: 10.1016/j.jfma.2018.02.005.

Xu, M. *et al.* (2019) 'Association between gut microbiota and autism spectrum disorder: A systematic review and meta-analysis', *Frontiers in Psychiatry*. Frontiers Media S.A. doi: 10.3389/fpsyt.2019.00473.

Xu, Z. and Knight, R. (2015) 'Dietary effects on human gut microbiome diversity', *British Journal of Nutrition*, 113, pp. S1–S5. doi: 10.1017/S0007114514004127.

Yadav, H. *et al.* (2013) 'Beneficial metabolic effects of a probiotic via butyrate-induced GLP-1 hormone secretion', *Journal of Biological Chemistry*, 288(35), pp. 25088–25097. doi: 10.1074/jbc.M113.452516.

Yadav, M. *et al.* (2018) ‘A review of metabolic potential of human gut microbiome in human nutrition’, *Archives of Microbiology*. Springer Berlin Heidelberg, 200(203), pp. 203–217. doi: 10.1007/s00203-017-1459-x.

Yang, B., Wang, Y. and Qian, P. Y. (2016) ‘Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis’, *BMC Bioinformatics*. BioMed Central, 17(1), p. 135. doi: 10.1186/s12859-016-0992-y.

Yang, L. *et al.* (2021) ‘Recent advances of intestinal microbiota transmission from mother to infant’, *Journal of Functional Foods*. Elsevier, 87, p. 104719. doi: 10.1016/J.JFF.2021.104719.

Yassour, M. *et al.* (2016) ‘Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability.’, *Science translational medicine*. NIH Public Access, 8(343), p. 343ra81. doi: 10.1126/scitranslmed.aad0917.

Yatsunenko, T. *et al.* (2012) ‘Human gut microbiome viewed across age and geography’, *Nature*. Nature Publishing Group, pp. 222–227. doi: 10.1038/nature11053.

Zhai, R. *et al.* (2019) ‘Strain-Specific Anti-inflammatory Properties of Two Akkermansia muciniphila Strains on Chronic Colitis in Mice’, *Frontiers in Cellular and Infection Microbiology*. Frontiers, 0(JUL), p. 239. doi: 10.3389/FCIMB.2019.00239.

Zhang, J. *et al.* (2020) ‘Improved High-Throughput Sequencing of the Human Oral Microbiome: From Illumina to PacBio’, *Canadian Journal of Infectious Diseases and Medical Microbiology*. doi: 10.1155/2020/6678872.

Zhang, L. *et al.* (2021) ‘The role of gut mycobiome in health and diseases’, *Therapeutic Advances in Gastroenterology*. SAGE PublicationsSage UK: London, England. doi: 10.1177/17562848211047130.

Ziętek, M., Celewicz, Z. and Szczuko, M. (2021) ‘Short-chain fatty acids, maternal microbiota and metabolism in pregnancy’, *Nutrients*. Multidisciplinary Digital Publishing Institute, p. 1244. doi: 10.3390/nu13041244.

Zijlmans, M. A. C. *et al.* (2015) ‘Maternal prenatal stress is associated with the infant intestinal microbiota’, *Psychoneuroendocrinology*. Pergamon, 53, pp. 233–245. doi: 10.1016/J.PSYNEUEN.2015.01.006.

Zimmermann, P. and Curtis, N. (2020) ‘Breast milk microbiota: A review of the factors that influence composition’, *Journal of Infection*. W.B. Saunders, 81(1), pp. 17–47. doi: 10.1016/J.JINF.2020.01.023.

– Chapter 2 –
Fecal sample collection methods and time of day impact microbiome and SCFA concentrations

2.1 Preamble

The way in which stool samples are collected for microbiome analysis varies widely in the literature, most notably a wide range is seen in: the mass of stool collected; the time and number of bowel movements collected; how stool is preserved, either in buffer at room temperature or frozen; as well as the timeframe in which it is frozen. In the third and fourth chapters of this thesis, stool samples were collected in two different ways by researchers associated with larger studies. In chapter 3, a large portion of stool (~8-20 g) was collected from one bowel movement using a large, scooped jar (AICES study). In chapter 4, small portions of stool (0.3 – 8 g) were collected in triplicate from a single bowel movement, with the intention of using one replicate per analysis (SYMBA study). In this chapter, I provide evidence for 3 different ways sampling bias can affect microbiome and SCFA profiles that are relevant to the way stool was collected in chapters 3 and 4. The results of this investigation were used to justify requesting all three sub samples collected for the SYMBA study, so they could be combined and analysed in a way that would better capture the faecal environment of each individual.

Furthermore, in this chapter the level of detail used to convey the procedure for stool collection and microbiome library preparation, as well as results describing data quality exceed the low standard that is currently seen in microbiome research. The current published literature often does not adequately describe the sample collection process, omits, or disregards the use of control samples to mitigate reagent contamination, and neglects a description of the quality of reads after bioinformatic processing. Each of these considerations would allow for better study replication, which is highlighted as lacking in molecular microbiome investigations, and improve the quality of the data thus improving interpretations of the data. In this study I set a high standard for data quality and use positive and negative controls to monitor and justify this claim. In chapter 3, and 4, I used the methods developed in chapter 2 to examine the microbiome and SCFA concentrations of larger more complex cohorts of children and adults. In particular, the use of control samples was invaluable in chapter 4, where I both mitigate and describing the potential impact of combining the large multi-sequence run datasets. Overall, chapter 2 provides the justification for how microbiome and SCFA data was collected, processed and interpreted in this thesis. Lastly, in keeping with the proposed reporting guidelines which have been tailored for microbiome research a “Strengthening The Organization and Reporting of Microbiome Studies’ (STORMS)” checklist has been

completed for each data chapter of the thesis. The published manuscript described in this chapter successfully hit 46 of 55 relevant criteria (Appendix 2).

2.2 Contributions and data accessibility

A statement regarding author contributions, acknowledgments and a link to the raw data used in this chapter are included within the manuscript.

2.3 Manuscript

The following section of chapter 2 contains the manuscript published in Scientific Reports, July 7th, 2021. https://www.nature.com/articles/s41598-021-93031-z

Fecal sample collection methods and time of day impact microbiome composition and short chain fatty acid concentrations

***Jacquelyn Jones[ab], Stacey Reinke[c], Alishum Ali[ab], Debra Palmer[de], Claus T. Christophersen[abcf]**
[a]Trace and Environmental DNA Laboratory, School of Molecular and Life Sciences, Curtin University, Bentley, WA, Australia
[b]The Western Australian Human Microbiome Collaboration Centre, Curtin University, Bentley, WA, Australia
[c]Centre for Integrative Metabolomics and Computational Biology, School of Science, Edith Cowan University, Joondalup, WA, Australia
[d]Telethon Kids Institute, University of Western Australia, Nedlands, WA, Australia
[e]School of Medicine, University of Western Australia, Crawley, WA, Australia
[f]School of Medical & Health Sciences, Edith Cowan University, Joondalup, WA, Australia

Associations between the human gut microbiome and health outcomes continues to be of great interest, although fecal sample collection methods which impact microbiome studies are sometimes neglected. Here, we expand on previous work in sample optimization, to promote high quality microbiome data. To compare fecal sample collection methods, amplicons from the bacterial 16S rRNA gene (V4) and fungal (ITS2) region, as well as short chain fatty acid (SCFA) concentrations were determined in fecal material over three timepoints. We demonstrated that spot sampling of stool results in variable detection of some microbial members, and inconsistent levels of SCFA; therefore, sample homogenization prior to subsequent analysis or subsampling is recommended. We also identify a trend in microbial and metabolite composition that shifts over two consecutive stool collections less than 25 h apart. Lastly, we show significant differences in bacterial composition that result from collecting stool samples in OMNIgene·Gut tube (DNA Genotec) or Stool Nucleic Acid Collection and Preservation Tube (NORGEN) compared to immediate freezing. To assist with planning fecal sample collection and storage procedures for microbiome investigations with multiple analyses, we recommend participants to collect the first full bowel movement of the day and freeze the sample immediately after collection.

2.4 Introduction

Our understanding of the relationship between the human gut microbiome and host continues to expand from explorations which describe inhabitants, to studies which demonstrate the involvement of the microbiome in human health and disease and disorders. Some examples include neurological disorders such as depression (Valles-Colomer *et al.*, 2019), Alzheimer's disease (Vogt *et al.*, 2017) and Autism Spectrum Disorder (Strati *et al.*, 2017), as well as inflammatory diseases such as food allergies

(Fieten *et al.*, 2018), and inflammatory bowel diseases (Venegas *et al.*, 2019). Advancements in microbiome studies have been accelerated by increased sequencing capabilities (D'Amore *et al.*, 2016), along with sensitive analytical techniques tailored for the quantification of metabolites in fecal material (Gratton *et al.*, 2016; Karu *et al.*, 2018). Short chain fatty acids (SCFA) are metabolites produced exclusively by resident bacteria, and are associated with dysbiosis, hypertension (De la Cuesta-Zuluaga *et al.*, 2018) and other inflammatory disorders. Investigating the gut microbiome by combining microbial sequencing data and metabolomic approaches has been an important step in unraveling associations between resident bacteria, SCFA, and health outcomes (Arrieta et al., 2018; De Filippis et al., 2016; Lagkouvardos et al., 2015).

However, stool, which is used as proxy for the distal colon microbiome, is a complex matrix of endo- and exogenous material containing a heterogeneous distribution of microorganisms (Donaldson, Lee and Mazmanian, 2015), which is susceptible to changes during and after collection. Microbiome profiles may be misrepresented due to subsampling of non-homogenized stool as seen in Gorzelak (2015) where large variations in bacterial abundance detected via qPCR in non-homogenized stool samples were significantly reduced after stool homogenization. In addition, the effects of sub-sampling stool may be further amplified when performing metabolomic analyses, as highly sensitive techniques are used (Couch *et al.*, 2013; Gratton *et al.*, 2016).

Stool collection by participants may be an undesirable yet necessary aspect of partaking in a microbiome study. Providing participants with a clean and simple collection method should increase compliance, but also maintain sample integrity. Some commercial stool collection tubes allow for easy collection and short term (~14 days) ambient temperature storage; however, some of these have been associated with changes in proportions of bacterial phyla (Penington *et al.*, 2018). A final consideration is the level of inter-individual differences that occur in the fecal microbiome over a week (Flores *et al.*, 2014), and even from day to day (Caporaso, Lauber, Costello, *et al.*, 2011), meaning that collection periods may need to span a number of days, or be collected at a particular time in the day to accurately capture the inherent variability. As far as the authors are aware spatial and short-term temporal variability of bacterial and fungal communities has never been evaluated together with SCFA composition. To address this gap, this study will assess the effects of four fecal sample collection methods, as well as consecutively collected whole stool samples (less than 25 hours apart), on the variability of the fecal microbiome. The comparisons will be drawn from bacterial and fungal community composition as well as SCFA profiling.

2.5 Results

2.5.1 Overview of microbiome taxonomy and SCFA concentrations

Stool samples yielded bacterial communities (bacteriome) from all individuals, while fungal communities (mycobiome) were successfully sequenced in 50 of 84 samples, but with uneven library size (~100x). Overall, the fecal bacteriome had a higher number of ASVs than the mycobiome (Supplementary Table 2.1). Across all individuals and collection methods, the most abundant bacterial families were Bacteroidaceae and Lachnospiraceae, which made up 38% and 10% of the bacteriome, while the most abundant fungal families were Saccharomycetaceae 90%, and Phaffomycetaceae 7%. To account for technical bias in library preparation, a single sample from one individual was also processed in duplicate for each gene region. Bacterial alpha diversity estimates for this replicate sample were more similar than the fungal replicate, while fungal replicates also had low richness, indicating that both the rarity of this community, and the library preparation may impact the interpretation of relative abundance of fungal communities (Supplementary Table 2.1).

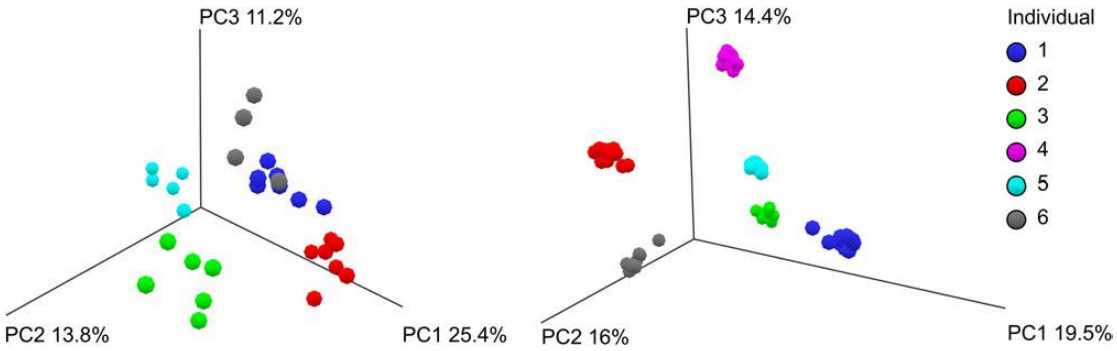


Figure 2.1. *Clustering of microbiome communities per individual from all directly frozen stool samples collected per individual. Plots show the mycobiome (A) of each individual is less distinct than bacteriome (B). Data was CLR transformed and ordination based on Euclidian distances.*

Bacterial and Fungal mock communities were also sequenced as positive controls, which allowed reads of the mock community samples to be compared to the known composition of the mock community (Supplementary Figure 2.1). Still, both mock communities contain a number of species that are not expected to be captured in the human fecal microbiome, and therefore, may not be amplified by the selected primers. Of the 20 bacterial species known to be in the mock community, 18 were correctly detected to family level and 15 to genus level, leaving two species of the mock community unidentified. At the level of order, the proportion of each ASV in the positive control was compared to the known percentage contribution of the mock community. Actinomycetales, Campylobacterales and Rhodobacterales were

underrepresented at 3.5, 3.2, and 4.6 times less that what was expected; while Bacillales and Clostridiales were over represented at 3.4 and 4.4 times more than expected. Of the 19 fungal species in the mock community 13 were sequenced and correctly resolved to family and genus (however, *R. irregularis* only had 11 reads), leaving six species of the mock community unidentified. However, four of these species (*Chytriomyces hyalinus*, *Rhizomucor miehei*, *Rhizoctonia solani*, and *Ustilago maydis*) are not expected to be part of the human microbiome, and were not detected with the ITS2 primers developed for use in the human fecal microbiome.

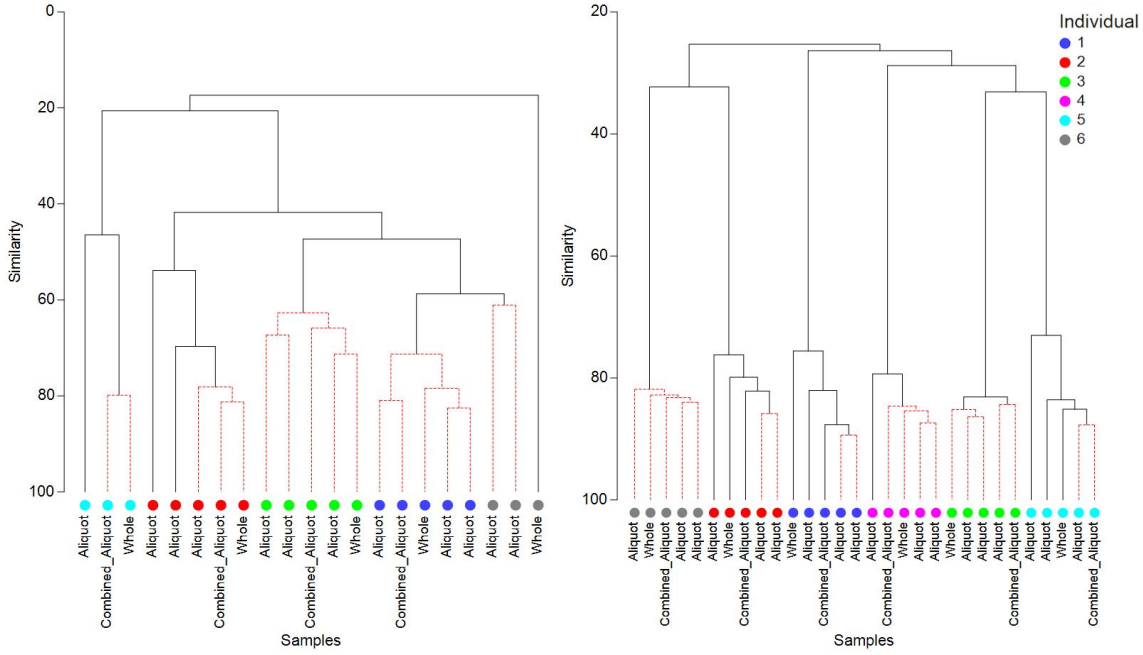


Figure 2.2. *Clustering mycobiome (A), and bacteriome (B) from whole stool, aliquot, and combined aliquot sampling methods. A SIMPROF test with alpha = 0.05 was used to determine significantly similar sub groupings within individuals, with red dashed lines grouping samples which are significantly similar, and solid black likes grouping samples which are not significantly similar. Group average cluster analysis (9999 repeats) on Bray– Curtis similarity.*

The mycobiome signature of each subject was not as distinct as the bacteriome (Figure 2.1). While fungal communities tended to cluster by subject, an analysis by PCO, shows the mycobiome of individuals 1 and 6 overlaps, which seems to be driven by both individuals having a composition of ≥99% Saccharomyces. An analysis of Beta diversity (Euclidian, and Bray-Curtis) between individuals over the collection period – with all sample types and sample points – showed significant differences in both bacterial and fungal communities (PERMANOVA $p < 0.02$).

SCFA concentrations were determined from whole stool as well as surface collected aliquots and overall, the average molar ratio of acetate, propionate, and butyrate was

78:12:10 respectively. The mean concentrations of individual or total SCFA in µmol per gram of faeces was not significantly different between collection methods; and in all subjects, acetic acid was most variable, ranging from an average of 103 to 697 µmol g^{-1}.

2.5.2 Comparison of surface aliquots and whole stool sampling methods

To assess the impact of sampling method on a-priori grouping by individual, a hierarchical cluster analysis was performed on bacterial and fungal communities (Figure 2.2). Bacterial communities from the same subject grouped together, with a SIMPROF test identifying significantly different sub clusters for 5 of 6 individuals. Fungal communities also clustered according to individual, but these groups were less similar than their respective bacterial groups. Furthermore, the aliquots from participant 6 clustered more closely to participant 1 than to its own respective whole stool sample. The mycobiome and bacteriome from the combined aliquot clustered according to individual however, did not seem to align consistently with the other aliquots.

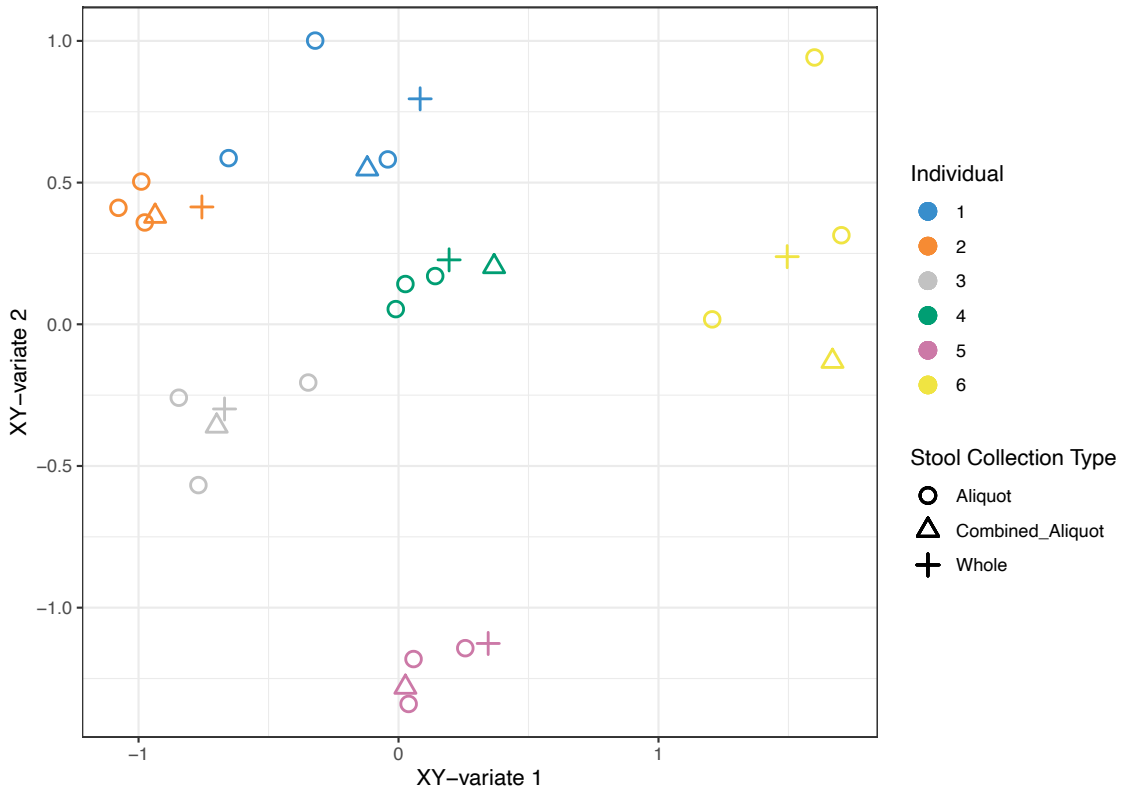


Figure 2.3. *Variability of bacterial community and SCFA composition due to stool homogenization methods. Datasets where combined using rCCA in the R package MixOmics, correlation coefficients were plotted in the shared X–Y space. Bacterial ASV counts were CLR transformed, and SCFA concentration were log10 transformed.*

To assess the heterogeneity of SCFAs and bacterial diversity within a single stool, the coefficient of variation (CV) for these measures was compared across three aliquots from a single stool and three separate stools (collected over 3 days) per individual (Supplementary Table 2.2). Acetic acid and valeric acid, were found to be as variable along a single stool as they were across three bowel movements, whereas propionic was more variable across bowel movements. Overall, the SCFA concentrations were more variable across 3 bowel movements than along a single stool except in individuals one and six. Shannon diversity was less variable along a single stool (five of six individuals), Chao1 species richness was more variable along a single stool (four of six individuals), and Phylogenetic diversity was equally variable along a single stool (three of three individuals). This trend was further assessed by integrating SCFA data with bacteriome data through rCCA (distance between features were relatively short, indicating the strong agreement between datasets), and the plotted canonical variates show the variability between the surface collected samples was still evident in subjects one, and six (Figure 2.3).

ASV	Taxa	log2 fold change
Bacteria		
591	*Anaerotruncus massiliensis*	6.62
790	Anaerovoracaceae	17.92[a]
802	Anaerovoracaceae	17.92[a]
70	*Eubacterium* sp.	2.89
4	*Fecalibacterium prausnitzii*	−3.41
27	*Fecalibacterium prausnitzii*	2.71
405	*Fecalibacterium prausnitzii*	17.47[a]
717	*Oscillibacter ruminantium*	19.52[a]
461	Rhizobiaceae	18.36[a]
Fungi		
19	*Alternaria alternata*	−2.9
219	*Aspergillus niger*	3.9
119	*Aureobasidium pullulans*	17.2[a]
8	*Cyberlindnera jadinii*	3.9
4	*Eremothecium sinecaudum*	−3.5
41	*Hanseniaspora uvarum*	−24.1
11	*Hanseniaspora uvarum*	−8.6
5	*Kazachstania barnettii*	−22.5
2	*Kazachstania servazzii*	−21.4[a]
132	*Rhodotorula mucilaginosa*	9.1
29	*Saccharomyces cerevisiae*	−23.5
6	*Saccharomyces cerevisiae*	4.9
17	*Sporopachydermia lactativora*	5.0
40	*Wickerhamomyces ciferrii*	7.3

Table 2.1. *ASVs identified with log2 fold change in gematric mean abundance greater than |2.5| between homogenized whole stool and stool aliquots.* [a]*significant enrichment.*

Within each individual the composition of microorganisms within the aliquots were not identical to each other, or to the whole stool from which they were sub sampled. DESEq2 was used to compare differentially abundant ASVs between aliquots and whole stool. This method identified 12 bacterial, and 16 fungal features with a $\log_2$ fold change greater than |2.5|. Of these, 5 bacterial and 1 fungal ASVs were significantly enriched in the whole stool compared to aliquots, and 1 fungal ASV was enriched in stool aliquots compared to whole stool (Table 1).

2.5.3 Bacterial community composition is affected by collection methods

Differences in bacterial communities due to collection method where visualized using PCO (Figure 2.4) showing a clear separation between directly frozen samples (method F) and those collected with either the Norgen (collection method N), or OMNIgene tubes (collection method O). Significant differences in beta diversity were tested using PERMANOVA, and stool collected with method F were significantly different ($p < 0.01$) to both those collected with the N or O methods. The N and O method where also significantly different to each other ($p = 0.04$). In whole stool samples, the overall most abundant families were Bacteroidaceae (F 38%, N 43%, O 39%), Ruminococcaceae (F 7%, N 17%, O 33%), and Lachnospiraceae (F 10%, N 15%, O 10%), with the abundance of Ruminococcacae significantly increased (ANOVA $P < 0.001$, FDR = 0.007) due to collection using the N and O methods compared to the F method. A number of taxa were also recovered differentially between the three collection methods (Supplementary Table 4), including some high-ranking taxa (Figure 2.5).

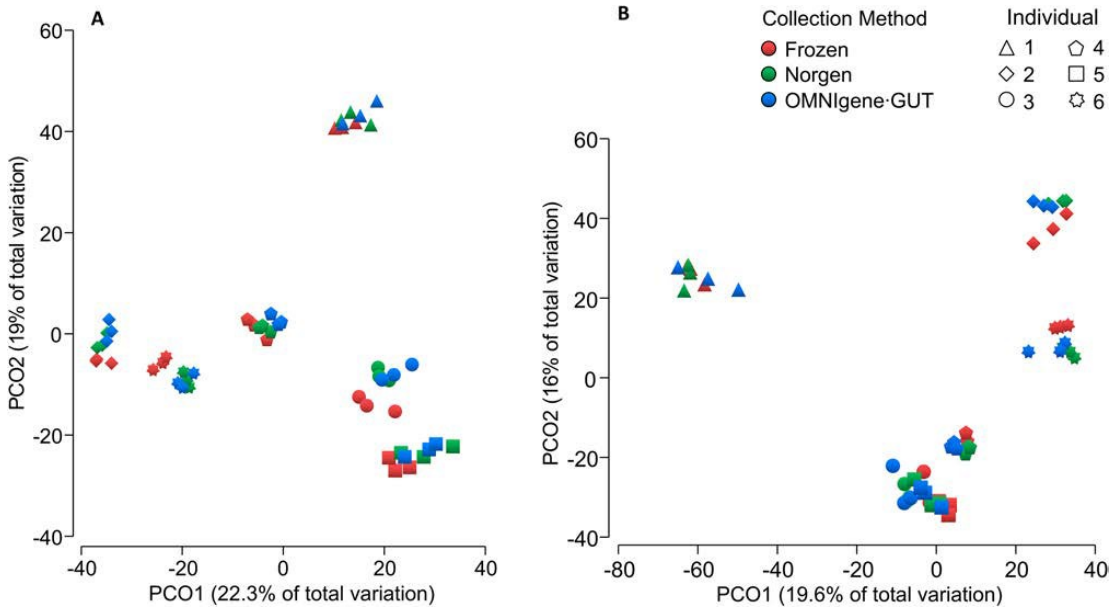


Figure 2.4. *Principal coordinates analysis of β-diversity showing stool samples collected in frozen (n = 18), Norgen (n = 18) and OMNIgene·Gut (n = 18) tube types. Clustering shows directly frozen samples are easily distinguished from stool collected in stabilizing liquid. Data shown in (A) was 4th root transformed, using Bray–Curtis similarity distance and (B) CLR transformed and Euclidian distance.*

2.5.4 Short term changes to microbiome composition and SCFA concentration

All 6 participants collected two bowel movements consecutively within a 25-hour period, with 5 of the 6 individuals producing two bowel movements within 10 hours. The total SCFA concentration ($p = 0.04$) and acetic acid concentration ($p = 0.03$) were significantly higher in the second stool sample compared to the first using a paired t-test. While not significant, bacterial richness ($p = 0.45$) and diversity estimates ($p = 0.95$) were also lower in the second stool collection for 4 of 6 individuals (Figure 2.6). Fungal communities from three individuals which were successfully profiled consecutively did not show any trend between richness and diversity measures.

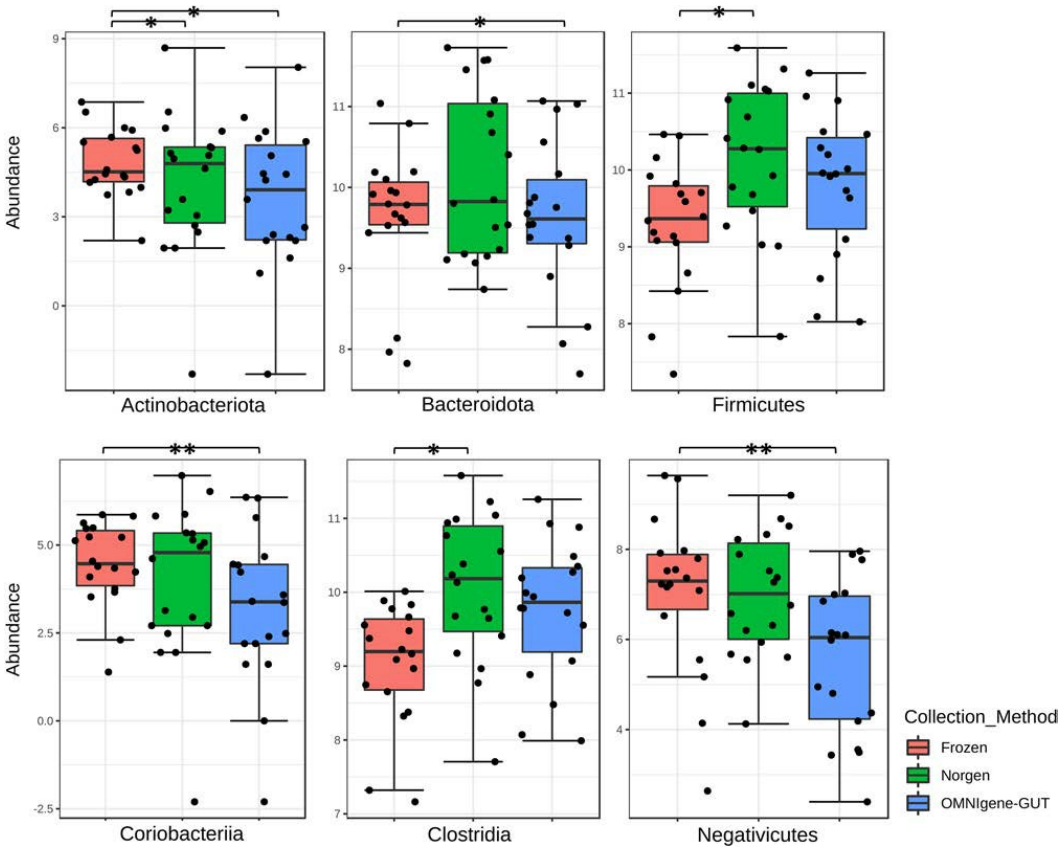


Figure 2.5. *Bacterial phyla with significantly different abundance due to collection method. Groups identified by * are significantly different prior to FDR correction and ** after FDR correction.*

Stool form according to the Bristol Stool Form Scale (BSFS) was also recorded during sample processing, and most individuals (4 of 6) did not have uniform stool types collected over the three time points. SCFA concentrations clustered in a PCA according to BSFS, and when the SCFA data was integrated with 16S ASV data using rCCA to maximise correlation, the resulting correlation coefficients also grouped loosely according to stool type (Supplementary Figure 2.2).

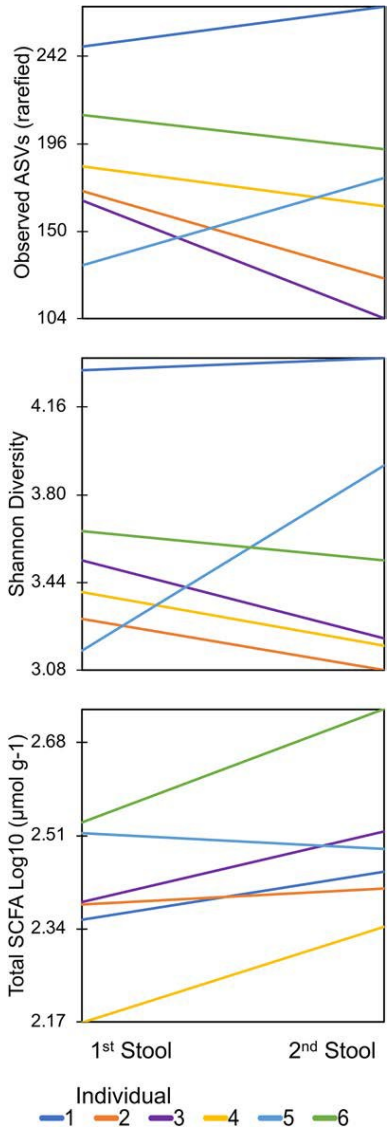


Figure 2.6. *Bacteriome diversity estimates and short chain fatty acid concentration for two consecutive stool collections. Time between stool collections was 8, 2, 4, 10, 4, and 25 h rounded to the nearest whole hour for individuals 1–6 respectively.*

2.6 Discussion

Analysis of the fecal microbiome is now commonly complemented by an additional analysis of microbial metabolites such as SCFA. To ensure these data can be represented together without the impact of spatial and temporal variability of the fecal material, collection and storage methods for stool samples must be considered. Our results found sporadic detection of low abundance bacterial and fungal species in unhomogenized stool. Further, SCFA concentrations were also shown to vary considerably across a single stool.

The level of variability (CV) in microbial diversity and SCFA concentration across a single stool, was com- pared per participant to the variability across three separate bowel movements. It was expected that temporal shifts in community structure over three timepoints would be larger than replicate sampling from a single stool. While Shannon diversity was more variable for five of six individuals among whole stool samples, Richness based on Chao1 was more variable along a single stool for four of six individuals. As well, SCFA concentrations were more variable within a single stool than across three separate bowel movements for two individuals. When the bacterial and SCFA data was integrated using rCCA, the intraindividual variability between the aliquots was also evident. Clustering of subsamples from individuals 2, 4, and 5 were very tight, indicating little difference in com- munity structure due to sampling method. Although, samples from subjects 1, 3, and 6 were less tightly clustered, implying community structure changes along the surface of the stool in these individuals that are sensitive to sampling method. This demonstrates that for some individuals, heterogeneity of microorganisms and microbial metabolites in stool may be as great as that observed over the course of 2 days, which will become apparent if samples are collected by sub-sampling a small volume of stool. This is consistent with reports of heterogeneity in mucosal bacteriome (Hong *et al.*, 2011), fecal microbiome (Wesolowska-Andersen *et al.*, 2014; Gorzelak *et al.*, 2015), and metabolite concentrations (Gratton *et al.*, 2016). As fecal material moves through the colon, the exterior surface is exposed to the mucus layer secreted by epithelial cells. This mucus (which is a niche for commensal microbes) accumulates in fecal material, and has been proposed as a mechanism for the patchy recovery of microbial species along the surface of stool (Donaldson, Lee and Mazmanian, 2015).

While our results show the surface of the stool may have more variability in richness and diversity, the β-diversity of bacterial communities between individuals remained significantly different, indicating that relative compositional differences due to subsampling are less pronounced than differences between individuals. This is consistent with similar work, where β-diversity (weighted UniFrac) was compared across nine stool subsampling locations with no significant differences observed (Liang *et al.*, 2020). Fungal communities however did not seem to be structured according to the individual to the same extent as bacterial communities, but was structured in one of two ways: Dominated by *Saccharomyces cerevisiae* (≥99%), or by hosting a more even abundance of genera including *S. cerevisiae*, and either *Kazachstania servazzii* and *Cyberlindnera jadinii*, or *Hanseniapora uvarum* and *Torulaspora delbrueckii*. In another study targeting fungi in the gut using the ITS1 region, three main mycobiome types were found: either dominated by *Candida albicans*, or *Saccharomyces cerevisiae*, or multi species type (Motooka *et al.*, 2017).

In the present study, *Candida* spp were not found at greater than 1% of the total community in any individual. However, *Candida apicola* was also not identified in the fungal mock community which could indicate an unknown technical bias against this group, although presumably not due to primer bias as low abundance was detected and this primer set has been used successfully for other Candida species (Heisel *et al.*, 2015).

Metabarcoding based microbiome investigations are inherently biased as the technique is limited by its semi- targeted design and the compositionality imposed by the NGS technology it uses. This fact is well known in the research community and thus it is encouraged to take corrective steps to reduce any distortion of the "true composition" of the microbial community by unintentional preferences in the workflow for some taxa over others (McLaren, Willis and Callahan, 2019). Bias arises at every stage of the microbiome workflow and has recently been recognized as multiplicative through to bioinformatics, although the largest contributors to this bias are upstream steps such as DNA extraction and PCR amplification (McLaren, Willis and Callahan, 2019)(Brooks *et al.*, 2015). The mock communities included in this study indicate that both *Bacillus* sp., and Clostridium sp., who are active members of the human gut microbiome, seemed to be preferentially targeted by amplification and sequencing. On the other hand, the proportion of *Rhodobacter sphaeroides*, and *Helicobacter pylori* were suppressed, and *Cutibacterium acnes* was not detected. However, suppression of *C. acnes* and *R. sphaeroides* is less concerning given they are not members of gastrointestinal community. It is also important to acknowledge that the sequence data presented here does not represent the actual number of DNA molecules recovered from the stool samples; and is limited by the capacity of the sequencing process. Therefore, the number of reads per sample, or read depth may impact the calculation of β-diversity indices by inflating the between sample diversity of samples with fewer reads (Gloor *et al.*, 2017). Despite the general move in the field towards accepting that microbiome is compositional, the question of compositionality being driven by NGS or microbiome versus the count origin of microbiome data remains a topic of discussion (Jeganathan and Holmes, 2021). Other work on the topic of bias and data correction states that the sensitivity of a β-diversity measure to sequencing effort can also be impacted by the thresholds used to remove rare species (McMurdie and Holmes, 2014), the data normalization approach, and the presence of samples with fewer than approximately 1000 sequences (Weiss *et al.*, 2017). In this work the widely used Bray–Curtis dissimilarity index was used as a distance measure to illustrate community differences between subjects and collection methods; this enabled us to directly compare our results with prior studies addressing the topic of sample collection. However, this distance measure may not always be the most reliable approach for compositional data with the characteristics previously described (Gloor *et al.*, 2017).

The microbiome is often scrutinized for small community changes in association with host-related biological factors such as diet or disease. These changes in microbial signatures are often detected in less abundant taxa, or only within particular groups of bacteria and can vary among individuals. Most bacterial ASVs with large differential abundance were found to be enriched in whole stool compared to surface aliquots, and all but one Alphaproteobacteria were classified as Clostridia. The Internal regions of stool have previously been shown to harbor significantly higher abundances of Firmicutes and *Bifiobacteria* spp compared to the external surface (Gorzelak *et al.*, 2015). In this study, the external surface of stool was likely targeted by surface aliquot collections, rather than the internal regions of stool, and if the internal regions of stool harbor larger abundances of Firmicutes, this might explain some of the differences seen between the surface aliquots and the homogenized whole stool. On the other hand, half of all fungi with large differential abundance were found to be reduced in whole stool compared to the surface aliquots; and of these all but one Dothideomycetes were classified as Saccharomycetes, indicating Saccharomycetes may be a mucosal associated fungus in the gut.

The long-term view of the healthy human gut microbiome seems to show a dynamic community which retains prolonged stability, but is punctuated by periods of disturbance (Voigt *et al.*, 2015; Fu *et al.*, 2019) On a shorter timescale, diet has been shown to cause fluctuations in microbial species (David *et al.*, 2014), as well as SCFA concentrations (McOrist *et al.*, 2011). Furthermore, the microbiome shift caused by daily food choices is highly personal, meaning the same food will elicit a unique response in each individual (Johnson *et al.*, 2019). What microbiome shift may look like across consecutive stools has not been previously explored. While only a small proportion of women defecate more than once a day, defecation frequency is known to be higher in men (Heaton *et al.*, 1992), and positively associated with vigorous physical activity, as well as plant based or high fiber diets (Sanjoaquin *et al.*, 2004). Therefore, the time of day that samples are collected may need to be indicated in sample collection protocols provided to participants. In this study, all women claimed to regularly defecate more than once a day, and the second stool of the day (collected on average 8 hours after) had significantly higher total SCFA concentrations, which seemed to be driven by significantly higher concentrations of acetic acid. The second stool also tended to have higher butyric acid concentrations, lower bacterial richness and lower Shannon diversity index compared to the first stool, although these differences were not significant. Similarly, a recent study assessing the microbiome and SCFA concentrations at a single timepoint in 441 adults found that lower bacterial diversity was associated with higher SCFA concentrations (De la Cuesta-Zuluaga *et al.*, 2018). It has been proposed that the gut microbiome has a certain level of volatility which may increase during times of stress (Bastiaanssen *et al.*, 2021), and the level of temporal variance between the two constitutive stools may indicate a normal level of volatility in the microbiome of each

individual. Another interesting observation was the similar trend in increase in butyrate producing Lachnospiraceae, and increased butyrate concentrations in the second bowel movement. The association between bacteria and SCFA concentration seen in this study also supports the idea that bacterial metabolites are linked to the circadian clock (Murakami and Tognini, 2020). Each participant collected the first bowel movement of the day in the morning, followed by the very next bowel movement; and as each woman claimed to regularly defecate at least twice per day, the natural volatility of the microbiome that seems to be linked to the circadian clock demonstrates why time of stool sample collection may be particularly important in individuals who defecate more than once per day.

Decreasing bacterial richness has also been found to correlate with decreasing stool firmness, or a higher Bristol Stool form value, based on fecal samples from 53 women (Vandeputte *et al.*, 2016). As well, the BSFS has also been shown to be a good predictor of whole-gut transit time, with high stool form values correlating to longer transit times (Heaton *et al.*, 1992; Degen and Phillips, 1996; Lewis and Heaton, 1997). A more recent study also found when stool form had a Bristol Stool value of less than three it was correlated with greater transit times, indicating that stool form can help predict whole-gut or colonic transit times (Saad *et al.*, 2010). While this study had a small sample size, it was interesting to note that both SCFA and bacterial phylogenetic diversity grouped according to stool form, and when these data were integrated through rCCA this trend was also observed. The link between transit time, microbial composition and SCFA concentration has been examined in an in-vitro system (Environmental Control System for Intestinal Microbiota). Here, it was shown that reducing the transit time from 48 to 96 h caused a significant decrease in Shannon diversity, as well as an increase in total SCFA concentration (Tottey *et al.*, 2017), as demonstrated in our study. Quantitative shifts in metabolic analysis between retention times also indicated a metabolic shift in the microbial communities. If microbial diversity and SCFA concentration are also linked to stool form and potentially transit time, assessing stool form at the point of sample processing could be a simple way to add valuable information to downstream multivariate analysis, and help explain sample clustering. Further, to reduce within-day variability that could potentially distort a long-term study, participants could be instructed to collect at a similar time, such as the first bowel movement of the day.

Directly freezing stool samples with no additional solution is considered the gold standard method for storing stool, while Norgen and in OMNIgeneGUT tubes offer a convenient method of collecting fecal material from remote participants. Regardless of collection method, all whole stool samples were dominated by Bacteroidaceae, but the second most abundant family Ruminococcaceae were significantly expanded in samples collected with both the O and N methods compared the F method, indicating that the two preservation methods may impact fecal microbiomes in a similar way. As expected,

bacterial β-diversity was mostly driven by inter-individual differences, and this is consistent with previous work where OMNIgene·Gut kits were compared to immediately frozen stool samples (Wang *et al.*, 2018). However, unlike Wang (2018) where no significant differences in bacterial β-diversity between these two methods was observed, our data shows significant differences between the three collection methods. The most obvious difference was between the directly frozen samples compared to either of the two other preservation methods (collected at room temperature), and this was observed consistently in all three bowel movements per participant.

Two previous studies have also compared fecal bacterial communities collected using OMNIgeneGUT kits which were frozen prior to processing with samples which were immediately frozen. One study found storage methods, contributed to the significant differences between samples based on Bray-Curtis dissimilarity measure, and that those collected in OMNIgeneGUT kits had a significant increase in Proteobacteria (Choo, Leong and Rogers, 2015); while another study found that samples stored in OMNIgeneGUT tubes resulted in microbiome profiles with decreased Actinobacteria and increased Lenthisphaerae compared to those that were frozen without stabilization (Penington *et al.*, 2018). Within our study, the preservation tubes were kept at room temperature – in accordance with manufacturer's instructions – and at the phyla level Actinobacteria were also reduced in fecal samples collected with both the O and N methods. It is more likely then, that the reduction in Actinobacteria is a result of storing in a preservation liquid, rather than the storage temperature.

Stool sample collection methods must not sacrifice sample "viability" for convenience, therefore, where possible we recommend collecting stool in bulk and freezing immediately. As well, during sample processing technicians can record the stool form according to BSFS, and homogenize the entire sample prior to subsampling for analysis. This method eliminates any subsampling bias due to heterogenous distribution of microbes in stool, and provides enough material for multiple assays. Additionally, because this method is less hands-on for participants, it may increase compliance if multiple collections are required. For studies where it is not possible to store a large quantity of bulk stool or where frozen transportation of stool is not viable, commercial preservation tubes may be an attractive alternative. In this circumstance it is recommended to only use a single tube type and insure a standard protocol. Furthermore, if OMNIgeneGUT or Norgen collection kits are used, researchers should be cautious in interpreting the reduced abundance of Firmicutes and Actinobacteria. Lastly, collection protocols should consider that some individuals can regularly have more than one bowel movement per day, and those participants should be instructed, where practical, to collect stool at a similar time.

2.7 Methods

2.7.1 Study design

Six healthy female volunteers, aged 25–40 years, who had not taken antibiotics in the last 3 months or probiotics in the last month prior to recruitment into this study provided fecal samples with written informed consent. The study protocol was approved by the Human Research Ethics Committee (HRE2018- 0791) from Curtin University, Western Australia, and methods were performed in accordance with the relevant guidelines and regulations. Each participant collected three fecal samples at two time points using the provided fecal sample collection kit. All stools were collected at the participants home and frozen immediately (−20 °C) in a portable freezer. Collection at the first time point (collection 1) required collecting one complete bowel movement, and from this stool collecting three small aliquots in the provided collection tubes. At the second time point, two consecutive bowel movements were collected individually (collection 2 and collection 3), with collection 2 preceding collection 3 (Figure 2.7). Once the collection was complete, the freezer was transported to Curtin University and the stool was transferred to a −80 °C freezer upon arrival.

2.7.2 Sample preparation

All stool samples were thawed at 4 °C, and transferred on ice to a EuroClone Biological safety cabinet to limit potential contamination. To assess variability between aliquots collected at home, each of the small aliquots were individually homogenized for 30 s with a sterile plastic scoop, and stool (0.25 g) was collected into separate tubes for each of two downstream analyses (metabarcoding, and SCFA quantification). The remaining stool from the initial three aliquots was combined and manually homogenized together for 30 s with a sterile plastic scoop, and collected again for two separate analyses. All samples were immediately frozen to −80 °C. Prior to preparation, whole stool samples were ranked on the Bristol Stool form chart. To assess collection methods, from each unhomogenized stool, feces were collected into one OMNIgene·Gut tube (DNA Genotec) (collection method O) and one Stool Nucleic Acid Collection and preservation Tube (NORGEN) (collection method N), and were stored at room temperature for 12 days. The remaining stool from each sample was individually homogenized while within the plastic collection bag for 1 minute, and then subsequent aliquots of stool (0.25 g) were collected for each of the three analyses and immediately frozen to −80°C (collection method F).

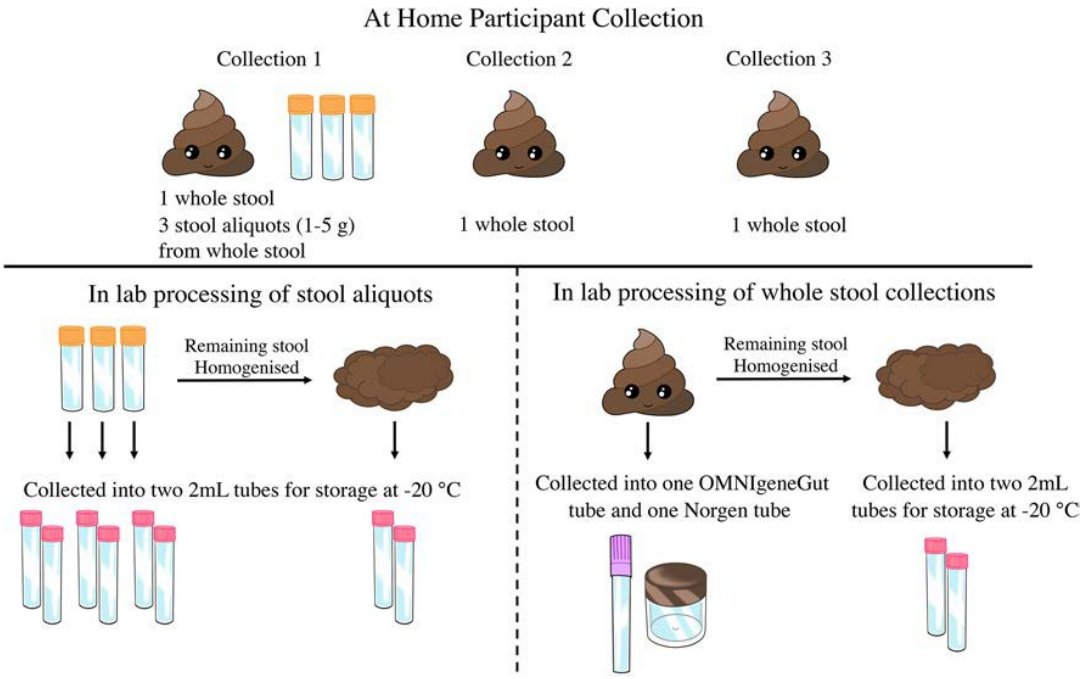


Figure 2.7. *Sample collection by participants at home (top panel), and subsequent sample processing in laboratory (bottom panel). The first bowel movement of the day was obtained for collection 1 and 2, and the very next bowel movement after collection 2 was obtained for collection 3.*

2.7.3 Short chain fatty acid quantification

Homogenized fecal material (0.50 ± 0.05 g) was transported on dry ice to the Science Analytical Facility at Edith Cowan University, Western Australia. Here SCFA were extracted and quantified as previously described (Zhao, Nyman and Jönsson, 2006). Briefly, an aqueous stock solution of standards containing acetic, propionic, iso-butyric, butyric, iso-valeric, valeric, and hexanoic acids was diluted to four levels, and used for analysis. SCFA were extracted using a solution of hydrochloric acid, methanol, ultrapure water, and 2-ethyl butyric acid which was used as an internal standard. The mixture was vortexed for 1 min, and then incubated at 4 °C for 1 h, and then vortexed a second time for 1 min. Finally, the solution was Centrifuge at 12000 rpm at 4 °C for 20 min, and the supernatant retained and stored at 4 °C for no more than 48 h prior to analysis on Thermo Scientific GC–MS (ISQ) using a Thermo Scientific TG-Wax column (30 m × 0.25 mm × 0.25 µm), and a seven- point calibration. A sample volume of 1.0 µL was injected with an inlet temperature of 220 °C, using Helium carrier gas (1.0 mL/min). The total run time was 18 min.

2.7.4 Fecal DNA extraction

Immediately prior to DNA extraction, frozen stool samples were thawed on ice, and stool samples stored in preservation tubes were shaken by hand for 10 sec. DNA was

extracted by using QIAamp Power Fecal DNA kit (QIAGEN, Hilden, Germany) using QIAamp Power Fecal DNA IRT protocol for QIAcube (QIAGEN), as well as OMNIgeneGUT microbial DNA purification protocol using QIAGEN QIAamp PowerFecal DNA kit, both according to the manufacturer's instructions with one modification at step 3 of the IRT protocol, tubes were vortexed for 20 sec to incorporate beads and stool prior to heating. Extraction controls were also processed following the same protocol as frozen stool samples.

2.7.5 Bacterial and fungal library preparation and sequencing

The V4 region of bacterial DNA and mock communities (ATCC MSA-1002 and MSA-1003) were targeted and amplified using 16S primers 515F (Turner *et al.*, 1999) and 806R (Caporaso, Lauber, Walters, *et al.*, 2011), each with a 6-8 bp unique barcode. The PCR reactions contained of 1x PCR buffer (Applied Biosystems), 2mM $MgCl_2$ (Applied Biosystems), 0.25nM dNTP (Bioline), 0.4mg/mL BSA (Thermo Fisher Scientific) 0.4 μM primer (Integrated DNA Technologies), 0.12x SYBER, and 2U AmpliTaq Gold™ DNA polymerase (Thermo Fisher Scientific). Reactions contained 2 μL of template DNA which was previously screened and optimized for efficiency by qPCR (Murray, Coghlan and Bunce, 2015), and had a final volume of 25 μL. Fungal DNA and mock communities (Bakker, 2018) were amplified using ITS2 primers FSeq and RSeq (Heisel *et al.*, 2015). PCR reactions were the same as for bacterial amplification except 3μL of template DNA was added to each reaction.

The reactions for both bacterial and fungal amplicons were performed on StepOnePlus Real-Time PCR system (Applied Biosystems), and under the following conditions for bacterial amplicons: denaturing at 94ºC for 3 min, followed by 30 cycles of 94ºC for 40 sec, annealing at 53ºC for 40 sec, and extension at 72ºC for 60 sec. The cycling program for fungi was as follows: denaturing at 94ºC for 3 min, followed by 35 cycles of 94ºC for 40 sec, 55 ºC for 40 sec, 72ºC for 80 sec. Both amplicons underwent a final extension at 72ºC for 10 minutes. Individual Bacterial and Fungal libraries were prepared by blending together in equimolar concentrations. Illumina compatible adaptors were ligated to the DNA fragments (Lucigen, Middelton, WI, USA), and the resulting amplicons were size selected using Pippin Prep (Sage Science). The QIAquick PCR purification column clean up kit (Qiagen, Germantown, MD), was used to purify the DNA library before sequencing, which was performed at Curtin University, Western Australia, using the Illumina MiSeq platform and V2 500 cycle kit (Illumina, San Diego, CA, USA) with 2 × 250 bp paired-end read lengths.

2.7.6 Deconvolution

Unique molecular barcodes were used to demultiplex reads with no mismatches allowed. Cutadapt (Martin, 2011) was used to remove primers, and the remaining reads were quality filtered, trimmed, and merged using DADA2 (Callahan *et al.*, 2016). Reads with ambiguous bases, or with more than two expected errors were discarded. Amplicon sequence variants (ASVs) were inferred from the reads using the pseudo-pooled method, and merged with a minimum overlap of 60bp allowing for one mismatch (16S V4), 30bp with no mismatches (ITS2). Amplicons were retained at a minimum length of 150, and 251, base pairs for ITS2, and 16S V4, datasets respectively. Chimera errors were also removed with DADA2 using the default method. Classification for 16S sequence variants was performed using the Genome Taxonomy reference database (release 95) formatted for use with DADA2 (https://zenodo.org/record/3951383#.X7Hs49sRVTY), while the UNITE general FASTA release for fungi Version 18.11.2018 (Kõljalg *et al.*, 2005) was used for ITS2 sequence classification, each with a minimum of 50% bootstrapping. Contamination was removed from all sequences with one run of the function remove.count in microDecon (McKnight *et al.*, 2019). Any ASVs with unassigned phylum, or with a prevalence less than 1 in 5% of samples were filtered out, as were fungal samples with less than 1000 reads.

2.7.7 Statistical analysis

Sequence counts were used to determine richness and α-diversity indices (Chao1, Shannon, and Faith's phylogenetic diversity (PD)) for bacterial microbiomes as applied in the "Phyloseq" package (McMurdie and Holmes, 2013) run in R studio with R version 3.6.3 (R Core Team, 2020). Correlation between library size and diversity estimate were tested for, and α-diversity measures with significant Pearson correlation ($p < 0.01$) to reads per sample were rarefied to lowest sample depth prior to calculation for those α-diversity measures (Chao1 and Faith's PD). β-Diversity was compared between collection methods with PERMANOVA in PRIMER-e v7 (Anderson, Gorley and Clarke, 2008)and visualized using PCoA using Euclidian distances of center log ratio (CLR) transformed data, as well as Bray–Curtis similarity calculated from 4th root transformed proportions of counts. SCFA concentration data were log10 transformed, and normality assumed using the Shapiro–Wilk test prior to paired t-test.

To evaluate differentially abundant taxa between homogenization method (aliquots and whole stool), the effect size estimate as a log2 fold change was calculated in DESeq2 statistical package (Love, Huber and Anders, 2014) with a Benjamini-Hochburg adjustment for multiple testing, and a design controlling for subject. Statistical differences between taxa abundance and community diversity due to homogenization methods and collection methods were further tested using ANOVA (MicrobiomeAnalyst) and PERMANOVA (PRIMER7) respectively. Projection to latent

structures discriminant analysis was used to test for differences in microbiome composition due to collection method; and regularize canonical correlation analysis (integrated to maximize correlation between latent variables) was used to integrate SCFA and bacteriome data, both in MixOmics (Rohart *et al.*, 2017).

Acknowledgments

We are grateful to all participants of this study. We would also like to extend our thanks to Johney Lo for his assistance with statistical analysis, Sinéad Jones for her artwork used in Figure 7, and Rose Lines for her technical assistance with laboratory work.

Author Contributions

C.T.C, S.R, J.J and D.P contributed to study concept. J.J recruited participants and performed all laboratory work. A.A and J.J participated in data processing, and J.J performed all analyses and drafted the manuscript. All authors contributed to reviewing the manuscript and approved the final manuscript.

Additional Information

Competing Interests: The authors declare no competing interests.

2.8 References

Anderson, M. J., Gorley, R. N. and Clarke, K. R. (2008) 'PERMANOVA+ for PRIMER: Guide to Software and Statistical Methods'. Plymouth, UK: PRIMER-E.

Arrieta, M. C. *et al.* (2018) 'Associations between infant fungal and bacterial dysbiosis and childhood atopic wheeze in a nonindustrialized setting', *Journal of Allergy and Clinical Immunology*. Mosby, 142(2), pp. 424-434.e10. doi: 10.1016/j.jaci.2017.08.041.

Bakker, M. G. (2018) 'A fungal mock community control for amplicon sequencing experiments', *Molecular Ecology Resources*. John Wiley & Sons, Ltd (10.1111), 18(3), pp. 541–556. doi: 10.1111/1755-0998.12760.

Bastiaanssen, T. F. S. *et al.* (2021) 'Volatility as a Concept to Understand the Impact of Stress on the Microbiome', *Psychoneuroendocrinology*. Elsevier Ltd, 124. doi: 10.1016/j.psyneuen.2020.105047.

Brooks, J. P. *et al.* (2015) 'The truth about metagenomics: Quantifying and counteracting bias in 16S rRNA studies Ecological and evolutionary microbiology', *BMC Microbiology*. BioMed Central Ltd., 15(1), p. 66. doi: 10.1186/s12866-015-0351-6.

Callahan, B. J. *et al.* (2016) 'DADA2: High-resolution sample inference from Illumina amplicon data', *Nature Methods*. Nature Publishing Group, 13(7), pp. 581–583. doi: 10.1038/nmeth.3869.

Caporaso, J. G., Lauber, C. L., Walters, W. A., *et al.* (2011) 'Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 108(1), pp. 4516–4522. doi: 10.1073/pnas.1000080107.

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

Choo, J. M., Leong, L. E. and Rogers, G. B. (2015) 'Sample storage conditions significantly influence faecal microbiome profiles', *Scientific Reports*. Nature Publishing Group, 5(1), p. 16350. doi: 10.1038/srep16350.

Couch, R. D. *et al.* (2013) 'The Approach to Sample Acquisition and Its Impact on the Derived Human Fecal Microbiome and VOC Metabolome', *PLoS ONE*. Edited by H. Ashktorab. Public Library of Science, 8(11), p. e81163. doi: 10.1371/journal.pone.0081163.

D'Amore, R. *et al.* (2016) 'A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling', *BMC Genomics*. BioMed Central, 17(1), p. 55. doi: 10.1186/s12864-015-2194-9.

David, L. A. *et al.* (2014) 'Diet rapidly and reproducibly alters the human gut microbiome', *Nature*. Nature Publishing Group, 505(7484), pp. 559–563. doi: 10.1038/nature12820.

Degen, L. P. and Phillips, S. F. (1996) 'How well does stool form reflect colonic transit?', *Gut*. BMJ Publishing Group, 39(1), pp. 109–113. doi: 10.1136/gut.39.1.109.

Donaldson, G. P., Lee, S. M. and Mazmanian, S. K. (2015) 'Gut biogeography of the bacterial microbiota', *Nature Reviews Microbiology*. NIH Public Access, pp. 20–32. doi: 10.1038/nrmicro3552.

Fieten, K. B. *et al.* (2018) 'Fecal microbiome and food allergy in pediatric atopic dermatitis: A cross-sectional pilot study', *International Archives of Allergy and Immunology*. Karger Publishers, 175(1–2), pp. 77–84. doi: 10.1159/000484897.

De Filippis, F. *et al.* (2016) 'High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome', *Gut*. BMJ Publishing Group, 65(11), pp. 1812–1821. doi: 10.1136/gutjnl-2015-309957.

Flores, G. E. *et al.* (2014) 'Temporal variability is a personalized feature of the human microbiome', *Genome biology*. BioMed Central, 15(12), p. 531. doi: 10.1186/s13059-014-0531-y.

Fu, B. C. *et al.* (2019) 'Temporal variability and stability of the fecal microbiome: The multiethnic cohort study', *Cancer Epidemiology Biomarkers and Prevention*. American Association for Cancer Research Inc., 28(1), pp. 154–162. doi: 10.1158/1055-9965.EPI-18-0348.

Gloor, G. B. *et al.* (2017) 'Microbiome datasets are compositional: And this is not optional', *Frontiers in Microbiology*. Frontiers, p. 2224. doi: 10.3389/fmicb.2017.02224.

Gorzelak, M. A. *et al.* (2015) 'Methods for improving human gut microbiome data by reducing variability through sample processing and storage of stool', *PLoS ONE*, 10(8). doi: 10.1371/journal.pone.0134802.

Gratton, J. *et al.* (2016) 'Optimized Sample Handling Strategy for Metabolic Profiling of Human Feces', *Analytical Chemistry*, 88(9), pp. 4661–4668. doi: 10.1021/acs.analchem.5b04159.

Heaton, K. W. *et al.* (1992) ‘Defecation frequency and timing, and stool form in the general population: A prospective study’, *Gut*, 33(6), pp. 818–824. doi: 10.1136/gut.33.6.818.

Heisel, T. *et al.* (2015) ‘Complementary amplicon-based genomic approaches for the study of fungal communities in humans’, *PLoS ONE*. Edited by M. D. Lenardon. Public Library of Science, 10(2), p. e0116705. doi: 10.1371/journal.pone.0116705.

Hong, P.-Y. *et al.* (2011) ‘Pyrosequencing-Based Analysis of the Mucosal Microbiota in Healthy Individuals Reveals Ubiquitous Bacterial Groups and Micro-Heterogeneity’, *PLoS ONE*. Edited by N. Ahmed. Public Library of Science, 6(9), p. e25042. doi: 10.1371/journal.pone.0025042.

Jeganathan, P. and Holmes, S. P. (2021) ‘A Statistical Perspective on the Challenges in Molecular Microbial Biology’, *Journal of Agricultural, Biological, and Environmental Statistics*. Springer, pp. 131–160. doi: 10.1007/s13253-021-00447-1.

Johnson, A. J. *et al.* (2019) ‘Daily Sampling Reveals Personalized Diet-Microbiome Associations in Humans’, *Cell Host and Microbe*. Cell Press, 25(6), pp. 789-802.e5. doi: 10.1016/j.chom.2019.05.005.

Karu, N. *et al.* (2018) ‘A review on human fecal metabolomics: Methods, applications and the human fecal metabolome database’, *Analytica Chimica Acta*. Elsevier, pp. 1–24. doi: 10.1016/j.aca.2018.05.031.

Kõljalg, U. *et al.* (2005) ‘UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi’, *New Phytologist*. John Wiley & Sons, Ltd (10.1111), 166(3), pp. 1063–1068. doi: 10.1111/j.1469-8137.2005.01376.x.

De la Cuesta-Zuluaga, J. *et al.* (2018) ‘Higher Fecal Short-Chain Fatty Acid Levels Are Associated with Gut Microbiome Dysbiosis, Obesity, Hypertension and Cardiometabolic Disease Risk Factors’, *Nutrients*. MDPI AG, 11(1), p. 51. doi: 10.3390/nu11010051.

Lagkouvardos, I. *et al.* (2015) ‘Gut metabolites and bacterial community networks during a pilot intervention study with flaxseeds in healthy adult men’, *Molecular Nutrition and Food Research*. Wiley-Blackwell, 59(8), pp. 1614–1628. doi: 10.1002/mnfr.201500125.

Lewis, S. J. and Heaton, K. W. (1997) ‘Stool form scale as a useful guide to intestinal transit time’, *Scandinavian Journal of Gastroenterology*. Informa Healthcare, 32(9), pp. 920–924. doi: 10.3109/00365529709011203.

Liang, Y. *et al.* (2020) 'Systematic Analysis of Impact of Sampling Regions and Storage Methods on Fecal Gut Microbiome and Metabolome Profiles', *mSphere*. American Society for Microbiology, 5(1). doi: 10.1128/msphere.00763-19.

Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*. BioMed Central Ltd., 15(12), p. 550. doi: 10.1186/s13059-014-0550-8.

Martin, M. (2011) 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet.journal*. EMBnet Stichting, 17(1), p. 10. doi: 10.14806/ej.17.1.200.

McKnight, D. T. *et al.* (2019) 'microDecon: A highly accurate read-subtraction tool for the post-sequencing removal of contamination in metabarcoding studies', *Environmental DNA*. John Wiley & Sons, Ltd, 1(1), pp. 14–25. doi: 10.1002/edn3.11.

McLaren, M. R., Willis, A. D. and Callahan, B. J. (2019) 'Consistent and correctable bias in metagenomic sequencing experiments', *eLife*, 8, p. e46923. doi: 10.7554/eLife.46923.

McMurdie, P. J. and Holmes, S. (2013) 'phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data', *PLoS ONE*. Edited by M. Watson. Public Library of Science, 8(4), p. e61217. doi: 10.1371/journal.pone.0061217.

McMurdie, P. J. and Holmes, S. (2014) 'Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible', *PLoS Computational Biology*. Public Library of Science, 10(4). doi: 10.1371/journal.pcbi.1003531.

McOrist, A. L. *et al.* (2011) 'Fecal Butyrate Levels Vary Widely among Individuals but Are Usually Increased by a Diet High in Resistant Starch', *The Journal of Nutrition*, 141(5), pp. 883–889. doi: 10.3945/jn.110.128504.

Motooka, D. *et al.* (2017) 'Fungal ITS1 deep-sequencing strategies to reconstruct the composition of a 26-species community and evaluation of the gut mycobiota of healthy Japanese individuals', *Frontiers in Microbiology*. Frontiers, 8(FEB), p. 238. doi: 10.3389/fmicb.2017.00238.

Murakami, M. and Tognini, P. (2020) 'The Circadian Clock as an Essential Molecular Link Between Host Physiology and Microorganisms', *Frontiers in Cellular and Infection Microbiology*. Frontiers Media S.A., p. 469. doi: 10.3389/fcimb.2019.00469.

Murray, D. C., Coghlan, M. L. and Bunce, M. (2015) 'From Benchtop to Desktop: Important Considerations when Designing Amplicon Sequencing Workflows', *PLOS ONE*. Edited by T. Gilbert. Public Library of Science, 10(4), p. e0124671. doi: 10.1371/journal.pone.0124671.

Penington, J. S. *et al.* (2018) 'Influence of fecal collection conditions and 16S rRNA gene sequencing at two centers on human gut microbiota analysis', *Scientific Reports*, 8(1). doi: 10.1038/s41598-018-22491-7.

R Core Team (2020) 'R: A Language and Environment for Statistical Computing'. Vienna, Austria: R Foundation for Statistical Computing.

Rohart, F. *et al.* (2017) 'mixOmics: An R package for 'omics feature selection and multiple data integration', *PLOS Computational Biology*. Edited by D. Schneidman. Public Library of Science, 13(11), p. e1005752. doi: 10.1371/journal.pcbi.1005752.

Saad, R. J. *et al.* (2010) 'Do stool form and frequency correlate with whole-gut and colonic transit results from a multicenter study in constipated individuals and healthy controls', *American Journal of Gastroenterology*, 105(2), pp. 403–411. doi: 10.1038/ajg.2009.612.

Sanjoaquin, M. A. *et al.* (2004) 'Nutrition and lifestyle in relation to bowel movement frequency: a cross-sectional study of 20 630 men and women in EPIC–Oxford', *Public Health Nutrition*. Cambridge University Press (CUP), 7(1), pp. 77–83. doi: 10.1079/phn2003522.

Strati, F. *et al.* (2017) 'New evidences on the altered gut microbiota in autism spectrum disorders', *Microbiome*. BioMed Central, 5(1), p. 24. doi: 10.1186/s40168-017-0242-1.

Tang, Z. Z. *et al.* (2019) 'Multi-omic analysis of the microbiome and metabolome in healthy subjects reveals microbiome-dependent relationships between diet and metabolites', *Frontiers in Genetics*. Frontiers, 10(MAY), p. 454. doi: 10.3389/fgene.2019.00454.

Tottey, W. *et al.* (2017) 'Colonic transit time is a driven force of the gut microbiota composition and metabolism: In vitro evidence', *Journal of Neurogastroenterology and Motility*. Journal of Neurogastroenterology and Motility, 23(1), pp. 124–134. doi: 10.5056/jnm16042.

Turner, S. *et al.* (1999) 'Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis', *Journal of Eukaryotic Microbiology*. John Wiley & Sons, Ltd (10.1111), 46(4), pp. 327–338. doi: 10.1111/j.1550-7408.1999.tb04612.x.

Valles-Colomer, M. *et al.* (2019) 'The neuroactive potential of the human gut microbiota in quality of life and depression', *Nature Microbiology*. Nature Publishing Group, 4(4), pp. 623–632. doi: 10.1038/s41564-018-0337-x.

Vandeputte, D. *et al.* (2016) 'Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates', *Gut*. BMJ Publishing Group, 65(1), pp. 57–62. doi: 10.1136/gutjnl-2015-309618.

Venegas, D. P. *et al.* (2019) 'Short chain fatty acids (SCFAs)mediated gut epithelial and immune regulation and its relevance for inflammatory bowel diseases', *Frontiers in Immunology*. Frontiers Media S.A., p. 277. doi: 10.3389/fimmu.2019.00277.

Vogt, N. M. *et al.* (2017) 'Gut microbiome alterations in Alzheimer's disease', *Scientific Reports*. Nature Publishing Group, 7(1), pp. 1–11. doi: 10.1038/s41598-017-13601-y.

Voigt, A. Y. *et al.* (2015) 'Temporal and technical variability of human gut metagenomes', *Genome Biology*. BioMed Central Ltd., 16(1), p. 73. doi: 10.1186/s13059-015-0639-8.

Wang, Z. *et al.* (2018) 'Comparison of Fecal Collection Methods for Microbiome and Metabolomics Studies.', *Frontiers in cellular and infection microbiology*. Frontiers Media SA, 8, p. 301. doi: 10.3389/fcimb.2018.00301.

Weiss, S. *et al.* (2017) 'Normalization and microbial differential abundance strategies depend upon data characteristics', *Microbiome*. BioMed Central Ltd., 5(1), p. 27. doi: 10.1186/s40168-017-0237-y.

Wesolowska-Andersen, A. *et al.* (2014) 'Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis', *Microbiome*, 2(19). doi: 10.1186/2049-2618-2-19.

Zhao, G., Nyman, M. and Jönsson, J. Å. (2006) 'Rapid determination of short-chain fatty acids in colonic contents and faeces of humans and rats by acidified water-extraction and direct-injection gas chromatography', *Biomedical Chromatography*. Biomed Chromatogr, 20(8), pp. 674–682. doi: 10.1002/bmc.580.

2.9 Supplementary Material

Supplementary Table 2.1. *Sequences of each bacterial and fungal library. Filtering involved all bioinformatic processing including qualify filtering, merging, and trimming, as well as decontamination and low prevalence filtering of ASVs. Richness and diversity estimates are of a single technical replicate sequenced for each gene region*

Gene region	**16S, V4**		**ITS2**	
Sample sequences post filtering	4,303,623		8,416,581	
Average reads per sample	53,387		150,296	
Minimum read count	5,387		1,073	
Maximum read count	243,233		1,431,685	
ASVs post filtering	904		70	
Technical replicates for each gene region				
Observed ASVs	285	279	37	39
Shannon	3.55	3.63	0.3	2.4

Bacterial Mock community composition	16S V4 region			Fungal Mock community composition	ITS2 region		
	F	G	S		F	G	S
Acinetobacter baumannii				Alternaria alternata			
Actinomyces odontolyticus				Aspergillus flavus			
Bacillus cereus				Candida apicola			
Bacteroides vulgatus				Chytriomyces hyalinus			
Bifidobacterium adolescentis				Claviceps purpurea			
Clostridium beijerinckii				Fusarium graminearum / G. zeae			
Cutibacterium acnes				Fusarium oxysporum			
Deinococcus radiodurans				Fusarium verticillioides / G. moniliformis			
Enterococcus faecalis				Mortierella verticillata			
Escherichia coli				Naganishia albida / C. albidus			
Helicobacter pylori				Neosartorya fischeri			
Lactobacillus gasseri				Penicillium expansum			
Neisseria meningitidis				Rhizoctonia solani			
Porphyromonas gingivalis				Rhizomucor miehei			
Pseudomonas aeruginosa				Rhizophagus irregularis			
Rhodobacter sphaeroides				Saccharomyces cerevisiae			
Staphylococcus aureus				Saitoella complicata			
Staphylococcus epidermidis				Trichoderma reesei			
Streptococcus agalactiae				Ustilago maydis			
Streptococcus mutans							

Supplementary Figure 2.1. *Occurrence of bacterial species of the mock community successfully identified (blue), incorrect classification but resolved correctly at a higher rank (pink), or not identified (grey). Ranks are family (F), genus (G), and species (S)*

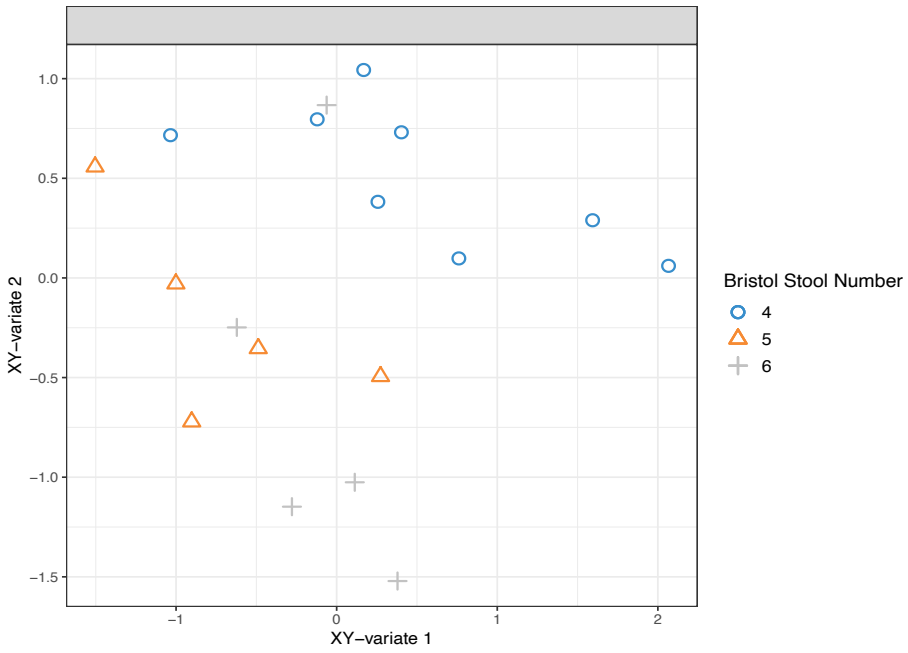
Supplementary Table 2.2. *Percent coefficient of variation of SCFA concentrations and bacterial diversity estimates among three aliquots collected from a single stool (aliquots) and from three whole stools collected from separate bowel movements (whole). Bolded values indicate the more variable measure per individual. Chao1 and Faith's phylogenetic distance was calculated on rarefied counts.*

Individual	1		2		3		4		5		6	
Aliquots	mean(SD)	CV	mean(SD)	CV	mean(SD)	CV	mean(SD)	CV	mean(SD)	CV	mean(SD)	CV
Chao1	284(27.6)	9.7	200.3(54.2)	**27**	220.3(17.3)	7.8	174.4(20.2)	**11.6**	163.1(40.2)	**24.7**	197.0(25.9)	**13.2**
ACE	330.4(36.4)	11	261.1(74.9)	**28.7**	273.4(28.3)	10.3	181.4(29.7)	**16.3**	182.6(38.3)	**21**	237.4(13.5)	5.7
Shannon	4.3(0)	1.3	3.2(0.2)	**6.5**	3.7(0)	0.9	3.3(0)	2.1	3.4(0.1)	4.9	3.4(0)	0.3
Faith's PD	26.8(1.2)	4.4	19.9(1.3)	6.6	21.5(2.3)	10.6	16.7(1.2)	**7.7**	16.7(3.4)	**20.7**	17.6(0.4)	**2.6**
acetic acid	127.5(34.7)	**27.1**	113(13)	11.5	143.9(26)	18.1	202.5(12)	5.9	229.6(23.9)	**10.4**	486.6(183.2)	**37.6**
propionic acid	28.1(4.6)	**16.4**	19.5(0.8)	4.2	24.8(2.8)	11.5	33(1.1)	3.5	58.2(3.7)	6.4	71.3(12)	16.8
butyric acid	28.4(4.2)	**14.9**	16.5(2.5)	15.3	26.7(3)	11.2	37.4(1.4)	3.8	33.6(2.3)	7	83.8(37.8)	**45.2**
valeric acid	4.7(0.8)	**19**	2.9(0.1)	3.5	2.2(0.4)	20	3.9(0.2)	5.9	2.3(0.1)	7.6	9.4(3.7)	**39.9**
total	188.8(41.5)	**22**	152.1(9.7)	6.4	197.8(32)	16.2	276.9(13.9)	5	323.8(27.6)	**8.5**	651.2(164.2)	**25.2**
Whole	mean(SD)	CV	mean(SD)	CV	mean(SD)	CV	mean(SD)	CV	mean(SD)	CV	mean(SD)	CV
Chao1	264.3(36.4)	**13.8**	178(38.2)	21.5	185.2(64.8)	**35**	197.6(19.7)	10	190.1(40.23)	21.2	244.7(3.1)	1.3
ACE	301(49.1)	**16.3**	232.9(57.2)	24.5	227(93.4)	**41.1**	232.8(12.9)	5.5	203.3(24.3)	11.9	280.2(17.3)	**6.1**
Shannon	4.2(0)	**2**	3.1(0.1)	3.3	3.4(0.1)	**5.7**	3.3(0.1)	**4.4**	3.5(0.3)	**10.6**	3.5(0)	**2.4**
Faith's PD	26.3(2.6)	**10**	19.3(3.4)	**17.8**	19.3(3.1)	**16.3**	17.9(0.9)	5.2	17.3(2)	11.8	19.8(0.4)	2.2

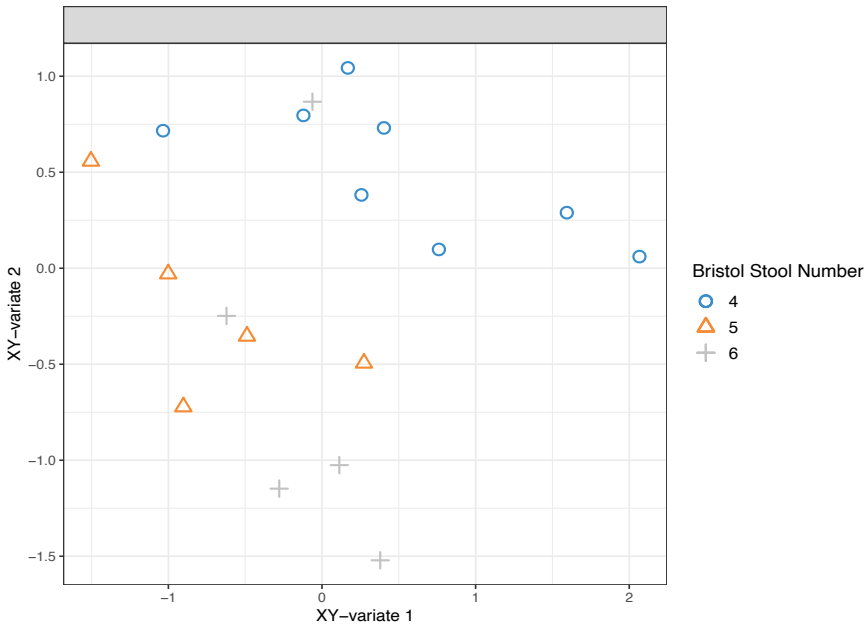
acetic acid	178.5(23.5)	13.1	171.7(40.4)	**23.5**	185.2(56.8)	**30.6**	144.8(42.3)	**29.2**	236.7(5.5)	2.3	347.8(103.4)	29.7
propionic acid	32.7(0.7)	2.2	22.9(1.2)	**5.5**	32.1(8.4)	**26.1**	26.7(9.9)	**37.2**	52.3(15.3)	**29.3**	52(16.1)	**31**
butyric acid	33.6(2.2)	6.7	27.3(7.3)	**27**	32.4(7.1)	**21.9**	38.5(11)	**28.5**	33.1(4.9)	**14.8**	72(30)	41.7
valeric acid	5.3(0.3)	7.1	2.5(0.7)	**27.7**	2.8(0.5)	**20.1**	3.4(1.2)	**35**	2.7(0.3)	**11.8**	6.6(1.3)	20.8
total	250.3(25.6)	10.2	224.6(45.6)	**20.3**	252.7(72.6)	**28.7**	213.6(63)	**29.5**	325(18.4)	5.6	478.6(118.4)	24.7

Supplementary Table 2.3. *Taxa identified in significantly different abundance in the O or N methods compared to the F method.*

	F vs N			F vs O	
	Pvalues	FDR		Pvalues	FDR
Phyla			**Phyla**		
Firmicutes	0.008	0.092	Actinobacteriota	0.013	0.160
Actinobacteriota	0.023	0.140	Bacteroidota	0.039	0.234
Class			**Class**		
Clostridia	0.003	0.059	Negativicutes	0.002	0.032
Lentisphaeria	0.031	0.229	Coriobacteriia	0.003	0.032
			Bacilli	0.030	0.185
			Bacteroidia	0.039	0.185
Order			**Order**		
Lachnospirales	0.001	0.056	Monoglobales	0.001	0.058
Oscillospirales	0.005	0.077	Coriobacteriales	0.003	0.068
Not_Assigned	0.006	0.077	Lactobacillales	0.010	0.131
Rhizobiales	0.017	0.174	Bacteroidales	0.042	0.381
Victivallales	0.031	0.250			
Monoglobales	0.041	0.254			
Family			**Family**		
Ruminococcaceae	0.001	0.041	Ruminococcaceae	0.000	0.003
Lachnospiraceae	0.001	0.041	Eggerthellaceae	0.000	0.003
Actinomycetaceae	0.002	0.047	UBA1381	0.001	0.035
Eggerthellaceae	0.008	0.138	Rikenellaceae	0.004	0.080
Oscillospiraceae	0.010	0.151	Actinomycetaceae	0.011	0.161
Rhizobiaceae	0.017	0.186	Streptococcaceae	0.015	0.184
Not_Assigned	0.018	0.186	Acutalibacteraceae	0.018	0.184
Victivallaceae	0.031	0.281	Veillonellaceae	0.025	0.221
UBA1381	0.041	0.327	Erysipelatoclostridiaceae	0.037	0.297



Supplementary Figure 2.2. *Bacterial communities and metabolites group according to stool form. SCFA and 16S ASV from three matching whole stool samples were combined using rCCA in the R package MixOmics, and plotted as correlation coefficients.*



Supplementary Figure 2.3. *Bacterial communities and metabolites group according to stool form. SCFA and 16S ASV from three matching whole stool samples were combined using rCCA in the R package MixOmics, and plotted as correlation coefficients.*

– Chapter 3 –
Changes to the gut microbiome in young children showing early behavioural signs of autism

3.1 Preamble

The gut microbiome of autistic children has been described, repeatedly showing impairments to the consortium of bacterial and fungal inhabitants, as well as SCFA profile. It is unknown weather characteristic shifts to the gut environment might be detectable prior to diagnosis; but to investigate these fine scale shifts sample and data processing must be of the highest standard. In chapter 2, I provide evidence for the way sampling bias can affect microbiome and SCFA profiles, concluding that large sample volumes should be frozen and homogenised appropriately. Stool of a sufficiently large volume was collected by from one bowel movement using a large, scooped jar, and each sample was frozen immediately. The samples were processed for two separate analyses using the protocol developed and presented in chapter 2, allowing for an analysis that would best capture the faecal environment of these young children. In line with the recommendations presented in chapter 2, a detailed description of the stool collection, microbiome library preparation, and decontamination steps were included in the manuscript. Furthermore, the use of control samples was expanded on to include a more detailed description of both the replicate sample and the mock community

In this study I describe structural and functional changes displayed by the gut microbiome that occur prior to the formal diagnosis of a disease. This research indicates that shifts in microbial structure are detectable at the earliest stages of a disease. In chapter 4, the gut microbiome of infants who are involved in a randomised control trail investigating a maternal prebiotic supplement will be described. While our aim is to look for changes that occur as a result of the maternal prebiotic supplement, the same dataset will be used to investigate changes in the microbiome that occur prior to the diagnosis of allergic severity. Chapter 3 therefore provides the preliminary framework for the way gut microbiome data can be used to investigate changes that happen prior to disease onset. Lastly, in keeping with the proposed reporting guidelines which have been tailored for microbiome research a “Strengthening The Organization and Reporting of Microbiome Studies’ (STORMS)” checklist has been completed for each data chapter of the thesis. The published manuscript described in this chapter successfully hit 47 of 55 relevant criteria (Appendix 2).

3.2 Contributions and data accessibility

A statement regarding author contributions, and a link to the raw data used in this chapter are included within the manuscript.

3.3 Manuscript

The following section of chapter 2 contains the manuscript published in Frontiers in Microbiology on July 28th 2022.
https://www.frontiersin.org/articles/10.3389/fmicb.2022.905901/full

Changes to the gut microbiome in young children showing early behavioural signs of autism

***Jacquelyn Jones[12], Stacey N Reinke[3], Mahsa Mousavi-Derazmahalleh[12], Debra J. Palmer[45], Claus T. Christophersen[1236]**

[1]Trace and Environmental DNA Laboratory, School of Molecular and Life Sciences, Curtin University, Bentley, WA, Australia
[2]The Western Australian Human Microbiome Collaboration Centre, Curtin University, Bentley, WA, Australia
[3]Centre for Integrative Metabolomics and Computational Biology, School of Science, Edith Cowan University, Joondalup, WA, Australia
[4]Telethon Kids Institute, University of Western Australia, Nedlands, WA, Australia
[5]School of Medicine, University of Western Australia, Crawley, WA, Australia
[6]School of Medical & Health Sciences, Edith Cowan University, Joondalup, WA, Australia

Keywords: microbiome[1], gut-brain-microbiota axis[2], autism spectrum disorder [3], short chain fatty acid [4], stool form [5].

The human gut microbiome has increasingly been associated with autism spectrum disorder (ASD), which is a neurological developmental disorder, characterized by impairments to social interaction. The ability of the gut microbiota to signal across the gut- brain-microbiota axis with metabolites, including short-chain fatty acids, impacts brain health and has been identified to play a role in the gastrointestinal and developmental symptoms affecting autistic children. The fecal microbiome of older children with ASD has repeatedly shown particular shifts in the bacterial and fungal microbial community, which are significantly different from age-matched neurotypical controls, but it is still unclear whether these characteristic shifts are detectable before diagnosis. Early microbial colonization patterns can have long-lasting effects on human health, and pre-emptive intervention may be an important mediator to more severe autism. In this study, we characterized both the microbiome and short-chain fatty acid concentrations of fecal samples from young children between 21 and 40 months who were showing early behavioral signs of ASD. The fungal richness and acetic acid concentrations were observed to be higher with increasing autism severity, and the abundance of several bacterial taxa also changed due to the severity of ASD. Bacterial diversity and SCFA concentrations were also associated with stool form, and some bacterial families were found with differential abundance according to stool firmness. An exploratory analysis of the microbiome associated with pre-emptive treatment also showed significant differences at multiple taxonomic levels. These differences may impact the microbial signaling across the gut-brain-microbiota axis and the neurological development of the children.

3.4 Introduction

Autism spectrum disorder (ASD) is a complex, chronic, neurological developmental disorder which is characterized by impairments to social interaction, as well as repetitive stereotyped behaviors (Fakhoury, 2015). Among children and adolescents, this disorder affects approximately 1 in every 58 - 166 children, and is 4 times more prevalent in males than females (Iglesias–Vázquez *et al*., 2020). Autism and autism severity are diagnosed by assessing the core behavioral symptoms using a number of different diagnostic tools which have been reviewed elsewhere (Falkmer *et al*., 2013). Children with ASD show a high comorbidity of gastrointestinal (GI) disturbance, which may be linked with neurological symptom severity (Chaidez, Hansen and Hertz-Picciotto, 2015), and infants who go on to be diagnosed with ASD are also more likely to have experienced GI symptoms including constipation and food intolerance between 6 – 18 months of age (Bresnahan *et al*., 2015). Autistic children are also often self-restricting when it comes to dietary preferences (Hyman *et al*., 2020); and diet is known to be an important factor in driving the composition of the microbiome in both autistic (Yap *et al*., 2021), and healthy individuals (Xu and Knight, 2015; Heiman and Greenway, 2016; Beaumont *et al*., 2017).

The gut microbiome is a complex and critical community (Dave *et al*., 2012), which has been well described (in terms of the bacterial residents) for healthy children and adults (Gilbert *et al*., 2018), and significant differences in the composition of the fecal microbiome between autistic and neurotypical children have also been shown (Kang *et al*., 2013; de Angelis *et al*., 2015). It is also now widely recognized that early-life colonization patterns can have long lasting effects on human health (Mesa *et al*., 2020), and that microbes play a crucial role in maintaining the normal functioning of their host (Gilbert *et al*., 2018). This is achieved in part by the production of exclusive microbial metabolites including short chain fatty acids (SCFA), which like the microbial members are found in dynamic quantities in the gut. The gut ecosystem will fluctuate in response to diet and nutrient availability, the presence of antimicrobials which affects patterns of cross-feeding (Ríos-Covián *et al*., 2016), as well as physical activity, and hygiene practices (Levy *et al*., 2017). Lifestyle choices can therefore result in shifts or impairments to the composition of the microbial community, and deficits in digestion, absorption, or metabolic imbalances can also feed into GI disturbance and exacerbate a breakdown of the microbial community. Under such conditions which favor the growth of opportunistic pathogenic bacteria, certain microbial products in the gut can induce an enhanced pro-inflammatory response, which can compromise the integrity of the gut epithelial barrier (Martin *et al*., 2018).

Some bacteria-derived toxins can enter the bloodstream through leaky gastrointestinal barriers and then pass through the blood brain barrier (BBB). These toxins include

enterotoxins and lipopolysaccharides (Lukiw, 2020), phenols such as 3-hydroxyphenyl acetic acid (Velásquez-Jiménez *et al*., 2021), whereas SCFA play a part in regulating the BBB (Parker, Fonseca and Carding, 2020). These events which begin in the gut can have far reaching effects due to bidirectional communication which takes place within gut-brain-microbiota axis, and involves signaling in neural, endocrine, and immune systems (Martin *et al*., 2018). The microbiome can be affected by modifications to gut motility, permeability, and intestinal secretions which are directed by the brain (Martin *et al*., 2018), and there is evidence for the microbiome to play a role in both the development and long term functioning of the brain (Sharon *et al*., 2016). Due to the potential disruption of normal brain development via the gut-brain-microbiota axis (de Angelis *et al*., 2015), it has been proposed that the gut microbiota plays a role in both the GI- and developmental symptoms that affect autistic children (de Angelis *et al*., 2015; Cryan *et al*., 2020).

An exciting prospect of the causal role the microbiome may play in the development and severity of autism, is that the microbiome is modifiable (Halmos *et al*., 2015; Pham *et al*., 2021). A temporary reduction in symptom severity has been observed after modulation of the gut microbiome using either antibiotics (Sandler *et al*., 2000), or fecal microbiome transplantation (Kang *et al*., 2017). Probiotics have also been shown to improve stool consistency and behavioral scores of autistic children (Parracho *et al*., 2010). If autism can be detected earlier, behavioral and dietary interventions can be implemented earlier, and would potentially be more effective. While the fecal microbiome of adolescent and older children with ASD has repeatedly shown shifts to the bacterial and fungal microbial community, it is still unclear whether these characteristic shifts might be detectable earlier. In a recent systematic review on the gut microbiota of children with autism, only 4 of 18 investigations included children at 2 years of age (Iglesias–Vázquez *et al*., 2020). In this study, we analysed stool samples from young children who took part in a larger study which compared two pre-emptive intervention treatments (Whitehouse *et al*., 2021). The aims of the present work were to characterize the gut microbiome and SCFA concentrations from young children who were showing behaviors in the first year of life, that are associated with later autism diagnosis (ASD-risk behaviors). The relationship between the microbiome composition and function, and its association with clinical measures for autism and neurodevelopment was investigated. We have also explored the microbiome for any association to the pre-emptive intervention these young children were receiving which may inform larger studies in the future.

3.5 Methods

3.5.1 Study design and sample collection

Samples were collected from young children who were enrolled in a broader study (Whitehouse *et al*., 2021), which was approved by the Child and Adolescent Health Service (HREC Ref: 2016008EP) in Perth, Western Australia. This mentioned study recruited infants between 9 – 14 months who were showing early social-communication delays as determined by Social Attention and Communication Surveillance–Revised (SACS-R) 12-month checklist. In this randomized control trial (RCT), the infants were randomized to one of two intervention arms 1) iBASIS-VIPP – a parent-mediated video-aided intervention supporting parent-child interaction, or 2) usual community care (UCC) – which varied and was comprised of services recommended by local health professionals, or included no additional treatment (Whitehouse *et al*., 2021). As part of this larger study, autism symptom severity and general development of the children was assessed at the time of stool sample collection using both the Mullens Scale of Early Learning (MSEL) (Mullen, 1995) and the Autism Diagnostic Observation Schedule, second edition (ADOS-2) (Lord *et al*., 2012) diagnostic tools (Table 3.1). ADOS-2 total scores were converted to calibrated severity scores (CSS), which range from 1-10 points (Shumway *et al*., 2012). The CSS scale were from scores 6 – 10 moderate-to-severe concern / autism classification, 4 – 5 mild-to-moderate concern / autism spectrum classification, and 1 – 3 little-to-no concern / non-spectrum classification.

Table 3.1. *Enrolment, follow-up timepoints and behavioural testing that took place in the AICES RCT. The number of stool samples received from this RTC and used in this current faecal microbiome study are also indicated. Six children provided stool samples at both timepoints A and B.*

AICES study timepoints	Enrollment	Follow up 1	Follow up 2	Follow up 3
Age (months)	9-14	15-20	21-26	33-38
ADOS-2 scoring		x	x	x
MSEL scoring	x	x	x	x
Current study timepoints			Timepoint A	Timepoint B
number of single stool samples received			3	15
number of replicate stool samples received			6	6
Boys:Girls			8:1	15:6

Stool samples were collected as part of the former mentioned study in Perth. Parents were instructed to collect a stool sample from a nappy or from a plastic lining which covered the toilet, and preferably free of urine, in a sterile screw-top container. The stool

samples were placed inside a sealable bag and frozen immediately in the household freezer. A provided freezer bag was used to transport the sample to the clinical assessment site (CliniKids, SUBIACO / Perth Children's Hospital, NEDLANDS) where they were frozen at −80°C until transfer to Curtin University on dry ice. A total of 30 stool samples were collected from 24 children at two timepoints during the RCT. Nine samples were collected one year after the study baseline when the children were between 21-28 months of age, and 21 samples were collected at two years postbaseline, when children were between 33 and 40 months of age. Six of the 24 children provided a stool sample at both one- and two-years post baseline (Table 3.1). Those stool samples were used in this current fecal microbiome study for which approval was granted by Human Research Ethics Committee (approval number HRE2020-0127) from Curtin University, Western Australia, and all research was conducted in accordance with the relevant regulations and guidelines.

3.5.2 Fecal DNA extraction and short chain fatty acid quantification

Fecal material was thawed at 4°C and homogenized manually for 1 min prior to sample collection for SCFA quantification (1 g ± 0.1), and microbial sequencing (0.25 g ± 0.05). The Bristol Stool Form Scale (BSFS) was used to categorize each stool form during homogenization (Mínguez Pérez and Benages Martínez, 2009). DNA was extracted from fecal samples immediately after homogenization using QIAamp PowerFecal Pro DNA kit (QIAGEN, Hilden, Germany) using the IRT protocol for QIAcube (QIAGEN), according to the manufacturer's instructions with three modifications: (1) prior to adding stool sample, three 3.5mm glass beads (Biospec) were added to bead-beating tubes, (2) after step 1, tubes were vortexed for approximately 20 sec to incorporate beads and stool, (3) followed by heating at 65°C for 10 minutes. Extraction controls were also processed following the same protocol as frozen stool samples. Fecal samples for SCFA analysis were frozen at −80°C immediately after homogenization, and then transferred on dry ice to the Science Analytical Facility at Edith Cowan University, Western Australia for SCFA quantification as previously described (Jones *et al.*, 2021).

3.5.3 Bacterial and fungal library preparation and sequencing

Bacterial DNA and gut microbiome mock community (https://www.atcc.org/products/msa-1006) were amplified using 16S primers 515F (Turner *et al.*, 1999) and 806R (Caporaso *et al.*, 2011), while fungal DNA was amplified using ITS2 primers FSeq and RSeq (Heisel *et al.*, 2015), each with a 6-8 bp unique barcode. The PCR reactions, library preparation and sequencing were performed according to methods previously described (Jones *et al.*, 2021).

3.5.4 Deconvolution and data quality filtering

Sequences were demultiplex using unique molecular barcodes with no mismatches allowed before they were removed along with primer sequences using Cutadapt (Martin, 2011). Quality filtering using DADA2 (Callahan *et al.*, 2016) was performed as previously described (Jones *et al.*, 2021). The Genome Taxonomy reference database (Version 202) formatted for use with DADA2 (https://zenodo.org/record/4735821#.YN18Om4RWis), and the UNITE general FASTA release for fungi Version 8.3 (Kõljalg *et al.*, 2005) were used to classify 16S and ITS2 sequence variants respectively, each with a minimum of 80% bootstrapping. Species were then assigned to 16S sequences with 100% identity using the same reference database. The bacterial and fungal species assignments for the top 50 ASVs were confirmed by BLAST using the same two databases, at 100% identity. ASVs with up to three matches were annotated to include all three potential species assignments, whereas any ASV with more than three identical matches was annotated with the genus name followed by "spp". Identical matches to species id numbers were not included. Where taxa are not fully resolved to a lower rank, the lowest available rank name and sp., gen., or fam. have been annotated for each lower taxonomic level. The package microDecon (McKnight *et al.*, 2019) was used to remove contamination from all sample sequences with one run of the function remove.count. Lastly, any ASVs with unassigned phylum, or with low prevalence (1 read in 5% of samples for bacterial ASVs and 1 read in 2.5% of fungal sample ASVs) were filtered out, as were fungal samples (1 sample) with less than 1000 reads. Performing decontamination in conjunction with filtering has been recommended (Cao *et al.*, 2021), and therefore these non-aggressive filtering and decontamination thresholds were chosen to balance noise reduction associated with sparse reads, while retaining less abundant but potentially important species.

3.5.5 Statistical analysis

To characterise the difference between autistic and non-spectrum children using CSS, the children were placed into the category defined by their score: CSS from 6-10 ASD, CSS from 4-5 non-autism autism spectrum disorder (NAASD), and CSS from 1-3 no developmental concern (NDC). Also, because stool form has previously been shown to be associated with alpha diversity and SCFA concentrations, the samples were placed into three groups determined by the Bristol stool form scale. Stool scoring 1-2 was considered firm, 3-4 was considered normal, and 5-6 was considered loose. No stools were scored as 7 in this study. Beta diversity was used to assess differences between CSS and intervention groups, and stool form using Euclidian distances of center log ratio transformed counts and visualized using Principal coordinates analysis (PCoA). PERMANOVA was performed in PRIMER-e v7 (Anderson, Gorley and Clarke, 2008) with 9999 unrestricted permutations of the raw data and type 3 sum of squares. ANOVA

was used to determine differences in SCFA concentration between CSS groups using Tukey multiple comparison method to adjust p values. For all SCFA concentrations, Q-Q plots were used to assume normality, and Levene's Test was used to check for homogeneity of variance. The total SCFA concertation was summed from all acids, with the concentration of valeric acid set to 0 for 3 individuals where the concentration was below detection.

Differential abundance testing between CSS groups at the ASV level was performed in DESeq2 (Love, Huber and Anders, 2014). At the family, genus, and species level, differences observed between intervention group and CSS, were determined using Mann-Whitney U test and Bristol stool groups using the Kruskal-Wallis test (all comparisons included only samples from timepoint B). These higher order comparisons were performed only on dominant reads, which were those with a sum greater than 200 reads across all bacterial samples (372 ASV), and greater than 25 reads across all fungal samples (136 ASVs). False discovery rate (FDR) due to multiple testing was corrected for with a Benjamini-Hochburg adjustment (BH) (Benjamini and Hochberg, 1995). The core bacteriome consisting of species with 20% prevalence and a relative abundance over 0.01% were also determined in MicrobiomeAnalyst (Dhariwal *et al.*, 2017). Pearson correlation was calculated for the linear regression between MSEL and alpha diversity estimates for bacterial ASV counts, with normality assumed by the Shapiro-Wilk normality test. A single sample was removed from this correlation analysis as no MSEL score was recorded for that sample.

Predictive functional profiling was inferred from the total bacterial 16S rRNA gene sequence data using Tax4Fun (Wemheuer *et al.*, 2020). Here, SILVA reference sequences were mapped to KEGG orthologs (KO), and the profile was then filtered to include only a core group (> 0.01% across all samples) of metabolic pathways, with photosynthisis (ko00195) also removed. Differences in the core metabolic profile between CSS groups were visualized in Primer using using PCoA of Bray-Curtis similarity, and using PCA in STAMP (Parks *et al.*, 2014). Significant differences between individual pathways were also assessed in STAMP using a two-sided Welch's test, and Benjamini-Hochberg FDR for multiple testing.

3.6 Results

3.6.1 Description of the total bacterial and fungal datasets

The bacteriome from all children in this study were dominated by Firmicutes (62%), Bacteroidota (25%), and Actinobacteriota (10%), and the mycobiome was dominated by Saccharomyces (92%) and Kazachatania (4%). A description of the data quality including read depth (Sup Table 3.1), replicate sampling and a positive control in the

form of a bacterial mock community (Sup Figure 3.1) are summarized in the supplementary results section.

3.6.2 Microbiome composition assessed by stool form, and intervention group

To explore the data principal coordinates analysis (PCoA) was used to visualize beta diversity between stool form, age, and dominant taxa from stool samples at both timepoints, while differences between intervention groups were assessed only from those samples taken at timepoint B. A significant difference in the beta diversity of intervention groups was determined by PERMANOVA at the level of species ($p = 0.014$) genus ($p = 0.008$), family ($p = 0.006$), and order ($p = 0.008$) (Sup Figure 3.2). A significant difference in dispersion of Phyla around the centroid between the two intervention groups was also determined by PERMDISP ($p = 0.04$). The difference in community composition seemed to be driven by a higher abundance of Lachnospirales ($p = 0.005$, FDR = 0.12) in the iBASIS-VIPP intervention group. A further 2 families, 8 genera, and 13 species were also differentially abundant between the two intervention groups (Sup Table 3.2).

The age in months did not have a clear impact on sample beta diversity (Sup Figure 3.3), and independent of timepoint, bacterial communities clustered according to stool form (Figure 3.1 A), with firmer stool samples clustering separately from loose stool samples. Bacterial families with differential abundance according to stool form were detected (Sup figure 3.4). Prior to FDR correction the abundance of Butyricicoccaceae, and Pasteurellaceae was significantly higher in loose stool, compared to firm stool ($p > 0.05$ FDR>0.60), the abundance of Enterobacteriaceae was significantly higher in loose stool compared to normal stool ($p = 0.03$ FDR = 0.42), and the abundance of Monoglobales A UBA1381, were significantly lower in firm stool compared to normal or loose stool ($p = 0.02$, FDR = 0.42). A further six genera were identified to have lower abundance in firm stool compared to lose or normal stool ($p < 0.05$, FDR < 0.59), and 3 genera had lower abundance in normal stool compared to firm or loose stool ($p < 0.03$, FDR0.44) (Sup Table 3.3). Fungal communities on the other hand did not group according to stool form, but clustered according to the dominant ASVs (Figure 3.1 B).

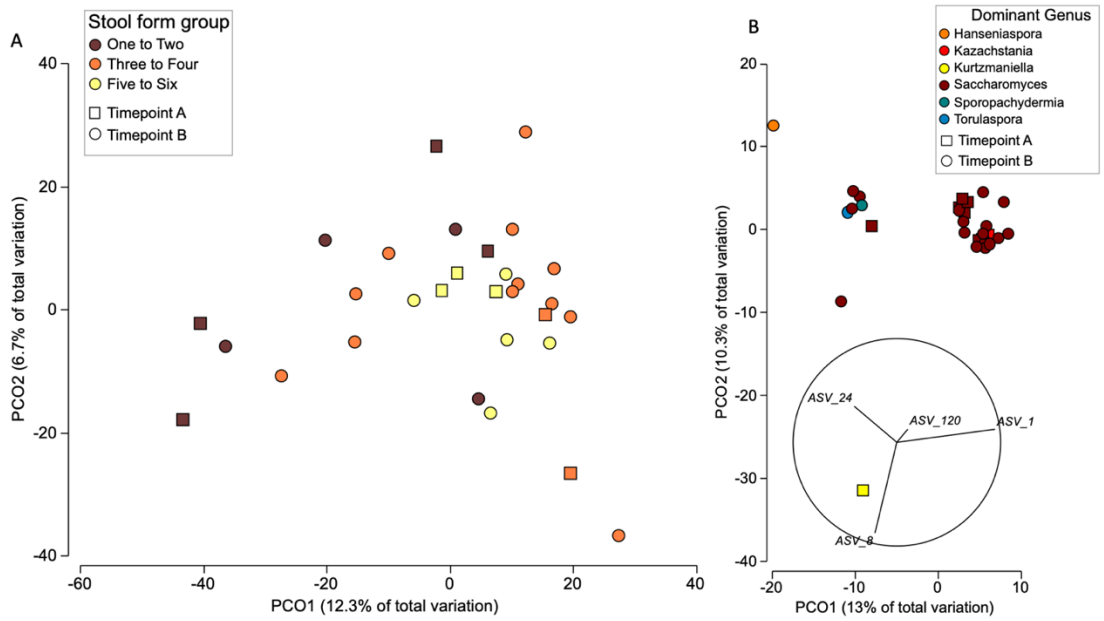


Figure 3.1. *Beta diversity of bacterial (A) and fungal (B) communities from all individuals at both timepoints using PCoA. A) Distribution of the bacterial communities are shown due to stool form and timepoint. B) Fungal communities are displayed based on dominant taxa and timepoint with abundant ASVs plotted as vectors. Beta diversity was estimated from Euclidian distances between CLR transformed counts.*

3.6.3 Comparison of microbiome diversity between diagnosis outcomes

The microbiome composition among CSS groups were compared at timepoint B only, and showed that the proportions of bacterial phyla were similar among autistic children and children in the NDC group, however, the abundance of Actinobacteria in the NAASD group was considerably lower than the other groups (Table 2). Strong interpersonal differences were observed between all children, and no significant differences were determined between CSS groups ($p>0.3$). Similar alpha diversity was observed between CSS groups for both bacterial and fungal communities, and alpha diversity also showed no significant relationship with MSEL (Figure 3.2).

Table 3.2. *Description of child and microbiome characteristics at timepoint A and B, and per diagnosis category. A single sample was removed from the fungal data, but not the bacterial data, and therefore the group characteristics for NDC are different for bacterial *1 and fungal *2 datasets. Abbreviated phyla names are Actinobacteriota (Act), Bacteroidota (Bac), Firmicutes (Firm), Proteobacteria (Pro), and abbreviated CSS groups are Autism (ASD), non-autistic autism spectrum disorder (NAASD), and no developmental disorder (NSD).*

Time A (n = 9)	mean Age	iBASIS-VIPP : UCC	Boys: Girls	Average Stool form	Average MSEL	Act (%)	Bac (%)	Firm (%)	Pro (%)
CSS									
ASD	25.5	2:2	3:1	2	74.8	2.9	29.2	66.9	0.5
NDC	24.2	2:3	5:0	3.8	108	2.5	30.3	64.5	0.9
Time B (n = 21)									
CSS									
ASD	36.4	7:4	8:3	3.1	78.7	15.3	17.5	65.6	0.3
NAASD	36.6	4:1	4:1	4.2	97.0	4.0	39.6	53.6	0.2
NDC*[1]	34.6	2:3	3:2	4.0	104.8	10.3	26.9	60.2	1.3
NDC*[2]	34.8	1:3	3:1	4.5	98.3				

Differential abundance testing was used to compare the microbiome composition at the ASV level among CSS groups at timepoint B only. In total, 19 Bacterial and 4 fungal ASVs were detected with significantly different abundance after adjusting for multiple testing. These ASVs were identified by the lowest taxonomic rank available (Table 3.3), and most bacteria were classified as either Bacteroidia (Bacteroidales) or Clostridia, most of which were Lachnospiraceae, while most fungal taxa were classified as Saccharomycetales. Six bacterial ASVs were enriched in autistic children, half of which belonged to Lachnospiraceae, and ten bacterial ASVs were significantly enriched in the NDC group. Two different fungal ASVs were significantly enriched in both the mycobiome of NDC and autistic children. Differences in the abundance of genera and species between CSS groups were also determined. After correcting for multiple testing (FDR), there were no significant differences, however prior to FDR, there were a number of taxa identified (Sup Table 4).

Table 3.3. *Bacterial and fungal ASV's identified with DESeq2 as having significant changes in abundance between CSS groups. Positive fold changes are enriched in NDC, and negative fold changes are enriched in ASD or NAASD groups.*

ASD compared to NDC group		NAASD compared to NDC group	
ASV# and lowest rank	**log2FoldChange**	**ASV# and lowest rank**	**log2FoldChange**
Bacteroidaceae		**Erysipelatoclostridiaceae**	
287 Bacteroides	22.4	194 Erysipelatoclostridium sp	15.6
Burkholderiaceae		**Debaryomycetaceae**	
184 Parasutterella	6.7	14 Debaryomyces hansenii	28.0
Lachnospiraceae		**Lachnospiraceae**	
77 CHKCI001	5.3	127 Ruminococcus A faecicola	24.0
Bacteroidaceae		**Lactobacillaceae**	
172 Bacteroides finegoldii	-22.2	202 Lactobacillaceae sp	19.7
Christensenellales fam		**Oscillospiraceae**	
285 Christensenellales sp	-20.7	268 Oscillospiraceae sp	21.3
Lachnospiraceae		**Veillonellaceae**	
220 Blautia sp	-24.1	211 Veillonella parvula A	24.1
162 Blautia sp	-22.3	**Bacteroidaceae**	
58 Lachnospiraceae sp	-7.5	172 Bacteroides finegoldii	-24.6
150 Lachnospiraceae sp	-23.8	**Lachnospiraceae**	
Saccharomycetaceae		220 Blautia sp	-15.8
21 Saccharomyces sp	-22.1	162 Blautia sp	-17.6
13 Eremothecium sinecaudum	-20.3	58 Lachnospiraceae sp	-6.9
		150 Lachnospiraceae sp	-23.1
		Saccharomycetaceae	
		21 Saccharomyces sp	-19.6

3.6.4 Temporal changes in fecal samples from six individuals

Six children provided samples at both timepoint A and timepoint B. When the 12 samples were clustered using Euclidian distances (Figure 3.3) the samples grouped based on participant, with a larger difference between one individual. This individual, and one other additional individual each received a different ADOS-2 score based on their behavior at timepoint B than they had previously received at timepoint A. Overall the richness, diversity, and phylogenetic diversity were observed to be higher at the second timepoint, when the children were on average 36 months of age. The SCFA concentrations did not shift in a particular direction between timepoints (Figure 3.3).

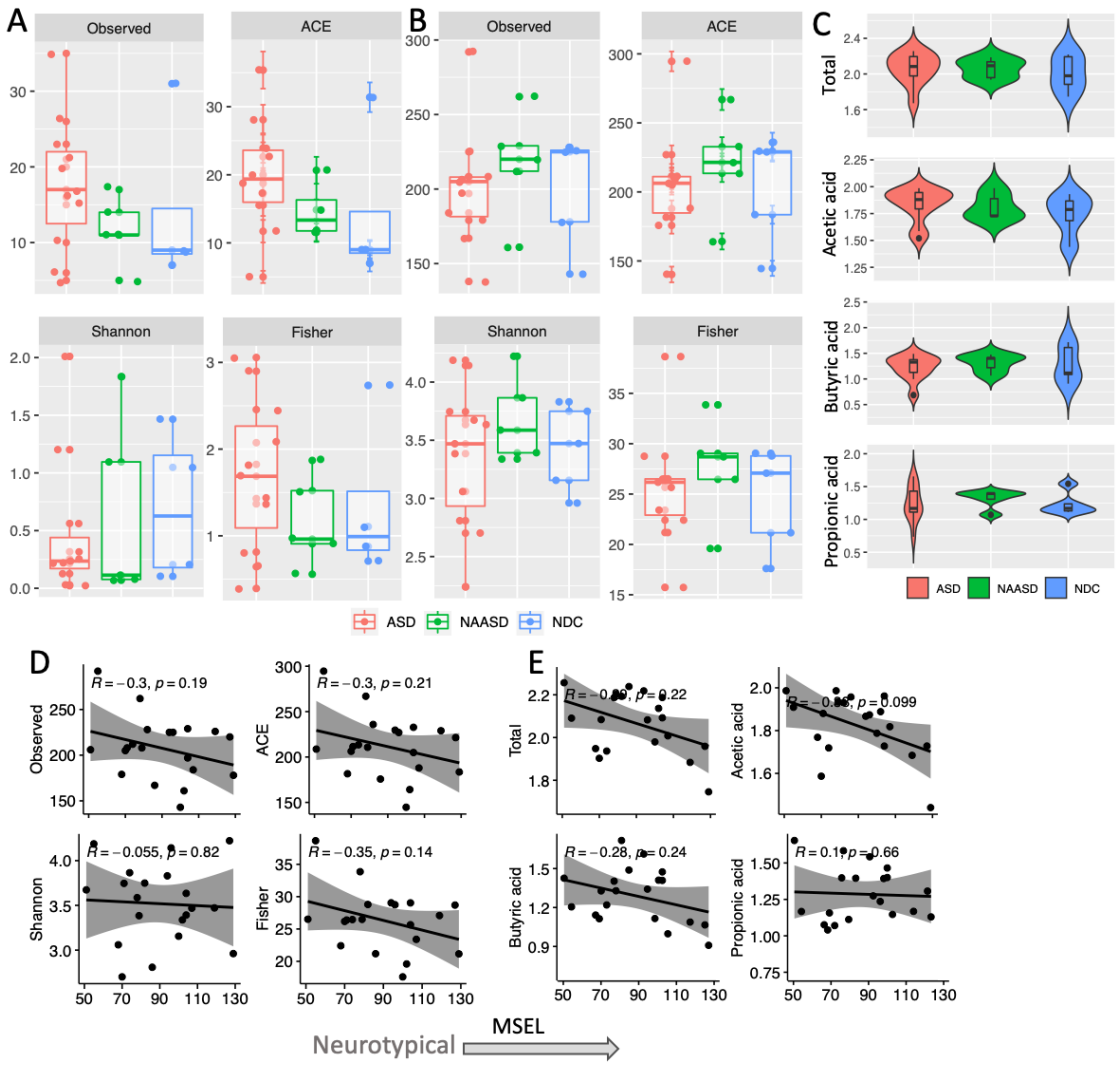


Figure 3.2. *Differences in alpha diversity and SCFA concentration in young children according to autism severity. A) Fungal community alpha diversity estimates according to CSS groups. B) Bacterial community alpha diversity estimates according to CSS groups. C) Total and individual SCFA concentration (log10) in CSS groups. D) Association between MSEL score and bacterial alpha diversity based on Pearson correlation. E) Association between MSEL score and SCFA concentration based on Pearson correlation.*

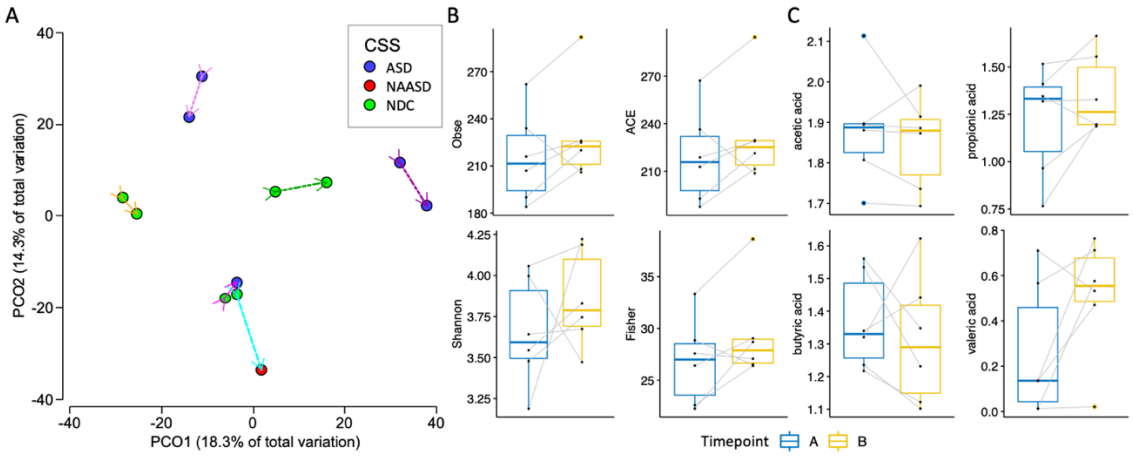


Figure 3.3. *Changes is the gut microbiome of 6 children between an average of 24-months of age (timepoint A), and 36-months of age (timepoint B). A) PCoA of bacterial beta diversity based on Euclidian distances of CLR transformed counts and the trajectory of the microbiome across time is shown with a uniquely colored arrow for each individual. B) Alpha diversity estimates between timepoint A and B. C) SCFA concentrations between timepoint A and B.*

The common core bacteriome between all children at timepoint A and B was identified using prevalence and abundance filtering. A total of 46 species were identified as core members at timepoint A and 50 at timepoint B, with 39 of those species present in both core communities. *Blautia* A sp, Lachnospiraceae sp, and Ruminococcaceae sp remained highly prevalent between both timepoints, and 50 other species changed only marginally or not at all between timepoints. 11 species that were present in the core microbiome at timepoint B were missing from the core at timepoint A, most of which were from Lachnospiraceae. *Phocaeicola* sp lost prevalence over time and *Blautia* A *faecis*, *Romboutsia timonensis/ilealis*, *Ruminococcus* C *callidus*, Clostridia sp, Dialisteraceae sp and *Faecalibacterium* sp all increased in prevalence over time (Figure 3.4).

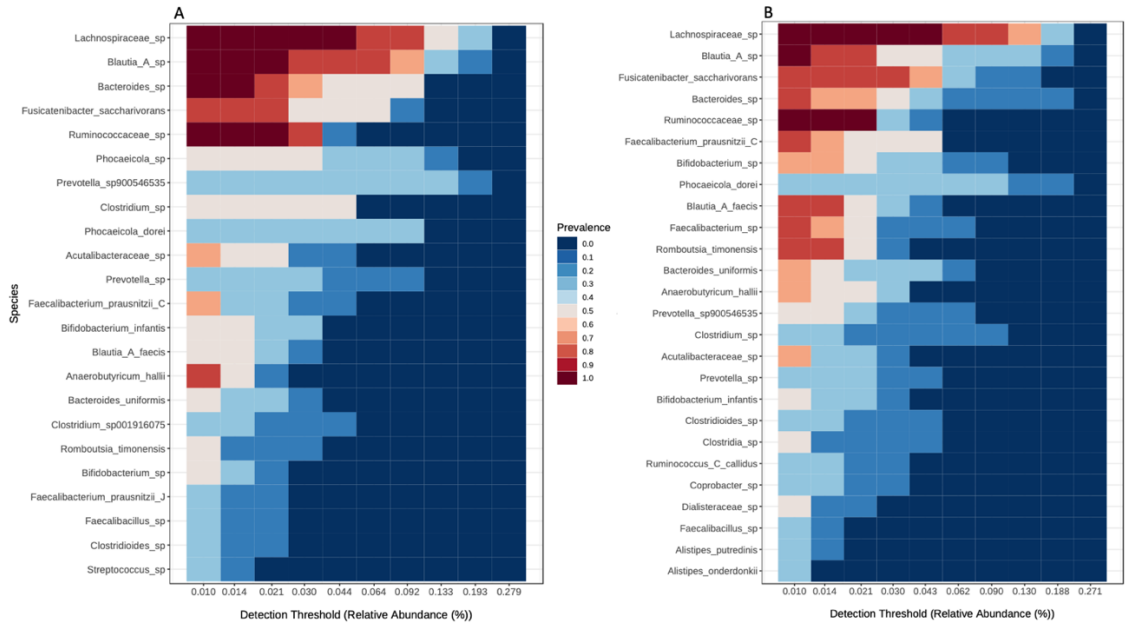


Figure 3.4. *Shifts in the core microbiome of 6 children between timepoints. A) Children between 21 and 28-months of age at timepoint A, and B) children between 33 and 40-months of age at timepoint B.*

3.6.5 SCFA concentrations

The average, total faecal SCFA concentration across all children was 119.01 µmol/g (95% confidence interval 104.29 - 133.73) and the average molar ratio of acetate, propionate, and butyrate was 67:19:20. No significant differences in total or individual SCFA concentration were found between CSS (Figure 3.2 C), however, children in the ASD group tended to have higher acetic acid, and total SCFA concentrations, and similarly, total and acetic acid concentration negatively corelated with MSEL score (Figure 3.2 E). The SCFA concentrations were also similar between the two treatment groups. When relating SCFA concentrations to stool form, the average concentration of acetate, propionate, and the total SCFA concentration were all highest in normal stool samples, compared to firm or loose stool samples, whereas iso-butyrate and iso-valerate were highest in firm stools, followed by normal and then loose stools. Butyric acid was the only acid that increased in concentration with stool looseness (Figure 3.5).

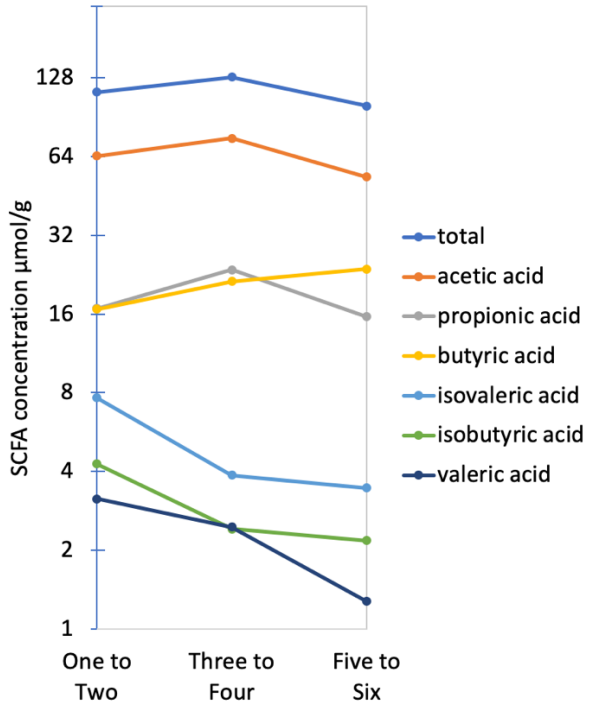


Figure 3.5. *Average total and individual SCFA concentrations across Bristol stool form groups. The Y axis is plotted on a* log_2 *scale.*

3.6.6 Functional analysis

Predictive community functional profiling using Tax4fun returned 321 functional pathways within: Cellular Processes, Environmental Information Processing, Genetic Information Processing, human diseases, Metabolism, and Organismal Systems. On average 38.5% of ASVs, and 52.7% of sequences per sample were mapped to a reference KO and used for prediction. After filtering to include only core metabolic functions, the profile consisted of 157 pathways within 13 classes. There was no visual clustering in the core metabolic profile between CSS groups using PCA (Figure 3.6), although a significant difference in the proportion of Tetracycline biosynthesis (after FDR correction), and a number of other pathways (prior to FDR correction) were identified between CSS groups using Welch's t-test (Sup table 3.5). Differences in the metabolic profile between intervention groups were also observed using PCA (Sup figure 3.5), as well as individual metabolic pathways.

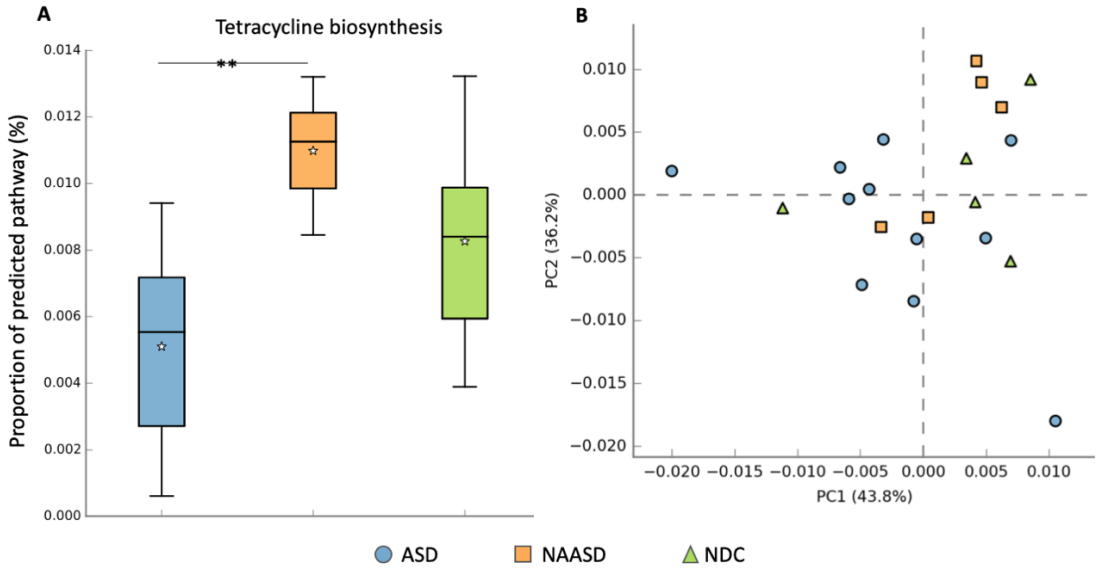


Figure 3.6. *Differences in metabolic pathways between CSS groups. A) Proportion of the predicted pathway tetracycline biosynthesis among CSS groups tested using Welch's t-test after FDR correction. B) PCA distribution of predicted core metabolic pathways by CSS group.*

3.7 Discussion

Identifying the behaviours associated with autism at a young age is possible, although predicting the trajectory of young children who are diagnosed with autism is typically not possible (Hyman *et al*., 2020). Roughly 80% of children who are diagnosed at less than three years of age will retain their diagnosis, however, because it is difficult to distinguish between either autism, or Pervasive Developmental Disorder-Not Otherwise Specified a conclusive diagnosis is difficult to make in children under 2 years of age (Akshoomoff, 2006). Therefore, in this study we used CSS, and MSEL scores as indicators of autism symptom severity to explore the gut microbiome and SCFA associated with these young children. It is also important to note that the children in the no-developmental-concern group may not be considered neurotypical controls, as all children in this study were showing early behavioural signs of ASD determined by SACS-R at 9 – 14 months of age.

3.7.1 Overall composition of the fecal microbiome and SCFA concentrations among CSS groups

Assessment of the bacteriome in relation to autism has led to conflicting results, especially when comparing Bacteroidetes and Firmicutes (Iglesias–Vázquez *et al.*, 2020); although ASD seems to be more often associated with a decrease in Firmicutes (Andreo-Martínez *et al.*, 2021). In this present study, there was no significant difference

in the relative abundance of any phyla between ASD and NDC children, and firmicutes were the most abundant and prevalent phyla, averaging 62% of the total proportion of reads across all children sampled. The abundance of Actinobacteriota dropped below 4% of the total community composition in the NAASD group, and a reduction in this phyla has also been observed previously in autistic children compared to neurotypical controls (Coretti *et al.*, 2018). The NAASD group also had the highest levels of butyrate and propionate in the study, and had a significantly elevated proportion of tetracycline biosynthesis based on predictive profiling. Tetracyclines are of the most commonly used, oral, broad spectrum antibiotics which can promote the proliferation of anaerobic bacteria (Kovtun *et al.*, 2020).

Significant differences in both the community composition (Strati *et al.*, 2017; Coretti *et al.*, 2018; Pulikkan *et al.*, 2018; Ma *et al.*, 2019), and SCFA concentration (Adams *et al.*, 2011; Wang *et al.*, 2012; Coretti *et al.*, 2018; Bojović *et al.*, 2020) have also been observed between autistic and neurotypical controls. Changes in composition are likely to affect the functional microbiome and may help reveal the mechanisms by which the microbial community may differentiate due to the ASD phenotype. SCFA concentrations in stool are good indicators of what was produced by resident bacteria as they reflect what was excreted after absorption by the host. Butyrate and propionate have been shown to be elevated in stool samples from children with ASD compared to controls (Coretti *et al.*, 2018). Considering children with a similar fibre intake, propionate has again been found in significantly higher concentrations in the stool of autistic children compared to neurotypical controls (Wang *et al.*, 2012). This is particularly important because both acids can cross the BBB and can interact with brain cells via G-protein-coupled receptors (Abdelli, Samsam and Naser, 2019). Although, in this study we found no such difference in either beta diversity or SCFA concentration between CSS groups. Instead, butyrate was the only SCFA which increased with stool looseness. The similar levels of SCFA across CSS groups but not Bristol stool groups may indicate that stool form may be a confounding factor that makes it difficult to determine differences between neurological development, especially with a small sample size. It may be that at this young age, differences in the community structure associated with neurological disorders are less pronounced, but still present at a finer scale.

The average richness and diversity of the bacterial community also increased from the first (A) to second (B) time point in the six children that provided stool samples at both timepoints; as did the number of bacterial species included in the core microbiome. As the diversity of a child's diet expands in early-life, the microbiome also continues to develop and increase in diversity (Matamoros *et al.*, 2013). Although, there was no significant difference in alpha diversity between CSS groups, and this is consistent with other studies with a larger number of young participants who were assessed for autism

using various methods (Strati *et al.*, 2017; Pulikkan *et al.*, 2018; Kong *et al.*, 2019; Fouquier *et al.*, 2021). In this study, there was a trend for increased bacterial diversity within the NAASD group, as well as a trend towards a negative correlation between bacterial diversity and MSEL score. This negative correlation is in agreement with another study which found three alpha diversity measures of the faecal microbiome at 1-year of age negatively correlated with both the Early learning Composite (a combination of 4 of the 5 standardized T-scores from the MSEL) and two of the five Mullen scales at 2-years of age (Carlson *et al.*, 2018).

While these trends in alpha diversity may indicate new avenues for future investigation, we caution that it may be difficult to use traditional diversity indices to describe changes in the gut that are linked with less understood, heterogeneous, modern diseases such as ASD. Not to mention that the sequencing depth – the basis on which the diversity measures are calculated – varies considerably between studies which may make it difficult to compare conclusions from different datasets (Willis, 2019). The inconsistency in alpha diversity in relation to autism has also been mentioned in a review in this area (Krajmalnik-Brown *et al.*, 2015). Here we propose that either an increase or decrease away from the diversity needed to maintain a healthy homeostatic microbiome may be disadvantageous. For example, an increase in diversity may indicate a bloom in pathobionts (Levy *et al.*, 2017), or an increase in usually more transient bacteria which may breakdown the normal community structure. Similarly, reduced diversity could indicate a loss of functionality important microbial members and a community which is disordered (Levy *et al.*, 2017).

3.7.2 Species and strain differences in the fecal microbiome between CSS groups

To detect bacteria associated with ASD, we compared differential abundance between CSS groups. Using the BH adjustment to control for false positives, we did not detect any significant features, although, our investigation lacks statistical power using nonparametric tests with small sample size, therefore, we discuss significant results prior to FDR correction. Overall members of Enterobacteriaceae, and Negativicutes, were increased only in the NDC group. Negativicutes are known for their contribution to propionate production (Reichardt *et al.*, 2014)*, and Veillonella* (Negativicutes), are known to fermenting lactate to produce SCFA (Kang et al., 2013). *Veillonella* have been found to be reduced in the faecal microbiome of autistic children compared to neurotypical controls (Strati *et al.*, 2017); whereas Enterobacteriaceae have been found to increase (De Angelis *et al.*, 2013). Lachnospiraceae including *Muricomes*, CHKCI001, and *Ruminococcus*, were also enriched in the NDC group compared to either the ASD or NAASD groups.

Taxa enriched in both non-neurotypical groups included members of Lachnospiraceae, including *Blautia* and an un unresolved Lachnospiraceae species. *Blautia* is bile-metabolizing, associated with tryptophan metabolism, and has been found at reduced abundance in a BTBR T^{+} *Itpr3*^{tf}/J mouse model of ASD (Golubeva *et al.*, 2017). This result is somewhat contradictory as *Blautia* are commonly associated with a typically developing infant microbiome (Hill *et al.*, 2017), a healthy adult microbiome (Tap *et al.*, 2009), and are well known for their contribution to SCFA production (Louis and Flint, 2017). *Clostridium* sp., were also more abundant in the ASD group, which is in line with other studies (Gondalia *et al.*, 2012; Coretti *et al.*, 2018; Ma *et al.*, 2019). *Clostridium* are one group of bacteria known to be able to produce toxins which can cross the BBB (Góra *et al.*, 2018), and are commonly found in both higher abundance (Iglesias–Vázquez *et al.*, 2020), and numbers based on CFU/g (Finegold *et al.*, 2017) in faecal samples from autistic children compared to controls.

Bacteroides caccae, has been associated with a decrease in abundance in the faecal microbiome of autistic children (Averina *et al.*, 2020) but was enriched in the ASD group in this study. Certain strains of *B. caccae* are mucolytic, which may place them at close proximity to the host, where it can both influence and be influenced by the host (Tailford *et al.*, 2015). Furthermore, myocytic bacteria may be opportunistic pathogens (Ganesh *et al.*, 2013), and because complete degradation of mucins may require co-metabolism evolving several species (Tailford *et al.*, 2015), shifts in a the abundance of some members could impact the functionality of the total myocytic community, which in turn would be either beneficial or harmful to the host. While this finding adds to the catalogue of autism associated bacterial taxa, it may also indicate the importance of identifying different bacterial strain, as well as using complementary metabolite data to fully understand community changes between groups. Furthermore, the inconsistency seen in the literature regarding taxa associated with autism, may be caused by differences in databases used for classification. This can be improved by a multi-omics study design, but also by collecting information that can influence microbiome composition, such as diet including fibre and protein content, as well as pre- or pro-biotic supplements, GI problems, and information of bowel movements.

3.7.3 Stool form may help explain inconsistencies in the bacteriome of children with ASD

Poor stool form, is an indication of GI distress, and issues such as constipation, diarrhea and flatulence are reported as more prevalent in children with ASD than neurotypically developing children (Chaidez, Hansen and Hertz-Picciotto, 2015). The way GI symptoms are documented could result in variation, for example, if study participants assess their own stool using the Bristol Stool scale, there may be individual bias, or error if recalling stool form after the collection has happened. In this study, stool form was

assessed during sample processing, which allowed for a non-bias assessment. We found differences in bacterial composition between firmer stool compared to normal or loose stool. This finding is in agreement with another study comparing autistic and neurotypical controls (including sibling controls), where stool consistency was among the factors most strongly associated with the microbiome composition (Yap *et al.*, 2021). However, Yap and colleagues proposed a top-down impact on stool composition beginning with reduced dietary diversity due to self-restricting eating habits, and subsequent reduced microbiome diversity resulting in loose stool consistency.

In our study, both richness and diversity were lower in firmer stool samples compared to normal or loose stool (data not shown). Stool becomes firm with a lower water content, and has been linked to longer whole gut transit times (Saad *et al.*, 2010), which may impose selective pressures on the microbial community, particularly if firm stool is experienced frequently. This same trend in reduced microbial diversity in firm stool has been observed in adults with functional constipation (Huang *et al.*, 2018). In our study SCFA concentrations (except for butyric acid) were also reduced in looser stools, which is likely due to both water content, and the potential for microbial members to influence and be influenced by gut motility (Zhao and Yu, 2016). Therefore, our results are in line with the emerging hypothesis that the microbiome is part of a circular feedback-loop, where behaviour and environment have a top-down effect on the microbiome and stool form, and the microbiome imposes a bottom-up effect. Collecting GI disturbance metadata but also recording stool form (as this may change even within a day) and using this information as a potential confounding variable might help explain some of the observed inconsistencies in ASD-associated bacteria.

3.7.4 Limitations

The number of children in this study was limited and the number of children in diagnostic and treatment groups was small and unbalanced, and no information regarding diet and gastrointestinal distress was collected. Also, because the children were first recruited prior to diagnosis, it was not possible to choose the number of children in each category. As a result, the number of children with no developmental concern was quite small. Additionally, the diagnosis of the children in the iBASIS-VIPP and UCC groups were mixed, and due to the sample size, it wasn't possible to examine differences between diagnosis within the intervention group. Using full siblings residing in the same household as the healthy control group has been shown to be the most reliable control group as it covers both environmental factors and genetic background; however, this would also not be possible with this pre-emptive study design. We would also like to point out that there are numerous methods used to diagnose autism severity, and therefore within the literature we discuss, a number of diagnostic tools have been used. Lastly, due to the small sample size we chose to display results that were

significant prior to correction for multiple testing. While this less stringent approach may lead to more false positives, it may also indicate those taxa that could be important in the shifts taking place in these young children that otherwise would be missed.

3.7.5 Conclusion and future research

In this study the relative abundance of the microbiome at the phylum level, and the diversity of bacterial and fungal communities were similar when viewed among neurological developmental groups. The microbiome composition and SCFA concentrations were instead found to be significantly associated with stool form, indicating that this factor might be important to consider when interpreting the microbiome composition of young children with autism – especially considering the high prevalence of gastrointestinal issues in children with ASD. While bacterial or fungal diversity could not be used to discriminate between neurological development, differential abundance in the community structure at the genus, species and strain level was detected between CSS groups. Together, these findings indicate that subtle changes in bacterial composition may occur in the microbiome of young children with autism. We also found that the amount of interindividual difference in the microbiome from the 6 children over two timepoints, did not seem to be consistent, or related to autism severity. This is likely due to the small number of samples, but it would be beneficial to collect a longitudinal series of samples along with diet and bowel movement history to establish the influence of these factors on the development of young children showing early behavioural signs of ASD.

Lastly, pre-emptive treatment for children at risk of developing autism is an important research area as there is currently no cure for autism, and treatments that begin before diagnosis – earlier in life – may be more effective than those started later in life. The children in this study were involved in a broader study examining the efficacy of behavioural interventions for children showing early signs of ASD (Whitehouse *et al.*, 2021). Although neither behavioural intervention arm involved dietary changes, and the sample size was very small, significant differences in bacterial community structure were observed between the intervention groups. While it is unclear how the behavioural intervention could impact the structure of the microbiome, the differences observed at multiple taxonomic levels between the two treatment groups may indicate an effect on the microbiome that warrants future research addressing the possibility of a top-down effect of cognitive or behavioural changes which selectively benefit some bacterial groups. However, this potential needs to be investigated in a much larger study to account for the many factors affecting the gut microbiome in the developing child.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

CTC, and JJ conceived of the study design; JJ conducted all laboratory work; JJ, CTC, SNR, and MM-D analyzed the data; JJ wrote the manuscript; and all authors revised the paper.

Acknowledgments

The authors would like to thank Andrew Whitehouse and Sarah Pillar for providing fecal samples, and for their support in drafting this manuscript. We also thank Alishum Ali for his DADA2 formatted GT database for taxonomy assignments, and for his assistance with multiple identical species assignments.

Data Availability Statement

The ASV tables, metadata and sequences reads used in this study can be found in the figshare online repository 10.6084/m9.figshare.19327475

3.8 References

Abdelli, L. S., Samsam, A. and Naser, S. A. (2019) 'Propionic Acid Induces Gliosis and Neuro-inflammation through Modulation of PTEN/AKT Pathway in Autism Spectrum Disorder', *Scientific Reports*. Nature Publishing Group, 9(1), pp. 1–12. doi: 10.1038/s41598-019-45348-z.

Adams, J. B. *et al.* (2011) 'Gastrointestinal flora and gastrointestinal status in children with autism - comparisons to typical children and correlation with autism severity', *BMC Gastroenterology*. BioMed Central, 11, p. 22. doi: 10.1186/1471-230X-11-22.

Akshoomoff, N. (2006) 'Use of the Mullen Scales of Early Learning for the assessment of young children with Autism Spectrum Disorders', *Child Neuropsychology*, 12(4–5), pp. 269–277. doi: 10.1080/09297040500473714.

Anderson, M. J., Gorley, R. N. and Clarke, K. R. (2008) 'PERMANOVA+ for PRIMER: Guide to Software and Statistical Methods'. Plymouth, UK: PRIMER-E.

Andreo-Martínez, P. *et al.* (2021) 'A Meta-analysis of Gut Microbiota in Children with Autism', *Journal of Autism and Developmental Disorders*. Springer, 52(3), pp. 1374–1387. doi: 10.1007/s10803-021-05002-y.

de Angelis, M. *et al.* (2015) 'Autism spectrum disorders and intestinal microbiota', *Gut Microbes*, 6(3), pp. 207–213. doi: 10.1080/19490976.2015.1035855.

De Angelis, M. *et al.* (2013) 'Fecal Microbiota and Metabolome of Children with Autism and Pervasive Developmental Disorder Not Otherwise Specified', *PLoS ONE*, 8(10). doi: 10.1371/journal.pone.0076993.

Averina, O. V. *et al.* (2020) 'The bacterial neurometabolic signature of the gut microbiota of young children with autism spectrum disorders', *Journal of Medical Microbiology*. Microbiology Society, 69(4), pp. 558–571. doi: 10.1099/JMM.0.001178/CITE/REFWORKS.

Beaumont, M. *et al.* (2017) 'Quantity and source of dietary protein influence metabolite production by gut microbiota and rectal mucosa gene expression: A randomized, parallel, double-blind trial in overweight humans', *American Journal of Clinical Nutrition*, 106(4), pp. 1005–1019. doi: 10.3945/ajcn.117.158816.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing', *Journal of the Royal Statistical Society: Series B (Methodological)*. John Wiley & Sons, Ltd, 57(1), pp. 289–300. doi: 10.1111/J.2517-6161.1995.TB02031.X.

Bojović, K. *et al.* (2020) 'Gut Microbiota Dysbiosis Associated With Altered Production of Short Chain Fatty Acids in Children With Neurodevelopmental Disorders', *Frontiers in Cellular and Infection Microbiology*. Frontiers Media S.A., 10, p. 223. doi: 10.3389/fcimb.2020.00223.

Bresnahan, M. *et al.* (2015) 'Association of maternal report of infant and toddler gastrointestinal symptoms with autism: Evidence from a prospective birth cohort', *JAMA Psychiatry*. American Medical Association, 72(5), pp. 466–474. doi: 10.1001/jamapsychiatry.2014.3034.

Callahan, B. J. *et al.* (2016) 'DADA2: High-resolution sample inference from Illumina amplicon data', *Nature Methods*. Nature Publishing Group, 13(7), pp. 581–583. doi: 10.1038/nmeth.3869.

Cao, Q. *et al.* (2021) 'Effects of Rare Microbiome Taxa Filtering on Statistical Analysis', *Frontiers in Microbiology*. Frontiers, 11, p. 3203. doi: 10.3389/fmicb.2020.607325.

Caporaso, J. G. *et al.* (2011) 'Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 108(1), pp. 4516–4522. doi: 10.1073/pnas.1000080107.

Carlson, A. L. *et al.* (2018) 'Infant Gut Microbiome Associated with Cognitive Development', *Biological psychiatry*. NIH Public Access, 83(2), p. 148. doi: 10.1016/J.BIOPSYCH.2017.06.021.

Chaidez, V., Hansen, R. L. and Hertz-Picciotto, I. (2015) 'Gastrointestinal problems in children with autism, developmental delays or typical development', *journal of Autism Dev Disord*. NIH Public Access, 44(5), pp. 1117–1127. doi: 10.1007/s10803-013-1973-x.Gastrointestinal.

Coretti, L. *et al.* (2018) 'Gut microbiota features in young children with autism spectrum disorders', *Frontiers in Microbiology*. Frontiers Media S.A., 9. doi: 10.3389/fmicb.2018.03146.

Cryan, J. F. *et al.* (2020) 'The gut microbiome in neurological disorders', *The Lancet Neurology*. Lancet Publishing Group, pp. 179–194. doi: 10.1016/S1474-4422(19)30356-4.

Dave, M. *et al.* (2012) 'The human gut microbiome: Current knowledge, challenges, and future directions', *Translational Research*. Mosby Inc., pp. 246–257. doi: 10.1016/j.trsl.2012.05.003.

Dhariwal, A. *et al.* (2017) 'MicrobiomeAnalyst: A web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data', *Nucleic Acids Research*. Oxford University Press, 45(W1), pp. W180–W188. doi: 10.1093/nar/gkx295.

Fakhoury, M. (2015) 'Autistic spectrum disorders: A review of clinical features, theories and diagnosis', *International Journal of Developmental Neuroscience*, 43, pp. 70–77. doi: 10.1016/j.ijdevneu.2015.04.003.

Falkmer, T. *et al.* (2013) 'Diagnostic procedures in autism spectrum disorders: A

systematic literature review', *European Child and Adolescent Psychiatry*. Dr. Dietrich Steinkopff Verlag GmbH and Co. KG, pp. 329–340. doi: 10.1007/s00787-013-0375-0.

Finegold, S. M. *et al.* (2017) 'Detection of Clostridium perfringens toxin genes in the gut microbiota of autistic children', *Anaerobe*. Elsevier Ltd, 45, pp. 133–137. doi: 10.1016/j.anaerobe.2017.02.008.

Fouquier, J. *et al.* (2021) 'The Gut Microbiome in Autism: Study-Site Effects and Longitudinal Analysis of Behavior Change', *mSystems*. Edited by J. A. Gilbert. American Society for Microbiology Journals, 6(2). doi: 10.1128/msystems.00848-20.

Ganesh, B. P. *et al.* (2013) 'Commensal Akkermansia muciniphila Exacerbates Gut Inflammation in Salmonella Typhimurium-Infected Gnotobiotic Mice', *PLoS ONE*. Edited by B. Ryffel. PLoS One, 8(9), p. e74963. doi: 10.1371/journal.pone.0074963.

Gilbert, J. A. *et al.* (2018) 'Current understanding of the human microbiome', *Nature Medicine*. Nature Publishing Group, 24(4), pp. 392–400. doi: 10.1038/nm.4517.

Golubeva, A. V. *et al.* (2017) 'Microbiota-related Changes in Bile Acid & Tryptophan Metabolism are Associated with Gastrointestinal Dysfunction in a Mouse Model of Autism', *EBioMedicine*. Elsevier B.V., 24, pp. 166–178. doi: 10.1016/j.ebiom.2017.09.020.

Gondalia, S. V. *et al.* (2012) 'Molecular Characterisation of Gastrointestinal Microbiota of Children With Autism (With and Without Gastrointestinal Dysfunction) and Their Neurotypical Siblings', *Autism Research*, 5(6), pp. 419–427. doi: 10.1002/aur.1253.

Góra, B. *et al.* (2018) 'Toxin profile of fecal Clostridium perfringens strains isolated from children with autism spectrum disorders', *Anaerobe*. Academic Press, 51, pp. 73–77. doi: 10.1016/j.anaerobe.2018.03.005.

Halmos, E. P. *et al.* (2015) 'Diets that differ in their FODMAP content alter the colonic luminal microenvironment', *Gut*. BMJ Publishing Group, 64(1), pp. 93–100. doi: 10.1136/gutjnl-2014-307264.

Heiman, M. L. and Greenway, F. L. (2016) 'A healthy gastrointestinal microbiome is dependent on dietary diversity', *Molecular Metabolism*. Elsevier GmbH, 5(5), pp. 317–320. doi: 10.1016/j.molmet.2016.02.005.

Heisel, T. *et al.* (2015) 'Complementary amplicon-based genomic approaches for the study of fungal communities in humans', *PLoS ONE*. Edited by M. D. Lenardon. Public Library of Science, 10(2), p. e0116705. doi: 10.1371/journal.pone.0116705.

Hill, C. J. *et al.* (2017) 'Evolution of gut microbiota composition from birth to 24 weeks in the INFANTMET Cohort', *Microbiome*. BioMed Central, 5(4). doi: 10.1186/s40168-016-0213-y.

Huang, L. S. *et al.* (2018) 'Analysis of fecal microbiota in patients with functional

constipation undergoing treatment with synbiotics', *European Journal of Clinical Microbiology & Infectious Diseases 2018 37:3*. Springer, 37(3), pp. 555–563. doi: 10.1007/S10096-017-3149-7.

Hyman, S. L. *et al.* (2020) 'Identification, evaluation, and management of children with autism spectrum disorder', *Pediatrics*. American Academy of Pediatrics, 145(1). doi: 10.1542/PEDS.2019-3447/36917.

Iglesias–Vázquez, L. *et al.* (2020) 'Composition of gut microbiota in children with autism spectrum disorder: A systematic review and meta-analysis', *Nutrients*. MDPI AG. doi: 10.3390/nu12030792.

Jones, J. *et al.* (2021) 'Fecal sample collection methods and time of day impact microbiome composition and short chain fatty acid concentrations', *Scientific Reports*. Nature Publishing Group, 11(1), pp. 1–13. doi: 10.1038/s41598-021-93031-z.

Kang, D.-W. *et al.* (2013) 'Reduced Incidence of Prevotella and Other Fermenters in Intestinal Microflora of Autistic Children', *PLoS ONE*. Edited by J. A. Gilbert. Public Library of Science, 8(7), p. e68322. doi: 10.1371/journal.pone.0068322.

Kang, D. W. *et al.* (2017) 'Microbiota Transfer Therapy alters gut ecosystem and improves gastrointestinal and autism symptoms: An open-label study', *Microbiome*. Microbiome, 5(1), pp. 1–16. doi: 10.1186/s40168-016-0225-7.

Kõljalg, U. *et al.* (2005) 'UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi', *New Phytologist*. John Wiley & Sons, Ltd (10.1111), 166(3), pp. 1063–1068. doi: 10.1111/j.1469-8137.2005.01376.x.

Kong, X. *et al.* (2019) 'New and preliminary evidence on altered oral and gut microbiota in individuals with autism spectrum disorder (ASD): Implications for ASD diagnosis and subtyping based on microbial biomarkers', *Nutrients*. MDPI AG, 11(9). doi: 10.3390/nu11092128.

Kovtun, A. S. *et al.* (2020) 'Antibiotic Resistance Genes in the Gut Microbiota of Children with Autistic Spectrum Disorder as Possible Predictors of the Disease', *Microbial Drug Resistance*. Mary Ann Liebert Inc., 26(11), pp. 1307–1320. doi: 10.1089/mdr.2019.0325.

Krajmalnik-Brown, R. *et al.* (2015) 'Gut bacteria in children with autism spectrum disorders: challenges and promise of studying how a complex community influences a complex disease', *Microbial Ecology in Health & Disease*, 26. doi: 10.3402/mehd.v26.26914.

Levy, M. *et al.* (2017) 'Dysbiosis and the immune system', *Nature Reviews Immunology*. Nature Publishing Group, pp. 219–232. doi: 10.1038/nri.2017.7.

Lord, C. *et al.* (2012) *Autism Diagnostic Observation Schedule, 2nd Edition (ADOS-2)*, *Western Psychological Services*. Torrence, CA. Available at:

https://www.carautismroadmap.org/autism-diagnostic-observation-schedule-ados/ (Accessed: 21 September 2021).

Louis, P. and Flint, H. J. (2017) 'Formation of propionate and butyrate by the human colonic microbiota', *Environmental Microbiology*. Blackwell Publishing Ltd, pp. 29–41. doi: 10.1111/1462-2920.13589.

Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*. BioMed Central Ltd., 15(12), p. 550. doi: 10.1186/s13059-014-0550-8.

Lukiw, W. J. (2020) 'Gastrointestinal (GI) Tract Microbiome-Derived Neurotoxins—Potent Neuro-Inflammatory Signals From the GI Tract via the Systemic Circulation Into the Brain', *Frontiers in Cellular and Infection Microbiology*. Frontiers Media S.A., 10, p. 22. doi: 10.3389/fcimb.2020.00022.

Ma, B. *et al.* (2019) 'Altered gut microbiota in Chinese children with autism spectrum disorders', *Frontiers in Cellular and Infection Microbiology*. Frontiers Media S.A., 9(MAR). doi: 10.3389/fcimb.2019.00040.

Martin, C. R. *et al.* (2018) 'The Brain-Gut-Microbiome Axis', *CMGH*. Elsevier, pp. 133–148. doi: 10.1016/j.jcmgh.2018.04.003.

Martin, M. (2011) 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet.journal*. EMBnet Stichting, 17(1), p. 10. doi: 10.14806/ej.17.1.200.

Matamoros, S. *et al.* (2013) 'Development of intestinal microbiota in infants and its impact on health', *Trends in Microbiology*, 21(4), pp. 167–173. doi: 10.1016/j.tim.2012.12.001.

McKnight, D. T. *et al.* (2019) 'microDecon: A highly accurate read-subtraction tool for the post-sequencing removal of contamination in metabarcoding studies', *Environmental DNA*. John Wiley & Sons, Ltd, 1(1), pp. 14–25. doi: 10.1002/edn3.11.

Mesa, M. D. *et al.* (2020) 'The evolving microbiome from pregnancy to early infancy: A comprehensive review', *Nutrients*. MDPI AG. doi: 10.3390/nu12010133.

Mínguez Pérez, M. and Benages Martínez, A. (2009) 'The Bristol scale - a useful system to assess stool form?', *Revista espanola de enfermedades digestivas*, 101(5), pp. 305–11. doi: 10.4321/S1130-01082009000500001.

Mullen, E. M. (1995) *Mullen Scales of Early Learning (Book, 1995) [WorldCat.org]*. AGS ed. Minnesota: Circle Pines, Minnesota. Available at: https://www.worldcat.org/title/mullen-scales-of-early-learning/oclc/860820407 (Accessed: 2 July 2021).

Parker, A., Fonseca, S. and Carding, S. R. (2020) 'Gut microbes and metabolites as

modulators of blood-brain barrier integrity and brain health', *Gut Microbes*. Taylor & Francis, pp. 135–157. doi: 10.1080/19490976.2019.1638722.

Parks, D. H. *et al.* (2014) 'STAMP: Statistical analysis of taxonomic and functional profiles', *Bioinformatics*. Oxford University Press, 30(21), pp. 3123–3124. doi: 10.1093/bioinformatics/btu494.

Parracho, H. M. R. T. *et al.* (2010) 'A double-blind, placebo-controlled, crossover-designed probiotic feeding study in children diagnosed with autistic spectrum disorders', *International Journal of Probiotics and Prebiotics*, 5(2), pp. 69–74. Available at: www.newcenturyhealthpublishers.com (Accessed: 20 May 2021).

Pham, V. T. *et al.* (2021) 'Antioxidant vitamins and prebiotic fos and xos differentially shift microbiota composition and function and improve intestinal epithelial barrier in vitro', *Nutrients*. MDPI AG, 13(4). doi: 10.3390/NU13041125/S1.

Pulikkan, J. *et al.* (2018) 'Gut Microbial Dysbiosis in Indian Children with Autism Spectrum Disorders', *Microbial Ecology*. Springer New York LLC, 76(4), pp. 1102–1114. doi: 10.1007/s00248-018-1176-2.

Reichardt, N. *et al.* (2014) 'Phylogenetic distribution of three pathways for propionate production within the human gut microbiota', *The ISME Journal 2014 8:6*. Nature Publishing Group, 8(6), pp. 1323–1335. doi: 10.1038/ismej.2014.14.

Ríos-Covián, D. *et al.* (2016) 'Intestinal short chain fatty acids and their link with diet and human health', *Frontiers in Microbiology*. Frontiers Media S.A., 7(185), p. 185. doi: 10.3389/fmicb.2016.00185.

Saad, R. J. *et al.* (2010) 'Do stool form and frequency correlate with whole-gut and colonic transit results from a multicenter study in constipated individuals and healthy controls', *American Journal of Gastroenterology*, 105(2), pp. 403–411. doi: 10.1038/ajg.2009.612.

Sandler, R. H. *et al.* (2000) 'Short-term benefit from oral vancomycin treatment of regressive-onset autism', *Journal of Child Neurology*. J Child Neurol, 15(7), pp. 429–435. doi: 10.1177/088307380001500701.

Sharon, G. *et al.* (2016) 'The Central Nervous System and the Gut Microbiome', *Cell*. Cell Press, 167(4), pp. 915–932. doi: 10.1016/J.CELL.2016.10.027.

Shumway, S. *et al.* (2012) 'The ADOS Calibrated Severity Score: Relationship to Phenotypic Variables and Stability over Time', *Autism Research*. Autism Res, 5(4), pp. 267–276. doi: 10.1002/aur.1238.

Strati, F. *et al.* (2017) 'New evidences on the altered gut microbiota in autism spectrum disorders', *Microbiome*. BioMed Central, 5(1), p. 24. doi: 10.1186/s40168-017-0242-1.

Tailford, L. E. *et al.* (2015) 'Mucin glycan foraging in the human gut microbiome',

Frontiers in Genetics. Frontiers, 0(FEB), p. 81. doi: 10.3389/FGENE.2015.00081.

Tap, J. *et al.* (2009) 'Towards the human intestinal microbiota phylogenetic core', *Environmental Microbiology*. Environ Microbiol, 11(10), pp. 2574–2584. doi: 10.1111/j.1462-2920.2009.01982.x.

Turner, S. *et al.* (1999) 'Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis', *Journal of Eukaryotic Microbiology*. John Wiley & Sons, Ltd (10.1111), 46(4), pp. 327–338. doi: 10.1111/j.1550-7408.1999.tb04612.x.

Velásquez-Jiménez, D. *et al.* (2021) 'Phenolic compounds that cross the blood-brain barrier exert positive health effects as central nervous system antioxidants', *Food and Function*, pp. 10356–10369. doi: 10.1039/d1fo02017j.

Wang, L. *et al.* (2012) 'Elevated fecal short chain fatty acid and ammonia concentrations in children with autism spectrum disorder', *Digestive Diseases and Sciences*. Springer US, 57(8), pp. 2096–2102. doi: 10.1007/s10620-012-2167-7.

Wemheuer, F. *et al.* (2020) 'Tax4Fun2: Prediction of habitat-specific functional profiles and functional redundancy based on 16S rRNA gene sequences', *Environmental Microbiomes*. BioMed Central Ltd., 15(1). doi: 10.1186/s40793-020-00358-7.

Whitehouse, A. J. O. *et al.* (2021) 'Effect of Preemptive Intervention on Developmental Outcomes Among Infants Showing Early Signs of Autism: A Randomized Clinical Trial of Outcomes to Diagnosis', *JAMA Pediatrics*, pp. e213298–e213298. doi: 10.1001/jamapediatrics.2021.3298.

Willis, A. D. (2019) 'Rarefaction, Alpha Diversity, and Statistics', *Frontiers in Microbiology*. Frontiers Media S.A., 10(OCT), p. 2407. doi: 10.3389/fmicb.2019.02407.

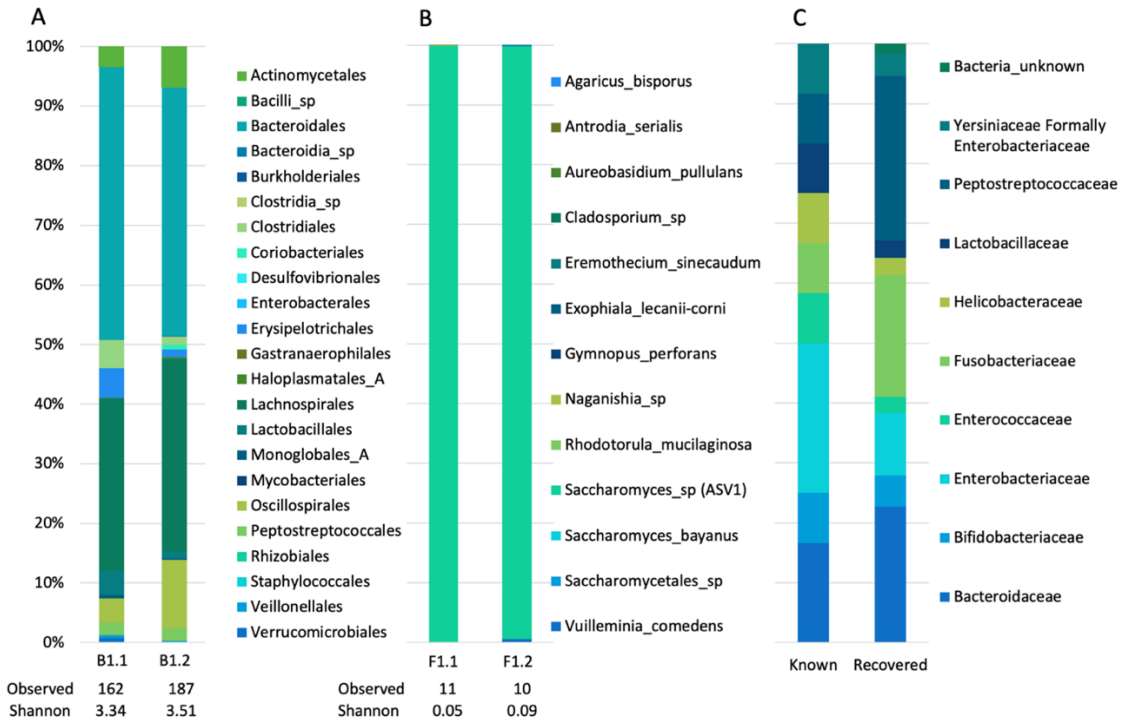
Xu, Z. and Knight, R. (2015) 'Dietary effects on human gut microbiome diversity', *British Journal of Nutrition*, 113, pp. S1–S5. doi: 10.1017/S0007114514004127.

Yap, C. X. *et al.* (2021) 'Autism-related dietary preferences mediate autism-gut microbiome associations', *Cell*. Elsevier, 184(24), pp. 5916-5931.e17. doi: 10.1016/j.cell.2021.10.015.

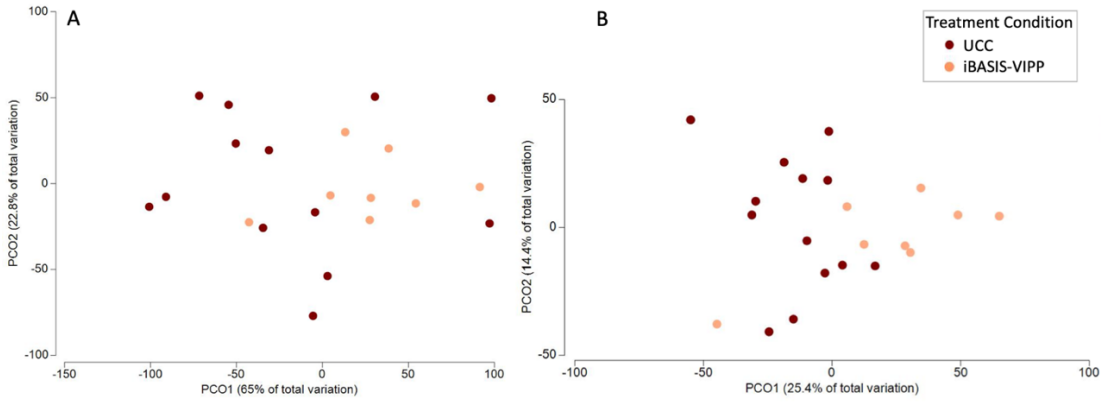
Zhao, Y. and Yu, Y. B. (2016) 'Intestinal microbiota and chronic constipation', *SpringerPlus*. Springer, p. 1130. doi: 10.1186/s40064-016-2821-1.

3.9 Supplementary Material

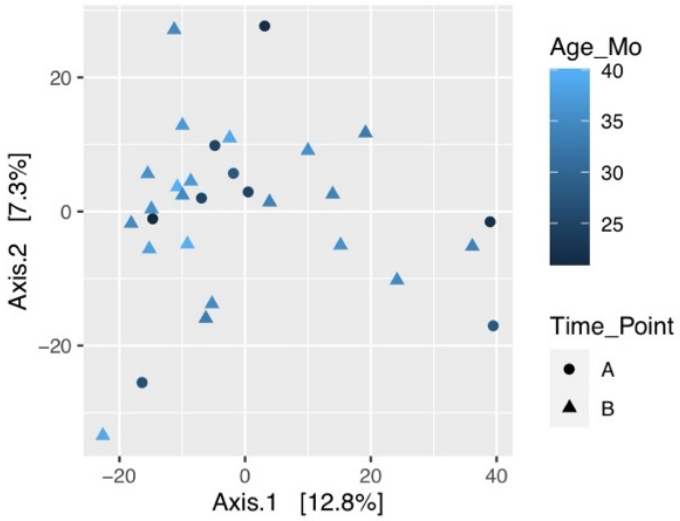
3.9.1 Supplementary figures



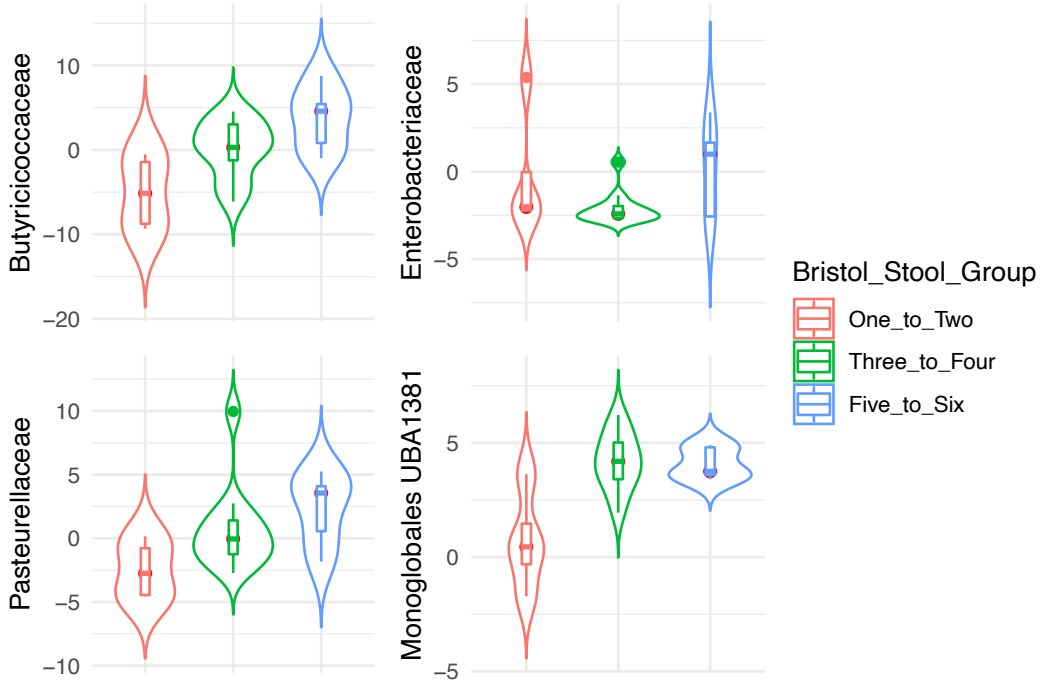
Supplementary Figure 3.1. *Replicate sampling of a single individual, showing the Bacteriome at the level of Order (A), and mycobiome at the level of ASV, with species assignments shown (B). The mycobiome replicate samples include ASV 1, which dominated both replicates F1.1 and F1.2 with 165,517, and 105,517 reads per sample respectively. Richness and Diversity were calculated on filtered non-rarefied counts. A Bacterial mock community (C) shows the proportion of both the known composition of the mock community positive control at the family level, and the recovered mock community sample composition.*



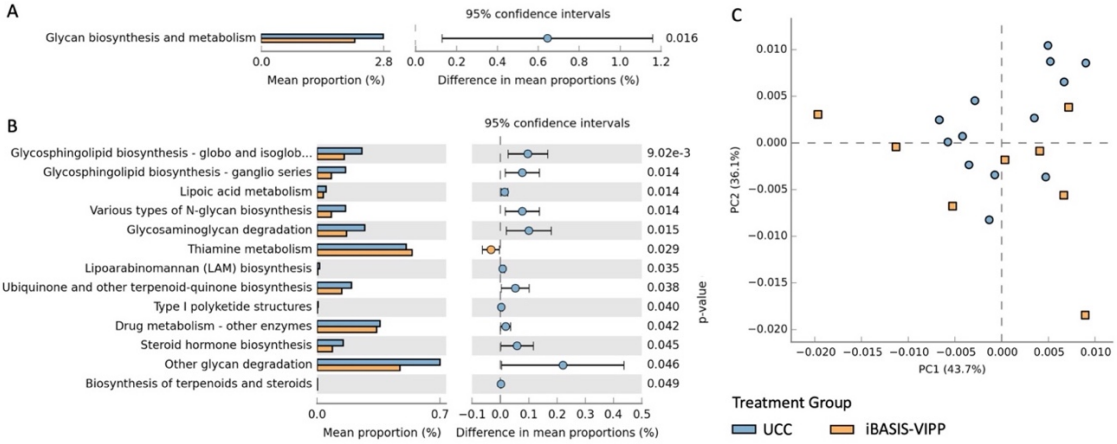
Supplementary Figure 3.2. *PCoA plots showing clustering of bacterial communities observed by treatment in the AICES study, at the level of phyla (A) and genus (B). Differences were calculated on Euclidian distance of CLR transformed counts.*



Supplementary Figure 3.3. *Clustering of bacteriome samples according to PCoA using Euclidian distances of CLR transformed counts. Samples are identified by age in months, and the sampling timepoint.*



Supplementary Figure 3.4. *Four bacterial families with differential abundance (CLR) according to stool form. The number of children in each stool form group was 4, 12, and 5 in the one-two, three-four, and five-six groups respectively.*



Supplementary Figure 3.5. *Differences in predicted pathway class (A), and individual pathways (B) associated with treatment type, and clustering via PCA of individual pathways based on treatment group (C). All plots were based on the filtered core metabolic predicted pathways.*

3.9.2 Supplementary tables

Supplementary Table 3.1. *Distribution of reads in each microbial dataset after quality filtering, and in the dominant microbiome subset.*

	Bacteriome	Mycobiome
minimum per sample	55,345	1,095
maximum per sample	114,085	286,688
Average	76,462	113,968
total reads	2,37,351	3,305,080
Total ASVs	1165	231
ASVs post filtering	584	207
ASVs in dominant Microbiome	372	136

Supplementary Table 3.2. *Taxa observed with differential abundance between preemptive treatment groups. Significant differences were determined using the Mann-Whitney U method using only dominant bacterial taxa.*

enriched in iBASIS-VIPP treatment group		enriched in UCC treatment group	
Order	Pvalue (FDR)	**Genus**	Pvalue (FDR)
Lachnospirales	0.00 (0.12)	*Enterocloster*	0.01 (0.56)
		Flavonifractor	0.02 (0.56)
Family		*Oscillibacter*	0.03 (0.56)
Lachnospiraceae	0.00 (0.21)		
Anaerovoracaceae	0.03 (0.74)	**Species**	
		Enterocloster sp	0.00 (0.12)
Genus		*Akkermansia* sp	0.00 (0.19)
Lachnospiraceae sp	0.00 (0.40)	*Flavonifractor* sp	0.02 (0.53)
Pasteurellaceae sp	0.03 (0.56)	*Oscillibacter* sp	0.03 (0.53)
Blautia	0.03 (0.56)		
Faecalibacillus	0.04 (0.56)		
Anaerovoracaceae, UBA1191	0.04 (0.56)		
Species			
Gemmiger sp	0.00 (0.12)		
Lachnospiraceae sp	0.00 (0.19)		
Dorea longicatena	0.02 (0.53)		
Pasteurellaceae sp	0.03 (0.53)		
Blautia obeum	0.03 (0.53)		
Lachnospiraceae, TF01 11 sp	0.04 (0.53)		
Anaerovoracaceae, UBA1191 sp	0.04 (0.53)		
Faecalibacillus sp	0.04 (0.53)		
Anaerobutyricum soehngenii	0.04 (0.53)		

Supplementary Table 3.3. *Genera identified as having differential abundance between lose (Bristol Stool number 1-2), normal (Bristol Stool number 3-4), and firm (Bristol Stool number 5-6) stool. P values and FDR corrected values are shown after Kruskal-Wallis test.*

Genera depleted in Firm stool	Pvalue (FDR)
Lachnospiraceae, *Agathobacter*	0.03 (0.44)
Butyricicoccaceae	0.03 (0.44)
Monoglobales A CAG41	0.02 (0.44)
Lachnospiraceae, CHKCI001	0.01 (0.44)
Pasteurellaceae *Haemophilus*	0.05 (0.59)
Lachnospiraceae, UBA9502	0.02 (0.44)

Genera depleted in normal stool	
Enterobacteriaceae, *Enterobacter*	0.03 (0.44)
Erysipelatoclostridiaceae, *Longibaculum*	0.03 (0.44)
Anaerovoracaceae, UBA1191	0.02 (0.44)

Supplementary Table 3.4. *Bacterial and Fungal taxa with differential abundance between CSS groups. Only dominant bacterial and fungal taxa were included.*

Taxa enriched in NDC vs ASD	p-value (FDR)
Bacteroides uniformis	0.02 (0.80)
CHKCI001 gen	0.02 (0.67)
CHKCI001 sp	0.02 (0.80)
Enterobacter D	0.00 (0.28)
Enterobacter D sp	0.00 (0.43)
Negativicutes *gen*	0.01 (0.45)
Negativicutes sp	0.01 (0.70)
Phocaeicola	0.02 (0.67)
Saccharomycetales gen	0.02 (0.34)
Saccharomycetales sp	0.02 (0.45)

Taxa enriched in ASD vs NDC	
Bacteroides caccae	0.03 (0.80)
Clostridium sp	0.03 (0.80)
Parabacteroides	0.04 (0.88)
Parabacteroides sp	0.04 (0.80)

Taxa enriched in NDC vs NAASD	
Bifidobacteriaceae gen	0.04 (0.96)
Bifidobacteriaceae sp	0.04 (0.96)
Enterobacter D gen	0.03 (0.96)
Enterobacter D sp	0.03 (0.96)
Muricomes gen	0.02 (0.96)
Muricomes sp	0.02 (0.96)
Veillonella sp	0.03 (0.96)

Supplementary Table 3.5. *Individual pathways which differ significantly prior to FDR correction between CSS groups using Welch's t-test.*

ASD vs NAASD groups	p-value	Corrected p-value
Biosynthesis of type II polyketide backbone	<0.001	0.062
Carbon fixation pathways in prokaryotes	<0.001	0.175
Drug metabolism - other enzymes	<0.001	0.191
Tetracycline biosynthesis	<0.001	0.035
ASD vs NDC groups		
Flavone and flavonol biosynthesis	0.001	1.43
Pentose and glucuronate interconversions	0.03	2.46
NAASD vs NDC groups		
Atrazine degradation	0.026	2.019
Polycyclic aromatic hydrocarbon degradation	0.017	2.619
Sulfur metabolism	0.035	1.79

3.9.3 Supplementary results

Bacterial data resulted from all 30 faecal samples, with a minimum of 55,418 reads per sample, and 1,165 ASVs from eight phyla. The Fungal data resulted from 29 samples and a minimum of 1095 reads per sample with 231 ASVs from two phyla prior to prevalence filtering (Sup Table 1). A single sample was also processed in duplicate to visualize the bias due to library preparation for both bacteria and fungi (Sup Figure 1). At the level of order, bacteriome communities showed high similarity, with the read count of Oscillospirales, most variable between the replicate samples (B1.1 = 4%; B1.2 = 11.5%). The mycobiome was compared at the level of ASV (species IDs shown) because of the lower diversity of this community. The total read count was much higher in the fungal samples >100,000 reads, and the communities were both dominated by ASV1, with a read counts 100x higher than all other ASVs. The mycobiome was also more variable with 57,587 more reads in F1.1 compared to F1.2 verses the bacterial replicates where B1.1 had 2,251 more reads than B1.2, which is typical of sequencing technology which generates stochastic variation in sequence counts. The American Type Culture Collection (ACTT) bacterial mock community (https://www.atcc.org/products/msa-1006), was also sequenced as a positive control (Sup Figure 1), containing 12 species in even concentrations from 9 families including multiple species in Bacteroidaceae, and Enterobacteriaceae. With greater than 500 reads, 17 ASVs were recovered from the sample mock community, from the correct nine families. Twelve ASVs were assigned to the correct 12 genera, and 11 ASVs were assigned to the correct species at 100% identify; with 7 of these ASVs matching to two or more

– Chapter 4 –
Assessment of the temporal changes to the maternal and infant microbiome in response to a maternal prebiotic supplement during pregnancy and lactation

4.1. Prologue

Chapter 2 explored the variability in faecal microbiome composition due to sample collection methods, as well as the benefit of using control samples. In chapter 3, these sample collection and data processing recommendations were used to explore the microbiome of very young children who were showing high behavioral risk of developing autism. Together these two studies demonstrated both a high level of attention to data integrity, and descriptive reporting to allow for study replication. Furthermore, chapter 3 described fine scale shifts to the gut environment prior to a formal diagnosis for autism, which provided a framework for the investigation of the microbiome data in chapter 4.

In the following chapter I assess the effectiveness of a maternal prebiotic supplement on both the maternal and infant microbiome composition, and SCFA concentrations. The human gut microbiome is most susceptible to change in early life, and therefore, supporting its development early (during pregnancy and lactation) would likely be most beneficial. The prebiotic supplement was assessed in mothers who were enrolled in a double-blinded randomized controlled trial, and the data presented in this chapter is part of a larger study (Palmer *et al*., 2022). Therefore, the data will be made available after publication. Lastly, with 74 mother-infant pairs, this chapter contains the largest dataset in this thesis, and I would like to extend my deepest gratitude to Georgia Nester and Mathew Heydenrych for their significant contribution to sample processing and DNA extraction.

4.2 Abstract

During pregnancy, the foetal gut microbiome may begin the important process of initial colonisation which will have a long-lasting impact on the eventual structure and resilience of the gut microbiome. This early life period may be a critical time in life to establish both gut health, and long-term immunological health as the immune system develops under the influence of the microbiome. The maternal microbiome is not only an important potential seeding location for microbial colonizers, but also a source of microbial products that may be transferred to the foetus. Short chain fatty acids (SCFA) in particular are microbial metabolites that are known to promote gut and metabolic

health, and dietary fibre supplied by the diet determines in large part the production and absorption of SCFA in the gut. Soluble fibres including fructo-oligosaccharide, and galacto-oligosaccharide resist digestion by the host, and are selectively fermented in the colon by commensal bacteria. Therefore, the composition of the maternal microbiome, and factors such as diet during pregnancy may carry forward and also impact the infant microbiome and developing immune system. Furthermore, there is wide agreement that the first 100 days of life are the most important for the development of the infant microbiome. The diet an infant is first exposed to will play an important role in this process and breastmilk – which changes over time and with maternal dietary habits – contains both microbes and human milk oligosaccharides that establish and support the growth of mutualistic microorganisms. Supporting the development of the gut microbiome may therefore also benefit the development of the immune system which shows characteristic shifts associated with autoimmune disease. Due to the encouraging results demonstrated using prebiotics in adults and infants there has been interest in trialling fibre supplementation during pregnancy for the benefit of the infant microbiome and immune system. In this study I aimed to determine the effect of a maternal prebiotic fibre supplement on both the maternal and infant microbiome composition based on illumina sequencing of the bacterial 16S rRNA gene (V4), and SCFA concentrations compared to a placebo. The prebiotic fibre was assessed based on the microbiome of 74 participants (which includes the mother and infant) who were enrolled in a double-blinded randomized controlled trial. I found significant differences in both maternal (PERMANOVA $p < 0.0001$, psudo-F 4.23), and infant (PERMANOVA $p < 0.001$; psudo-F 1.7) microbial beta diversity between the prebiotic and placebo groups over the intervention period. The prebiotic fibre was found to increase the abundance of commensal *Bifidobacteria* ($p \leq 0.02$) and *Parabacteroides merdae* (DESeq, padj < 0.001) in the maternal microbiome, as well as *Aerococcus* spp (Closteridia) in both maternal and infant microbiomes (ANCOM-BC padj ≤ 0.05). The shifts in SCFA concentration over pregnancy were also significantly different between the prebiotic and placebo groups, and acetic acid significantly increased ($p = 0.008$; Pseudo-F = 7.09) in the prebiotic group but not the placebo group after the start of the intervention. The infant microbiome was also observed to go through a period of significant expansion in alpha diversity (ANOVA, p>0.001). which coincided with a significant increase in butyric and propionic acid concentrations after the introduction of solid foods (ANOVA, p>0.001).

4.3 Introduction

Pregnancy is now understood to be an important period for the development of the infant gut microbiome, and immune system. As the immune system develops under the influence of the microbiome, this early life period may be a critical time in life to

establish long term immunological health and gut health (Gray *et al.*, 2017). Gastrointestinal bacteria stimulate the development of the gut-associated lymphoid tissue, which is the largest collection of lymphocytes in the body (Faria, Reis and Mucida, 2017), and during this time the immune system is trained to be tolerant to aspects of the environment. During pregnancy, both the maternal gut and vaginal microbiomes undergo dramatic, yet predictable compositional changes (Mesa *et al.*, 2020), and there is now evidence summarized in two recent reviews which suggest that maternal gut health during pregnancy influences the health of the infant (Gray *et al.*, 2017; Mesa *et al.*, 2020). This includes susceptibility to metabolic diseases, infection, and allergic disease as well as adverse pregnancy outcomes (Mesa *et al.*, 2020).

The assumption that an infant is first exposed to the microbial world at birth is being challenged with recent findings of microbial genetic material recovered from of the placenta, anionic fluid, and meconium. Well described commensal bacteria including *Lactobacillus*, and members of Enterobacteriaceae have been detected in the placenta (Mesa *et al.*, 2020), while predominant members including *Enterobacter*, *Escherichia/ Shigella* and *Propionibacterium* have been detected in the amniotic fluid (Collado *et al.*, 2016). Further support comes from isolating live bacteria from the first meconium, predominantly *Enterococcus fecalis, E. coli,* and *Staphylococcus epidermidis* (Jiménez *et al.*, 2008). These microbes are theorized to have been swallowed with amniotic fluid by the foetus, as evidenced by further investigations in animal (mice) models. Here, genetically labelled bacterial strains were recovered from the meconium of mice pups after oral administration of the bacterial strain to the pregnant mother. This observation provides evidence that maternal gastrointestinal bacteria may travel a hematogenous route to reach the foetal digestive tract (Jiménez *et al.*, 2008). While the evidence does suggest we should be open to the possibility of a non-sterile foetal gut, there is still debate regarding the origin of genetic material recovered from these sources. Contamination of low biomass samples by exogenous bacterial DNA is a major concern, and therefore the use of strict protocols to avoid or eliminate contamination is essential (Stinson *et al.*, 2019).

The maternal microbiome is not only an important potential seeding location for microbial colonizers, but also a source of microbial products that may be transferred to the foetus, depending on the regulatory action of the placenta. Genetic material and metabolites that originate in the gut microbiome can be absorbed into the maternal blood, and depending on the regulatory action of the placenta may act as important environmental exposures which stimulate foetal immune programming (Gray *et al.*, 2017). SCFA in particular are known to promote tolerance of T lymphocytes by regulating the antigens presented to them by dendritic cells (Gray *et al.*, 2017). Furthermore, in mice, maternal derived SCFA were found to interact with the sympathetic nervous system, intestinal tract, and pancreas of the foetal mice, which

impacted energy homeostasis, and regulation of insulin levels in offspring (Mu *et al.*, 2015). Therefore, the composition of the maternal microbiome, and factors directly affecting it during pregnancy may carry forward and also impact the infant microbiome and developing immune system (Gray *et al.*, 2017).

Diet has been consistently shown to be a strong influencer of the gut microbiome, with animal protein, and fibre being two of the most influential nutrient types on the composition of the microbial community (Ríos-Covián *et al.*, 2016; Beaumont *et al.*, 2017; Martin and Li, 2017; Moles and Otaegui, 2020). High protein diets have been shown to shift the activity of the microbial community from carbohydrate fermentation to protein fermentation, which changes both the pH and the composition of microbes in the gut (Beaumont *et al.*, 2017). Dietary fibre is typically the main source of carbon utilized by the microbiota in the human gut (Scott, Duncan and Flint, 2008), and therefore, fibre supplied by the diet determines in large part the production and absorption of SCFA in the gut. The benefits of fibre-based dietary supplements on the microbiome and SCFA concentrations in adults have been discussed in a number of reviews (Scott, Duncan and Flint, 2008; Christodoulides *et al.*, 2016; Yan *et al.*, 2021). Soluble fibres including inulin-type prebiotics, fructo-oligosaccharide (FOS), and galacto-oligosaccharide (GOS) have been studied extensively in more recent times. These substrates are also commonly used as prebiotic fibres, and each has been reviewed for the ability to maintain a healthy gut ecosystem by benefiting commensal bacteria, regulating levels of SCFA, and reducing gut pH (Sawicki *et al.*, 2017). FOS and GOS resist digestion in the upper digestive tract, and are selectively fermented in the colon by resident bacteria, most notably *Bifidobacterium* spp (Liu *et al.*, 2017). This allows for selection of microbes that can utilise this exclusive energy source, and may also benefit other microorganisms through cross-feeding. Diet and fibre content will also impact weight and weight gain through regulation of metabolism and reductions in blood glucose concentrations (Mayengbam *et al.*, 2019). Therefore, maintaining appropriate fibre intake during pregnancy may also assist mothers in gaining appropriate weight during pregnancy, which will optimize health outcomes for both mother and infant (Moore Simas *et al.*, 2013).

During pregnancy, the composition of the gut microbiome can also be negatively affected by any antibiotics that are taken. Antibiotics reduce both richness and diversity, however a healthy established community is usually quite resilient to moderate doses of antibiotics (Martínez *et al.*, 2018). During pregnancy shifts to the maternal microbiome resulting from antibiotics could theoretically impact the colonization process by limiting microorganisms available for seeding the infant gut. Also, antibiotics used during pregnancy will cross the placenta and enter the foetal bloodstream where they may further impact microbial assembly (Pacifici, 2006). Due to the effect antibiotics have on the composition of the microbial community, they may impact negatively on the

development of the immune system, and the use of antibiotics during pregnancy has been shown to be associated with an increased risk of asthma (Stensballe *et al.*, 2013).

After birth, mother and infant are in intimate-close contact, and while not all researchers agree that the pregnancy period seeds the microbiome, there is wide agreement that the first 100 days of life are the most important for the development of the infant microbiome (Gritz and Bhandari, 2015; Ferretti *et al.*, 2018; Mesa *et al.*, 2020). Birth mode is a strong contributing factor for the development of the infant microbiome, and caesarean born infants have been shown to have reduced abundance of *Bacteroides* in their gut for up to 6-months after birth (Galazzo *et al.*, 2020). Caesarean birth is often accompanied by prophylactic antibiotics, preferentially given prior to surgery, which will allow them to cross the placenta. Caesarean and vaginal birth modes also differ with respect to exposure to the first bacterial community, with caesarean born infants exposed to skin and air microbial communities, while vaginally born infants are first exposed to the vaginal microbiome of the mother (Dominguez-Bello *et al.*, 2016).

Breastfeeding confers many health benefits to both the mother and infant (Turck *et al.*, 2013), and for the first 6-months of life is the recognised normative standard for infant nutrition (Ballard and Morrow, 2013). Breast milk is composed of potentially thousands of bioactive components (Ballard and Morrow, 2013), many of which are multifunctional, working not only as enzymes, growth factors, and anti-inflammatory agents, but also glycoproteins that act as antimicrobials by binding directly to pathogens including *Escherichia coli*, *Salmonella* spp. *Helicobacter pylori*, and *Burkholderia cepacia*as (Cacho and Lawrence, 2017). Breastmilk also contains both microbes and human milk oligosaccharides (HMOs) that act as pre- and pro-biotics respectively to establish and support the growth of mutualistic microorganisms (Cacho and Lawrence, 2017). Additionally, the composition of breastmilk changes over time, and with maternal dietary habits (Bravi *et al.*, 2016). Infants fed with formula that was supplemented with GOS and FOS prebiotic fibre have been found to have a higher concentration of secretory IgA, and a higher percentages of Bifidobacteria compared to a control group receiving formula without the fibre supplement (Scholtens *et al.*, 2008). In another RCT, trialling GOS/FOS supplemented infant formula an increase in the CFUs of Bifidobacteria in infants' stool was found, as well as significantly reduced incidence of atopic dermatitis compared to the control group (Moro *et al.*, 2006).

The gut microbiota shows characteristic shifts associated with a range of diseases, as well as autoimmune pathological conditions, especially in early life, that are predicted to be contributing to the rise of non-communicable diseases (Gray *et al.*, 2017). Therefore, supporting the development of the gut microbiome in early life may also benefit the development of the immune system during this important period of life. Infant supplementation with prebiotic formula feeding has been a successful way to shift the

gut microbiome, however these effects may not be long lived. Six of the studies included in a recent review of infant formulas, discussing prebiotic supplementation, found differences in microbiota composition could be detected for up to 6-weeks, and differences in intestinal pH were detected up to a maximum of 16-weeks (Fabiano *et al.*, 2021). Using animal models, it has also been demonstrated that maternal supplementation may be an unexplored avenue to influence the developing infant gut microbiome.

Due to the encouraging results demonstrated using prebiotics in adults and infants – as well as the potential therapeutic use of diet during pregnancy – there has been interest in trialling this framework of fibre supplementation during pregnancy for the benefit of the infant microbiome and immune system. Therefore, mothers have been enrolled in a double-blinded randomized controlled trial investigating the effects of maternal prebiotic supplementation during pregnancy and breast feeding (Palmer *et al*., 2022).

I have specifically characterized the resident bacteria and the SCFA concentration in mother-infant pairs to determine the effect of maternal prebiotic fibre (GOS and FOS) supplementation on both the maternal microbiome composition, and its effect on the development of the infant microbiome. It is hypothesised that the prebiotic supplement will increase both the relative abundance of *Bifidobacterium* and the concentrations of acetic and butyric acid in the maternal gut of those mothers compared to a placebo, and these markers will be investigated withing the maternal gut. The positive effect of the prebiotic supplement on the material gut is also expected to support the development of the infant microbiome through exposure to healthy levels of SCFA during pregnancy. The presence of commensal colonisers is also predicted to be enriched in the infant gut in association with maternal prebiotic supplement, and a differential abundance assessment between prebiotic and placebo infants over time will be used to identify these bacteria. Lastly, it is hypothesised that the gut microbiome will change over the course of pregnancy, and to illustrate this enterotyping will be used.

4.4 Methods

4.4.1 Study design and faecal sample collection

Stool samples from pregnant mothers and their infants were collected as part of the SYMBA study, which aimed to recruit 652 pregnant women. The participants of the SYMBA study were randomized into two groups, one group received a daily prebiotic supplement consisting of GOS (8.1g) and FOS (0.9g), and the other group received a placebo consisting of maltodextrin (8.7g). Not all SYMBA study participants were included in the microbiome analysis due to financial and time related constraints. A subset of mother-infant pairs was selected based on time series completeness (i.e.

participants with samples at all maternal and infant timepoints), and to insure a minimum of 30 samples within the prebiotic and placebo groups each, from the available samples as of June 2019 (Table 4.1). The total sample size (n=52) required to test (with a power of 80% and two-tailed alpha = 0.05) for a significant difference in acetate concentrations was determined using G*Power (Faul *et al*., 2007); and the effect size (r = 0.81) was based on publicly available data (Liu *et al*., 2017). In this current study stool samples were received from 74 participants, with complete stool sampling all timepoints for 65 mother-infant pairs. At 20-weeks a baseline was established by collecting a single stool sample from each mother prior to consumption of the supplement or placebo. A single stool sample was then collected from each mother at 5 timepoints during and after birth, and a single infant stool sample was collected at 4 timepoints during the first year of life (Table 4.1). At each sampling point, mothers collected stool into three identical specimen jars, and froze them in the household freezer within 15 minutes of collection. The samples were then taken to the Telethon Kids Institute, located in the Perth Children's Hospital, Nedlands WA in a supplied Styrofoam pack within 7 days of collection. All samples were bio-banked at -80°C, at the Telethon Kids Institute until sample processing. Amplicon sequencing and statistical analyses were completed while blinded from the treatment group allocation.

Table 4.1. *Sample collection timepoints and number of samples collected as of June, 2019 within the SYMBA study, and the number of stool samples received and analysed in this current study at each timepoint.*

Collected for SYMBA as of June, 2019	20-weeks	28-weeks	36-weeks	2-months	4-months	6-months	12-months	Total
Maternal	246	219	205	128	136	100		1034
Infant				147	143	133	77	500
Maternal samples recieved								
Intervention group 1	32	32	32	32	32	32		192
Intervention group 2	39	39	39	39	39	38		233
Infant samples received								
Intervention group 1				32	32	32	32	128
Intervention group 2				39	39	39	37	154

4.2.2 Stool sample processing, DNA extraction and SCFA quantification

Immediately prior to DNA extraction, frozen stool samples were thawed at 4°C, and transferred on ice to a EuroClone Biological safety cabinet to limit potential contamination. The three tubes that made up one sample were combined in a larger sterile pot, and homogenized together for 30 seconds with a sterile plastic scoop. The Bristol Stool Form Scale (BSFS) was used to categorize each stool form during homogenization (Mínguez Pérez and Benages Martínez, 2009). Stool was then collected into separate tubes for each of two downstream analyses: metabarcoding 0.25 g ± 0.05 g and SCFA quantification 0.50 g ± 0.05 g. Samples for SCFA analysis were immediately frozen at -80°C following homogenisation, and then transferred on dry ice to the

Science Analytical Facility at Edith Cowan University, Western Australia for SCFA quantification using GC-MS as previously described (Jones et al., 2021). DNA was extracted by using QIAamp PowerFecal Pro DNA kit (QIAGEN, Hilden, Germany) and the IRT protocol for QIAcube (QIAGEN), according to the manufacturer's instructions with three modifications: (1) prior to adding stool sample, three 3.5mm glass beads (Biospec) were added to bead-beating tubes, (2) after step 1, tubes were vortexed for approximately 20 sec to incorporate beads and stool, (3) followed by heating at 65°C for 10 minutes. Contamination was accounted for using NTCs during the DNA extraction and PCR steps which were also processed following the same protocol as frozen stool samples. Due to the large number of NTCs created during DNA extraction, the DNA extracted from NTCs were combined in to a single "pooled extraction blank" which was uniquely tagged twice, and spiked into each sample library. The pooled extraction blanks and NTCs were used during data processing to remove contamination.

4.2.3 Library preparation and sequencing

Bacterial DNA, NTCs and mock communities (https://www.atcc.org/products/msa-1006) were amplified using 16S primers 515F (Turner *et al.*, 1999) and 806R (Caporaso *et al.*, 2011), each with a 8-10 bp unique barcode. Mock communities were sequenced to evaluate variability across sequencing runs. The PCR reactions were performed according to methods previously described (Jones *et al.*, 2021). PCR products were then pooled by amplification efficiency, and then blending in equimolar concentrations based on the concentration of each pool. Illumina compatible adaptors were ligated to the DNA fragments (Lucigen, Middelton, WI, USA), which were then were size selected using Pippin Prep (Sage Science). The DNA library was purified before sequencing using the QIAquick PCR purification column clean up kit (Qiagen, Germantown, MD), and sequencing was performed at Curtin University, Western Australia, using the Illumina MiSeq platform and V2 500 cycle kit (Illumina, San Diego, CA, USA) with 2 × 250 bp paired-end read lengths.

4.2.4 Deconvolution and data merging

Sequences were demultiplex with no mismatches allowed in the unique molecular barcodes, and then all non-biological regions were removed using Cutadapt (Martin 2011). Initial quality filtering included discarding sequences with ambiguous bases, or those with more than two expected errors; sequences were then merged with a minimum overlap of 60 bp allowing for one mismatch using DADA2 (Callahan *et al.*, 2016). Amplicons less than 251 bp were discarded and amplicon sequence variants (ASVs) were inferred using the pseudo-pooled method. Chimeric errors were removed using the default method, and sequences variants were classified to the using the Genome Taxonomy reference database (Version 202) formatted for use with DADA2 (https://zenodo.org/record/4735821#.YN18Om4RWis). Assignments to the genus level

were made with a minimum of 80% bootstrapping using “assignTaxonomy”, and species assignments were then added using “addSpecies”. The species assignments for the top 200 ASVs were confirmed by BLAST using the same databases, at 100% identity. ASVs with up to three matches were annotated to include all three potential species assignments, whereas any ASV with more than three identical matches was annotated with the genus name followed by “spp”. Identical matches to species id numbers were not included. Where taxa are not fully resolved to a lower rank, the lowest available rank name and sp., gen., or fam. have been annotated for each lower taxonomic level. To remove potential contamination from all samples, the package microDecon (McKnight *et al.*, 2019) was used. NTCs and pooled extraction blanks (n=7) were selected to inform the decontamination process using one run of the function “remove.count”. Lastly, any ASVs with unassigned phylum, or with a prevalence less than 1 read in 2% of samples were filtered out, reducing the number of ASVs from 6,587 to 1,462.

4.2.5 Statistical analysis

To allow all individuals to be included in correlation and grouping analyses proxy values were added in two circumstances. (1) where the age introduced to formula was greater than 1 year, or (2) the duration of breastfeeding was greater than year, 13 months was used to replace NA so the factor could be included in the model. Bristol stool number was also determined during sample processing, and stools were grouped by firm (1 to 2), normal (3 to 4) and loose (5 to 7) stool types. A 2-tailed McNemar’s test was used to compare the counts of Bristol stool form over time in maternal samples where data was available at all timepoints. The approximate gestational weight gain was calculated for mothers using the weight at randomization (20-weeks) and the weight at 36-weeks. Enterotypes were assigned to the maternal microbiome using proportions of genera as an input, and by fitting to 278 MEtaHIT samples as a reference http://enterotypes.org/. A 1-tailed McNemar’s test was used to compare the counts of Firmicutes dominated enterotypes at baseline to late pregnancy (36-weeks).

DistLM with step-wise selection criteria over 9999 repeats, was used to identify factors with a significant contribution to the composition of the maternal and infant microbiome. Factor information was missing from some individuals, therefore, to perform the DistLM, either the factor or the individual with missing information had to be removed from the analysis. To select maternal factors for inclusion the follows steps were taken (1) Any factor with complete participant information was included, (2) factors with missing information were ranked from most to least important (3) factors were added step-wise to the analysis until a maximum of 5% of the participants were removed, the remaining factors (fibre intake and parity) were not included. Correlations over 95% between similar factors were used to identify and remove redundant variables.

Infant factor selection followed the same 3 steps, and significant maternal factors were also included for consideration at step 1 along with infant factors. For visualization in dbRDA, only factors that contributed more than 1.4% in marginal tests were included.

Differential abundance testing was performed in DESeq2 (Love, Huber, & Anders, 2014) on filtered counts of ASVs. Differences between intervention and placebo groups were determined using both maternal age and antibiotic use as co-factors for maternal samples, and for infant samples gender (2m-6m) or breastfeeding (1-year). P values were corrected using Benjamini-Hochburg adjustment (BH) (Benjamini & Hochberg, 1995), and significance was set at $p = 0.05$. At higher taxonomic levels, ANCOM-BC (Lin and Peddada, 2020) was used to test for differences between intervention groups. After correcting for multiple testing using the BH adjustment method, an effect size (W statistic) cutoff of 0.1 was used to further refine significant results. The microbial abundance at individual timepoints was also bias corrected using ANCOM-BC, which introduces a sample-specific offset term which accounts for the sampling fraction which varies across samples. The bias-corrected abundance was then used to visualize the differentially abundant taxa between intervention groups. Lastly, a Wilcoxon signed rank test was used to test both the prebiotic and placebo groups for an increase in *Bifidobacterium* and *Lactobacillus*. The pre-birth comparisons were made from baseline (20-weeks) to 28- and 36-week timepoints with a BH adjustment, and post-birth comparisons were made from baseline to 2-, 4-, and 6-month timepoints with a BH correction.

A random forest analysis using 10 predictors over 5000 trees was performed on microbiome analyst (Dhariwal *et al.*, 2017) at the genus and family level. The model was used to both estimate the accuracy of predicting the classification of the group, and identify taxa that were most responsible for driving differences between the intervention groups. Pattern search available within Microbiome Analyst was also used to identify groups of genera with corelated enrichment (spearman rank correlation) to Bristol stool types.

Beta diversity was visualized using Principal coordinates analysis (PCoA), and was calculated using Euclidian distances of centre-log-ratio transformed counts of ASVs or higher taxonomic ranks. Differences in beta diversity were determined with PERMANOVA, performed PRIMER-e v7 (Anderson, Gorley and Clarke, 2008) with 9999 permutations under a reduced model (microbiome data), and type 3 sum of squares (Dhariwal et al., 2017). Interactions between the intervention and time were tested by nesting intervention within participant. One-way and two-way-mixed repeated measures ANOVAs were performed in R to compare alpha diversity estimates, and fold changes in diversity over time. Any participant who was not sampled over the entire study period, and extreme outliers (values above or below $Q3 \pm 3xIQR$) were excluded from

the repeated measures analysis. Normality was assumed with QQ plots showing good correlation, and. Homogeneity of variance and covariances of between-subject factors was checked with Levene's test, and Box's M-test respectively. Spearman correlation was used to determine the relationship between maternal and infant diversity indexes. Extreme outliers were removed from this analysis using the same method mentioned above.

Individual SCFA concentrations were analysed as mM/g of feces, as well as relative proportions of the sum of all acids quantified (acetic, propionic, butyric, isobutyric, valeric, and isovaleric acid). The concentration of any acid below the detection limit was set to 0. The variance of each SCFA concentration was stabilized by a log10 transformation with a constant of 1 was added to the concentration of propionic, butyric, isobutyric, valeric, and isovaleric acid. Repeated measures ANOVA was used to compare SCFA concentrations over time within each intervention group, and including only those participants who had SCFA data at all timepoints. Any participants without SCFA data at all timepoints was removed. Extreme outliers were identified for each, and normality, homogeneity of variance and homogeneity of covariances was confirmed using the methods previously described. Covariates determined using DistLM for SCFA data (probiotic use, and history of allergic disease for maternal samples) were included in the ANOVA model when testing individual SCFA concentrations over time. Pairwise tests were restricted to 4 timepoint comparisons ("20-28", "20-36","36-2","2-6") The p value of multiple post-hock comparisons was corrected using the Bonferroni method. Lastly, PERMANOVA was used to compare the resemblance of each of the three most predominant SCFA (acetic acid, propionic acid, and butyric acid) between the intervention and placebo groups over the study period. The same parameters as described previously for the PERMANOVA analysis were utilised, except unrestricted permutations of the raw data were allowed rather than permutations under a reduced model.

4.5 Results

4.5.1 Data quality and participant characteristics

Read depth is an important factor in accurate diversity estimates, and read depth is expected to vary considerably across sequencing runs. Across all sequencing runs, the minimum read depth was 20,820, and the maximum was 105,769. The sizable read depth was due to both the high percentage (minimum 83%) of reads with greater than 99.9% accuracy in base calls, and the proportion of amplicon clusters that were successfully generated during bridge amplification. A larger proportion of reads were lost from library LGN62 during read filtering, although during chimera removal, only a small proportion of reads were lost compared to the other libraries (Sup Table 4.1).

To ensure that high quality sequence data was consistently produced, several control samples were used during extraction, PCR, and across sequencing runs. Mock communities with a known composition were tagged and added at equimolar concentration to each sequencing run.

Prior to prevalence filtering, but after decontamination there were 19 ASVs that were consistently present across all replicates of the mock community. Of these, 17 ASVs had more than 100 read counts per sample on average. There were another 20 ASVs present in the mock community samples with between 10 and 100 counts per sample on average. These were likely reads with spurious errors that clustered as separate ASVs or poorly identified at the genus and species level using DADA2 rather than contamination. All ASVs with an average of 10 or more counts per sample were from one of the families of the mock community standard, and ASVs with less than 10 counts per sample were primarily Fermicutes. The relative proportions of families identified in each sample was consistent across replicates (Sup Figure 4.1, A), demonstrating good reproducibility across sequencing runs. All 12 species that were contained in the mock community standard were detected although, both *Clostridioides difficile* and *Fusobacterium nucleatum* were over-represented and *Enterobacter cloacae* were under-represented based on relative abundance (Sup Figure 4.1, B).

The number of total reads was expected to vary across sequencing runs, and with the largest sequencing depth, the samples within LGN68 had, the largest library size compared to all other sequencing runs (Sup Figure 4.2, A). To assess how this may affect alpha and beta diversity, eight randomly selected individuals were pooled together and then blended in equimolar concentrations into each sequencing run. Richness was affected by sequencing run, with the replicates in LGN59 having the highest number of ASVs (Sup Figure 4.2, B). Both alpha-diversity measured by Shannon index (Sup Figure 4.2, D), and beta-diversity between replicates were not affected by sequencing run (Sup Figure 4.2, C). The number of ASVs also varied by sequencing run (Sup table 4.1), and as anticipated, varied between maternal and infant samples. On average, maternal samples hosted 285.6 ± 69.6 ASV, and infant samples hosted 78.5 ± 48.7.

PCoA was used to observe Beta diversity across sequencing runs among all samples in this study. Maternal samples did not show any relationship to sequencing run, with sample density per sequencing run is distributed homogeneously across two-dimensional PC space (Figure 4.1, A). Infant samples also did not show any bias due to sequencing run, although the distribution on PC1 is less homogenous due to the significant differences between the different age groups, particularly the 1-year-old samples (Figure 4.1, B).

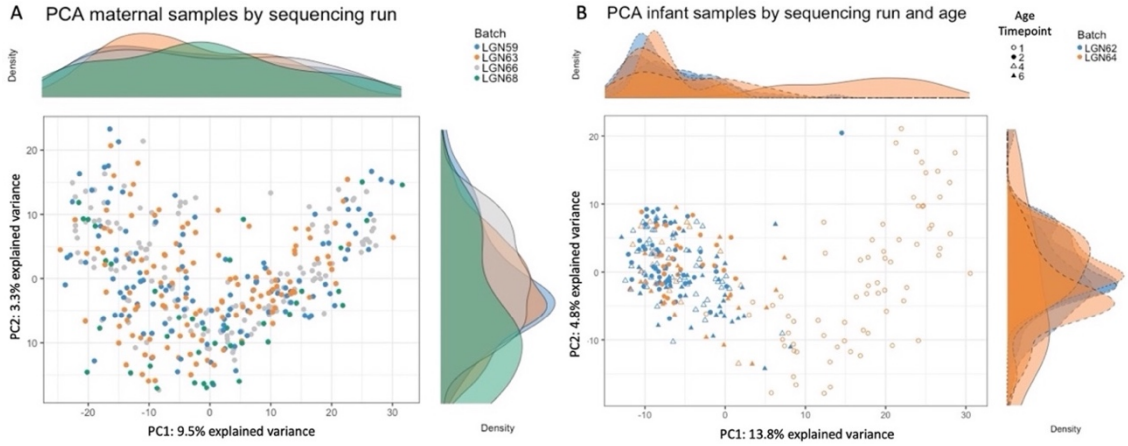


Figure 4.1. *Distribution densities of maternal (A) and infant (B) microbiome samples according to sequencing batch number. Infant samples are also shown according to age, as 1-year old samples were all sequenced on run LGN 64.*

To describe the participant cohort, each of the randomised groups where characterised at baseline, and again after randomisation (Table 4.2). Some participants did not provide information for every variable. Also, mothers and infants who received antibiotics over the course of the study were not excluded from the study. While typically antibiotic use is used as an exclusionary criterion, for this type of longitudinal study, it was impractical to exclude a large proportion of women and infants after enrolment due to antibiotic use. This is because there is a high incidence rate of antibiotic use during later pregnancy, and during the first few years of life. Significantly more mothers in the placebo group had previously used prebiotics prior to the start of the intervention (Chi-squared, $p = 0.03$), and infants in the prebiotic group weight significantly more at 6-months of age (t-test, $p = 0.02$) compared to the placebo group.

Table 4.2. *Baseline and post-randomisation characteristics for mother infant pairs. Averages are presented as mean and standard deviation. Percentages are used to show the proportion of individuals for which the characteristic is true. P values are for ANOVA or Chi-Squared test, and * denotes significant differences.*

Baseline characteristic	**Placebo (n = 33)**	**Prebiotic (n = 41)**	P val
Maternal age in years	32.97 ± 4.05	33.09 ± 3.79	0.89
Ethnicity	1(n = 28) 2(n = 5)	1 (n = 40), 2(n = 1)	0.12
Previous use of probiotic supplements	21 (63.6%)	23 (56.1%)	0.67
Previous use of prebiotic supplements	15 (45%)	8 (19%)	0.03
Pre-pregnancy BMI (kg/m2)	24.8 ± 4.87	24.9 ± 4.79	0.93
Weight at randomization (kg)	71.64 ± 13.57	74.42 ± 13.61	0.38
BMI at randomization (kg/m2)	26.61 ± 4.78	26.28 ± 4.72	0.95
Maternal history of allergic disease	27 (81.2%)	33 (80.5%)	0.88
Infant sex ratio female:male	18:15	20:21	0.8
Post-randomisation characteristic			
Maternal weight at 36 weeks gestation in kg	79.98 ± 13.5	83.20 ± 14.10	0.33
Maternal BMI at 36 weeks gestation	29.72 ± 4.75	29.88 ± 4.77	0.89
Maternal consumption of any probiotics during the intervention period	4 (12.12%)	10 (24.4%)	0.3
Maternal antibiotic use prior to 36 weeks gestation	5 (15.2%)	7 (17.1%)	0.92
Maternal antibiotic use during the intervention period	20 (60.6%)	29 (70.7%)	0.5
Infant birth weight in kg	3.36 ± 0.37	3.44 ± 0.46	0.39
Infant gestational age at birth in weeks	38.63 ± 1.16	38.5 ± 1.24	0.66
Preterm birth < 37 weeks	0	3 (7.3%)	NA
Infant birth vaginal	15 (45.5%)	17 (41%)	0.73
Infant birth caesarean section	18 (54.5%)	24 (58.5%)	0.52
Ever breastfed	33 (100%)	40 (97.6%)	NA
Infant breastfed until 6 months of age	27 (81%)	28 (68%)	0.29
Average breastfeeding duration in months	9.20 ± 3.67	8.03 ± 4.07	0.20
Infant given any infant formula during intervention period	28 (84.8%)	34 (82.9%)	0.92
Age at introduction to infant formula in months in infants given formula	2.14 ± 2.86	2.28 ±3.27	0.87
Infant consumption of any prebiotics (in formula) during the intervention period	15 (45.5%)	18 (43.9%)	0.92
Infant consumption of any probiotics during the intervention period	18 (54.5%)	17 (41.4%)	0.38
Infant antibiotic use during the intervention period *	12 (36.4%)	24 (58.5%)	0.09
Age at introduction to solid foods in months	5.05 ± 0.65	4.91 ± 0.69	0.36
Infant weight at 3 months of age in kg	6.04 ± 0.77	6.38 ± 0.75	0.06
Infant weight at 6 months of age in kg	7.56 ± 0.89	8.07 ± 0.91	0.02

*Dose not include antibiotics given prophylactically during labour

4.5.2 Impacts of the intervention on the maternal microbiome community structure and SCFA concentrations

The intervention had a significant impact on the maternal microbiome composition (PERMANOVA $p < 0.0001$, psudo-F 4.23), and there was no evidence of an interaction between intervention and time (PERMANOVA $p = 1$). The maternal microbiome over the entire intervention period were plotted using PCoA (Figure 4.2, A). The infant microbiome was also found to be significantly impacted by the intervention (PERMANOVA $p < 0.001$; psudo-F 1.7) as well as time (PERMANOVA $p < 0.001$; psudo-F 13.2). Due to the underdeveloped and more similar microbiome of 2- and 4-month old infants, a strong impact of time was also observed on the PCoA of the infant microbiome when viewed across all timepoints up to 12-months of age (Figure 4.2, B). To better observe the tightly clustered infant microbiome at 2 and 4-months of age, these timepoints were also plotted separately (Figure 4.2, D). The prebiotic supplement

was also found to reduce the alpha diversity of the gut microbiome by 8-weeks into the intervention compared to the placebo group, and this difference was maintained until 16-weeks after the start of the intervention (Figure 4.2, C). However, the fold change difference in all alpha diversity measures between the two groups after 8-weeks of intervention was not significant (ANOVA: Fisher, p = 0.075; Shannon, p = 0.14; PD, p = 0.16; Chao1, p = 0.12).

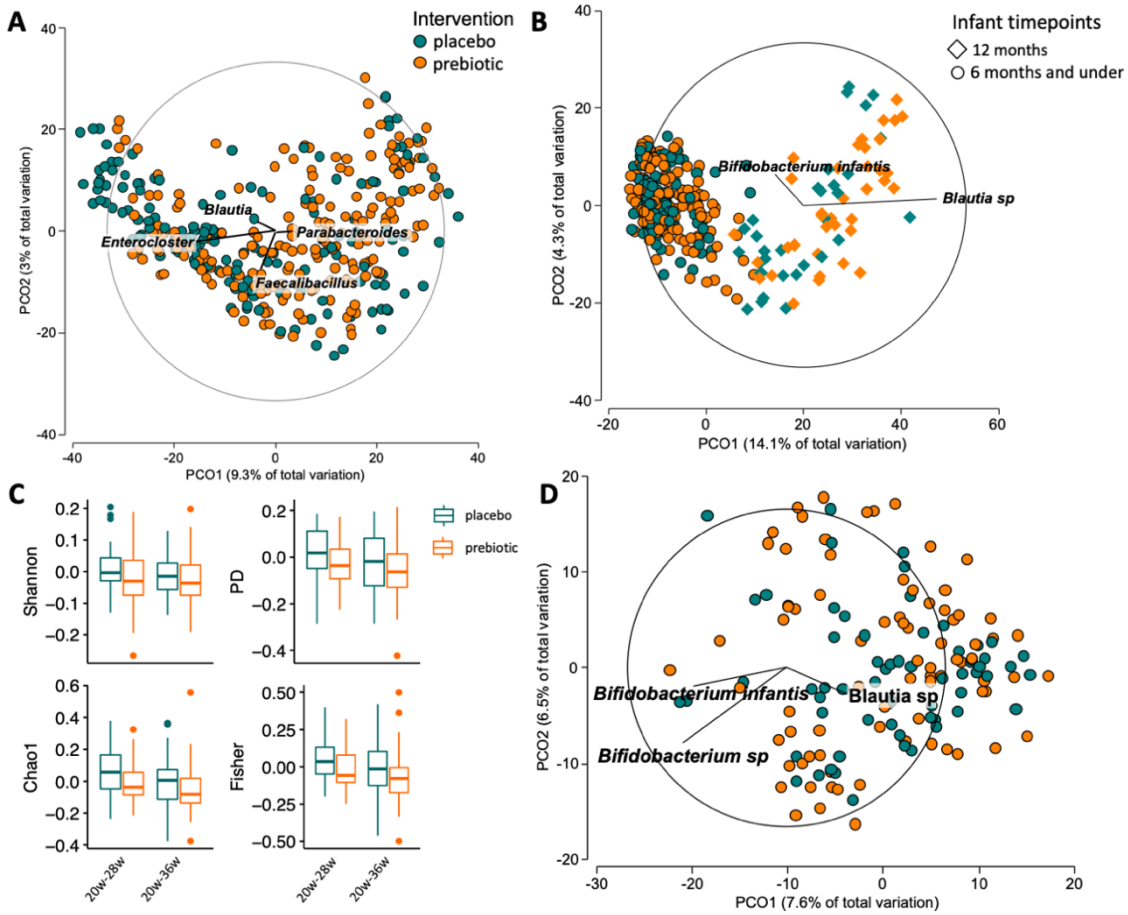


Figure 4.2. *Maternal microbiome composition over the intervention period (A), and infant microbiome composition both over the entire study period (B), and only at the 2 and 4-month timepoints (D). The vectors on the infant plots are ASV 1 and 156 (Bifidobacterium), and ASV 414 (Blautia), and vectors on the maternal plot are ASV 17 (Faecalibacillus), 26 (Parabacteroides), 302 (Enterocloster), and 387 (Blautia). The change in alpha-diversity from baseline to 8 and 16 weeks within both the prebiotic and placebo groups (C).*

Confounding factors that need to be considered when assessing the impact of the intervention over the study period were identified using a distance-based linear model (DistLM). Maternal microbiomes were assessed only at 5 of the 6 timepoints due to high similarity of the microbiome composition at 2- and 4-months. Antibiotic use prior to 36-weeks' gestation, as well as age and ethnicity (which is assumed to impact diet), and age

were determined to explain a significant portion of the total variability. Infant samples were assessed at all timepoints, and birth mode, gender, and breastfeeding habits explained a significant proportion of the total variability (Table 4.3).

Table 4.3. *Factors identified using DistLM that explain a significant amount of variability in the microbiome composition. Each timepoint was assessed independently using the same method, and the proportion of contribution (Prop) for each factor at each timepoint is shown.*

Maternal variables	p-val	Prop.
20-week sequential tests		
Antibiotics prior to 36weeks	0.01	2.3
Age	0.01	2.2
Ethnicity	0.03	1.9
28-week sequential tests		
Antibiotics prior to 36weeks	>0.01	2.5
Age	0.03	1.9
Ethnicity	0.04	1.8
36-week sequential tests		
Antibiotics prior to 36weeks	0.01	2.1
Age	0.04	1.8
2-month sequential tests		
Ethnicity	0.01	2.3
6-month sequential tests		
Ethnicity	0.01	2.2
Age	0.03	2.0
Antibiotics prior to 36weeks	0.02	2.1

Infant variables	p-val	Prop.
2-months sequential tests		
Duration of breastfeeding	>0.01	2.6
Birth mode	>0.01	2.3
Ever breastfed	0.02	2.2
4-month sequential tests		
Duration of breastfeeding	>0.01	2.9
Gender	>0.01	2.7
Birth mode	0.03	2.0
6-month sequential tests		
Duration of breastfeeding	>0.01	3.3
Gender	>0.01	2.1
12-month sequential tests		
EverBreastfed	0.02	2.5

To assess patterns in SCFA concentration during the intervention period, PERMANOVA was used to compare the resemblance of each of the three most predominant SCFA among the intervention groups. Maternal acetic acid concentrations were found to differ by the intervention ($p = 0.008$; Pseudo-F = 7.09) and over time ($p = 0.0025$; Psudo-F 7.1), with pairwise tests showing differences between the groups at the 28-week and 4-month timepoints. Maternal butyric ($p > 0.001$; Psudo-F 6.1) and propionic (pp = 0.004; Psudo-F 3.9) acid differed by time but not by intervention group. Infant SCFA concentrations also differed significantly by time ($p > 0.004$), but not by intervention, and butyric acid showed the larges affect size over time (Psudo-F 92.65).

The SCFA concentrations were also split by randomised group, and plotted over the study period. Concerning the maternal samples within the placebo group, no significant shifts were detected, while in the prebiotic group, butyric acid increased from baseline up until 36-weeks' gestation ($p = 0.04$). There was also a significant drop in acetic acid ($p = 0.002$) and butyric acid ($p = 0.001$) concentrations from the pre- to post-pregnancy

period in the prebiotic group (Figure 4.3, A). The change in maternal SCFA concentration from baseline throughout the pregnancy period was also compared.

In the prebiotic group, 59% of the participants had a positive fold change in acetic acid concentrations from baseline to 28 weeks compared to 31% of the participants in the placebo group. From baseline to 36-weeks, 69% and 46% of the participants in the prebiotic and placebo groups respectively experienced a positive fold change in acetic acid concentration. The concentrations of all three acids increased from baseline in the prebiotic group, while in the placebo group, acetic acid concentrations decreased from the baseline. A significant difference in the baseline shift of acetic acid concentration was observed between the randomised groups during pregnancy (Figure 4.3, B).

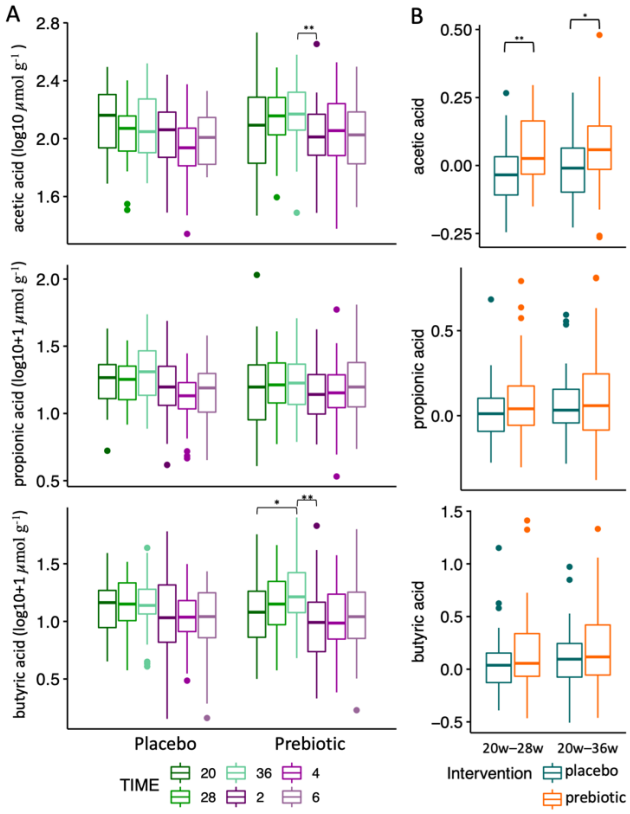


Figure 4.3. *Shifts in individual SCFA concentrations (μmol g^{-1}) in maternal samples over the study period. ** indicate significant differences p>0.01, and * indicate significant differences p ≥ 0.05 determined using a one-way repeated measures ANOVA (A). Fold change in SCFA concentration from baseline to 8 and 16 weeks into the intervention (B).*

The change in SCFA concentration over time in infant samples was also assessed within each randomised group. (Figure 4.4). Butyric and propionic acid both significantly increased from 6-months to 12-months of age in both groups, while in the placebo group, the concentration of butyric acid also increased significantly from 4 to 6-months.

Acetic acid concentrations did not increase significantly during the study, but remained at a relatively high level over the first 6-months of life, averaging 310 μmol g^{-1}. A significant drop in acetic acid from 6 to 12-months was seen in the prebiotic group.

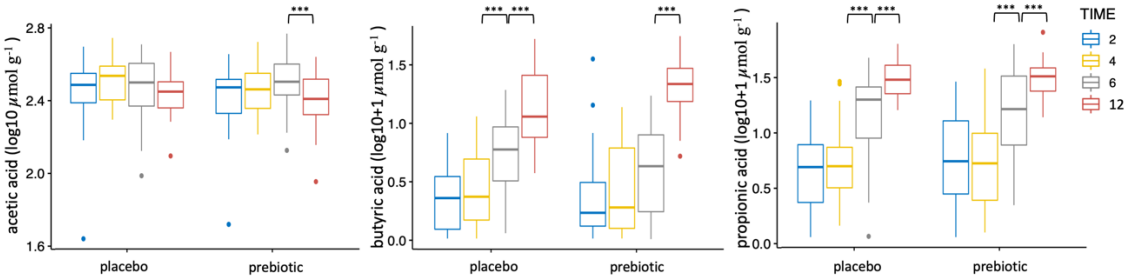


Figure 4.4. *Individual SCFA concentrations (µmol g^{-1}) in infants over the study period, separated by intervention group. Significant differences (padj>0.001) between timepoints are identified as ***. Timepoints compared were 2 – 4-months, 4 – 6-months, and 6 – 12-months.*

Enterotypes were assigned to maternal samples to observe shifts in these important taxonomic groups over time and between randomised groups. Twenty-eight maternal samples (mostly from the post birth period) were not compositionally similar to the MEtaHIT reference samples and therefore confidence in those assignments is lower, however, the gut microbiome of pregnant mothers is expected to look characteristically different to that of a non-pregnant female, which might explain some of the low similarity. Of the 417 assignments, 190 were Firmicutes enriched (F-type), 143 were *Bacteroides* enriched (B-type), and 84 Prevotella enriched (P-type). At the 20-week baseline, 65% and 63% of participants had a Firmicutes dominated Enterotype in the placebo and prebiotic groups respectively. At the 36-week timepoint, the proportion of Firmicute dominated Enterotypes decreased to 52% in the placebo group, and 45% in the prebiotic group. Over the pregnancy period, the shift from a Firmicutes dominated Enterotype to either other (P or B) Enterotypes was significant in the prebiotic group ($p = 0.048$) but not the placebo group ($p = 0.45$).

Random forest (RF) was used to estimate weather samples could be classified into intervention groups based on the composition of genera at the 28-week timepoint and at the 6-month timepoint. At 28-weeks the OOB error was 0.479, with *Eubacterium*, Lachnospirales gen, and *Schaedlerella* enriched in the placebo group and contributing most to the difference between the groups. At 6-months, the OOB error was 0.53, and the genera that most contributed most to the classification were *Blautia* and Lactobacillales sp, both of which were more abundant in the placebo group. Similarly, sparse partial least squares discriminant analysis (sPLSDA) was used to determine the level of accuracy in distinguishing intervention groups. After 100 repeats of validation,

the proportion of samples incorrectly classified to the placebo group 1 was 57%, and 38% to the prebiotic group. Therefore, the microbiome composition could not be used to predict the randomised group using RF.

4.5.3 Differential abundance of taxa between intervention groups

Differentially abundant taxa were identified only within abundant (filtered to 969 ASVs) members using ANCOM-BC. A total of 12 phyla, 76 families, and 219 genera were scrutinized for significant differences between intervention groups at the two timepoints during pregnancy, and first two infant timepoints. At 28-weeks four families and five genera were identified, and at 36-weeks five families and 10 genera were identified as differentially abundant. No taxa with significantly different abundance at the phyla level were identified in the maternal microbiome samples (Table 4.4). The largest affect size was associated with the enrichment of Selenomodaceae in the maternal placebo group at 36-weeks. Similarly, Negativicutes at 4-months were enriched in the infant placebo group with the largest effect size. These taxa were plotted considering maternal antibiotic use and infant birth mode (significant categorical confounders) and a similar trend was observed in both maternal and infant samples. The placebo group had enriched abundance of Negativicutes, regardless of antibiotic use or birth mode (Figure 4.5).

One phylum, two families, and five genera had significantly different abundance when comparing the two randomised groups in the infant microbiome at 2-months. At 4-months there was a further single phylum, three families, and five genera. Genera from Closteridia were identified in each group, and genera within Lachnospiraceae were more abundant in the placebo group, while genera within Peptostreptococcaceae were more abundant in the prebiotic group (Table 4.5). Lastly, to examine the relationship between the class Clostridia and butyric acid concentrations, the abundance of the significantly enriched members of this class at 28 weeks were correlated with butyric acid concentrations in the same sample. Of the six taxa tested, no correlations had R values greater than 0.24.

Table 4.4. *Families and genera identified with differential abundance between the maternal microbiomes in the 2 randomised groups at 28- and 36-weeks. Bolded taxa indicate the taxonomic level tested, and all results shown are significant after BH correction (padj <0.001). The average bias corrected (corrected for differences in sampling fraction) abundance for maternal genera ranged from 0.007 – 8.7.*

28-weeks						**36-weeks**					
Enriched in placebo group				average bias corrected abundance							average bias corrected abundance
Phylum/Class	*Family*	*Genus*	effect size	Placebo	Prebiotic	*Phylum/Class*	*Family*	*Genus*	effect size	Placebo	Prebiotic
Actinobacteriota	**Atopobiaceae**		-0.33	0.13 (0.47)	0.08 (0.58)	Bacilli	Erysipelatoclostridiaceae	**Erysipelatoclostridiaceae gen**	-2.03	0.59 (1.55)	-0.01 (0.38)
Bacilli	**Staphylococcaceae**		-0.67	0.25 (1.09)	0.08 (0.68)	Bacilli	Lactobacillaceae	**Lacticaseibacillus**	-1.70	0.48 (1.12)	0.07 (0.65)
Clostridia	Lachnospiraceae	**Hungatella**	-0.26	0.66 (1.27)	0.56 (0.98)	Bacilli	UBA660	**CAG-1000**	-0.33	0.32 (1.05)	0.23 (1.01)
Clostridia	Lachnospiraceae	**Sellimonas**	-0.96	1.45 (2.10)	0.92 (1.62)	Clostridia	Lachnospiraceae	**Merdimos**	-0.23	0.21 (0.94)	0.16 (0.64)
Clostridia	Lachnospiraceae	**Schaedlerella**	-0.34	2.48 (2.04)	1.47 (1.69)	Negativicutes	**Selenomodaceae**		-1.70	0.63 (1.71)	0.04 (0.58)
Desulfobacterota	Desulfovibrioceae	**Bilophila**	-0.22	2.90 (1.65)	2.66 (1.72)						
Enriched in prebiotic group											
Clostridia	**Eubacteriaceae**		1.35	0.13 (0.53)	0.47 (1.22)	Actinobacteriota	Atopobiaceae	**Lancefieldella**	0.50	0.13 (0.57)	0.21 (0.57)
Clostridia	**UBA9506**		1.27	0.23 (0.74)	0.59 (1.31)	Actinobacteriota	**Atopobiaceae**		0.43	0.13 (0.65)	0.21 (0.64)
Clostridia	Peptoniphilaceae	**Aerococcus**	1.60	0.10 (0.39)	0.14 (0.58)	Bacteroidota	**Barnesiellaceae**		0.62	0.11 (0.76)	0.26 (1.07)
						Bacteroidota	Barnesiellaceae	**Barnesiella**	0.68	0.11 (0.71)	0.27 (1.03)
						Bacteroidota	Muribaculaceae	**UBA7173**	0.51	0.12 (0.49)	0.24 (1.24)
						Bacilli	UBA660	**DUPI01**	2.57	0.27 (0.79)	1.07 (1.58)
						Clostridia	Acutalibacteraceae	**Pseudoruminococcus**	0.40	0.58 (1.79)	0.76 (1.96)
						Clostridia	**UBA9506**		0.95	0.19 (0.55)	0.40 (1.07)
						Clostridia	UBA9506	**UBA9506**	1.02	0.19 (0.53)	0.41 (1.04)
						Proteobacteria	**Enterobacteriaceae**		1.16	0.14 (0.94)	0.52 (1.56)

Table 4.5. *Phyla, families, and genera identified with differential abundance between the randomised groups at the 2- and 4-month timepoints in infants. Bolded taxa indicate the taxonomic level tested, and all results shown are significant after BH correction (padj <0.001). The average bias corrected (corrected for differences in sampling fraction) abundance for infant genera ranged from -0.007 – 8.4.*

2-months						**4-months**					
Enriched in the placebo group				average bias corrected abundance						average bias corrected abundance	
Phylum/Class	*Family*	*Genus*	Effect size	placebo	prebiotic	*Phylum/Class*	*Family*	*Genus*	Effect size	placebo	prebiotic
Clostridia	Lachnospiraceae	**Sellimonas**	-0.87	0.34 (0.93)	0.08 (0.61)	Bacilli	**Lactobacillales_sp**		-0.40	0.23 (0.71)	0.13 (0.88)
Clostridia	Lachnospiraceae	**Schaedlerella**	-0.22	0.33 (1.36)	0.19 (0.65)	Clostridia	**Cellulosilyticaceae**		-0.76	0.25 (0.99)	0.04 (0.75)
Clostridia	Lachnospiraceae	**Hungatella_A**	-0.10	0.29 (0.63)	0.19 (1.08)	Clostridia	Lachnospiraceae	**Mediterraneibacter**	-1.57	0.38 (1.52)	0.08 (0.52)
Desulfobacterota			-0.21	0.35 (1.23)	0.27 (0.99)	Clostridia	Cellulosilyticaceae	**Niameybacter**	-0.41	0.20 (1.14)	0.22 (0.78)
Desulfobacterota	**Desulfovibrioceae**		-0.32	0.35 (0.16)	0.27 (0.85)	**Negativicutes**			-3.31	6.10 (1.96)	4.03 (2.05)
Desulfobacterota	Desulfovibrioceae	**Bilophila**	-0.11	0.35 (1.15)	0.25 (0.85)						
Negativicutes	**Megasphaeraceae**		-0.64	0.19 (0.70)	0.06 (0.47)						
Enriched in the prebiotic group											
Clostridia	Peptoniphilaceae	**Aerococcus**	1.73	0.16 (0.79)	0.55 (1.16)	**Desulfobacterota**			0.14	0.22 (0.96)	0.27 (1.19)
						Desulfobacterota	**Desulfovibrioceae**		0.82	0.22 (0.82)	0.47 (1.28)
						Desulfobacterota	Desulfovibrioceae	**Bilophila**	1.27	0.17 (0.98)	0.65 (1.51)
						Clostridia	Peptostreptococcaceae	**Terrisporobacter**	1.14	0.19 (0.97)	0.75 (1.99)
						Clostridia	Peptostreptococcaceae	**Peptostreptococcaceae_gen**	0.84	0.10 (0.65)	0.50 (1.74)

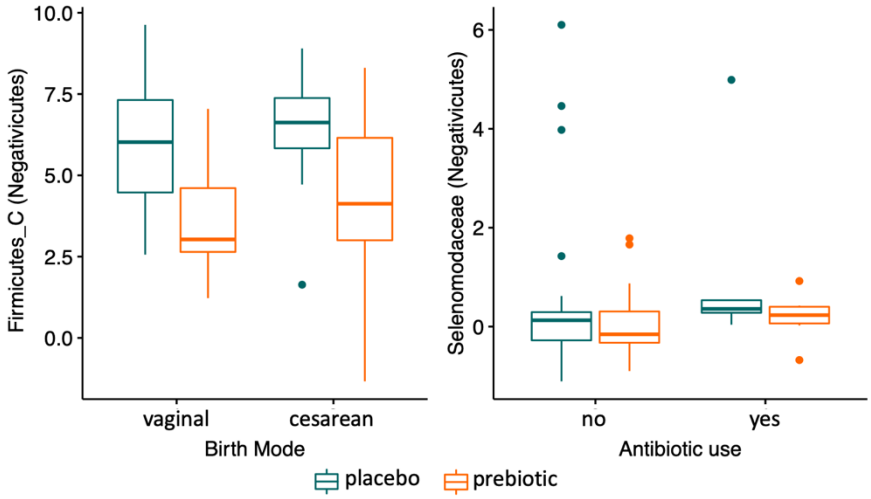


Figure 4.5. *Bias corrected abundance of Selenomodaceae and Negativicutes, which were both significantly different (padj <0.001), and had the largest effect size difference between intervention groups in maternal and infant microbiomes respectively. The intervention groups have been split by significant confounding factors that were identified previously using DistLM.*

Differentially abundant ASVs were also detected between the intervention groups at the 28-week, 36-week, and 6-month timepoints (Table 4.6). The same ASV identified as *Blautia* was enriched in the placebo group at both 28 and 36-weeks. While also controlling for individual, eleven taxa were found to be enriched in the prebiotic group compared to the placebo group across all timepoints during the intervention.

Table 4.6. *Maternal ASVs with differential abundance between placebo (group 1) and prebiotic (group 2) groups at 3 timepoints during the intervention, and across the entire intervention period. Differences were determined using DESeq, and p values have been adjusted (padj) using the Benjamini-Hochburg adjustment.*

ASV	taxa	log2 Fold Change	padj	enriched group
28 weeks				
387	Blautia sp	29.99	<0.001	1
2345	Oscillospirales sp	29.91	<0.001	1
36 weeks				
387	Blautia sp	29.99	<0.001	1
573	Lachnospiraceae	10.62	0.09	1
2862	Oscillospirales sp	17.71	<0.001	2
1779	Ruminococcaceae	30	<0.001	1
6 months				
1795	Burkholderiaceae sp	29.96	<0.001	1
2980	Butyricicoccaceae sp	29.88	<0.001	1
302	Enterocloster sp	30	<0.001	1
2953	Lawsonibacter sp	30	<0.001	1
Over the entire intervention period				
Bacilli				
17	Faecalibacillus	14.56	<0.001	2
75	Turicibacter	14.98	<0.001	2
Bacteroidia				
26	Parabacteroides merdae	11.01	<0.001	2
87	Alistipes finegoldii	13.73	<0.001	2
126	Alistipes obesi	10.75	0.012	2
Clostridia				
31	Romboutsia timonensis	18.54	<0.001	2
146	Lachnospiraceae	18.27	<0.001	2
182	Lachnospiraceae	12.1	<0.001	2
204	Lachnospiraceae	12.24	0.006	2
386	Lachnospiraceae	7.86	0.012	2
815	Oscillospirales	6.56	<0.001	2

The infant microbiome was also scrutinized at each timepoints individually for differences between the intervention groups. Overall, 39 ASVs were found with significantly differential abundance between the two groups. ASVs within Gammaproteobacteria, Negativicutes, and Verrucomicrobiae were only enriched in the placebo group, and primarily were enriched at the 2, and 4-month timepoints. The prebiotic group saw ASVs within Bacteroidia and Clostridia more frequently enriched. *Bacteroides caccae* (ASV 44) was enriched in the prebiotic group at the 2, 4, and 6-month timepoints, and *Prevotella* sp (ASV 47) was enriched in the placebo group at the 6 and 12-month timepoints (Figure 4.6).

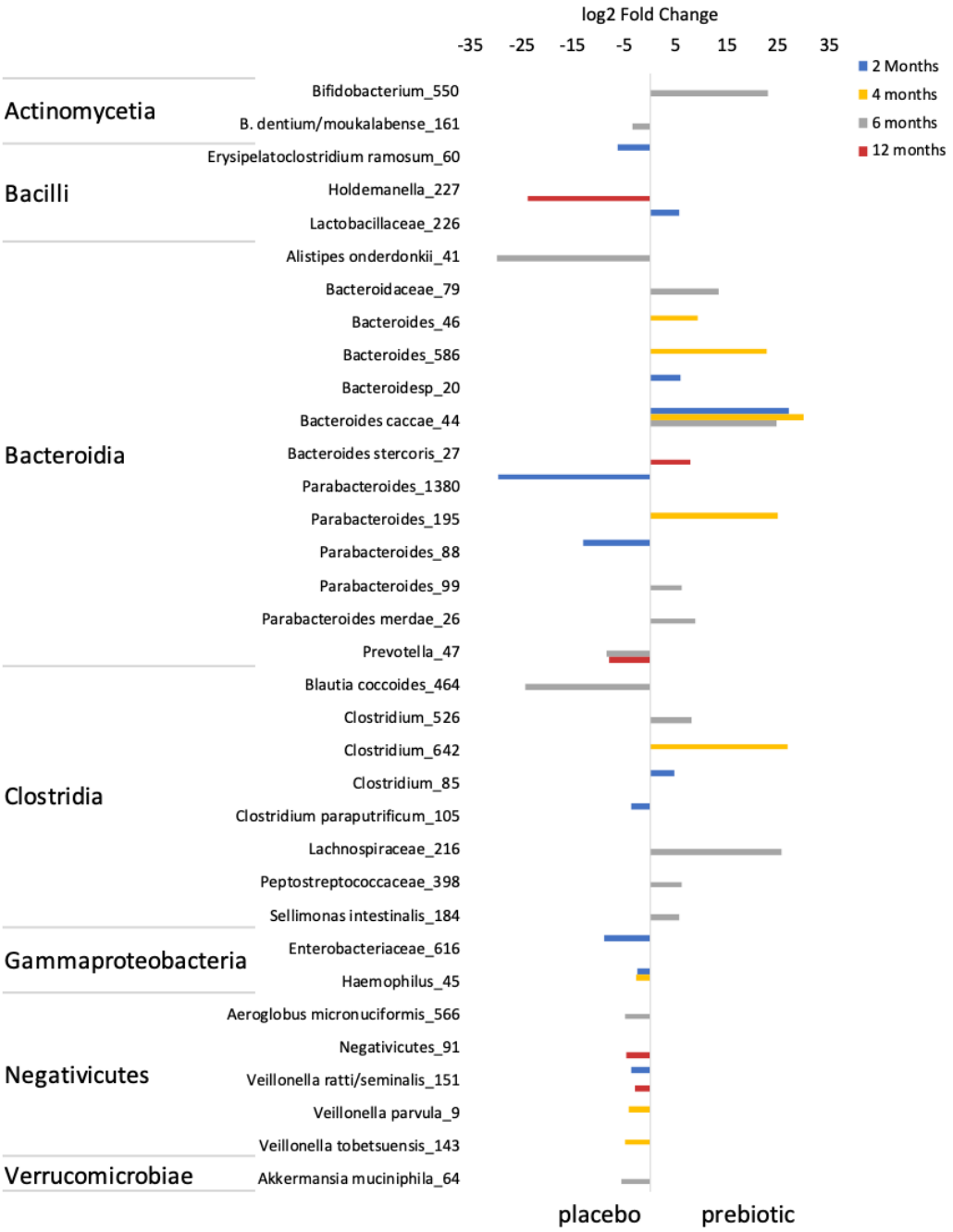


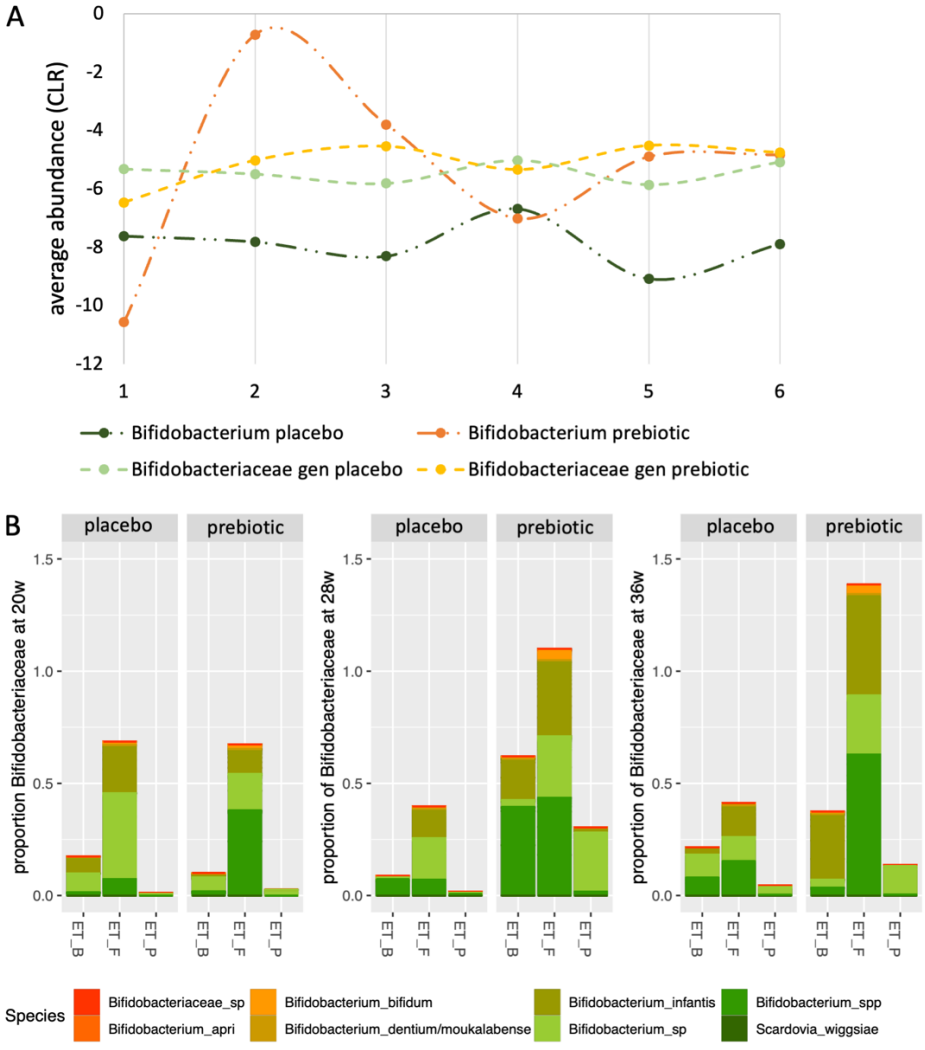
Figure 4.6. *Log2 fold change of significantly enriched bacterial ASVs between infants in the placebo and intervention groups. The lowest taxonomic ID is followed by the ASV number. Each of the age groups are plotted in the same descending order from youngest to oldest for each taxon.*

Lactobacillus, and *Bifidobacterium* are expected to respond with increased abundance to prebiotics. Therefore, a Wilcoxon signed rank test was used to compare the abundance of these genera at baseline (20-weeks) to the abundance at 28- and 36-weeks in each intervention group. *Lactobacillus* did not change in abundance in either group, except for a decrease in the placebo group from 20-weeks to 6-months. However,

Bifidobacteriaceae gen and *Bifidobacterium* both significantly increased from 20- to 28-weeks ($p<0.02$) and from 20- to 36-weeks ($p<0.01$) in the prebiotic group only. To determine if the abundance of these two genera would remain elevated after birth, the maternal baseline abundance was compared to the 2, 4, and 6-month timepoints. In the placebo group, the abundance of *Bifidobacterium* spp. remained relatively consistent over the study period, only increasing non-significantly at the 2-month timepoint. In the prebiotic group both Bifidobacteriaceae gen and *Bifidobacterium* remained significantly higher than baseline at the 4 and 6-month timepoints, but at the abundance of *Bifidobacterium* dropped at the 2-month timepoint only, where it was no longer enriched compared to baseline (Figure 4.7, A). The increase in abundance of Bifidobacterium species after prebiotic supplement may be associated with maternal enterotype. At baseline there was a higher proportional abundance of species within Bifidobacteriaceae in the F dominated enterotype compared to the B and P dominated enterotypes regardless of randomised treatment group. From 20- to 28-weeks, Bifidobacterium infantis and Bifidobacterium spp predominantly increased in proportional abundance in all enterotype groups. At 36 weeks, the relative proportion of Bifidobacterium infantis increased in both the F (11.3%) and B (10.8%) enterotype groups but decreased in the P enterotype group (-2.14%). Also, the relative proportion of Bifidobacterium spp decreased in both the P and B enterotype groups (-1.36%, and -36% respectively), but continued to increase in the F enterotype group (19.2%) from 28 to 36 weeks (Figure 4.7 B).

Infants are expected to have higher diversity of *Bifidobacterium* species in the gut during the first year of life compared to adults, and at 2-months of age, infants hosted six *Bifidobacterium* species with an average abundance greater than 0.3%. Of these six species, *B. infantis* was most abundant (average 27%), and *B. dentium*/*moukalabense* was least abundant (average 0.3%). A comparison of these species between all infants in each randomised group at 2-months of age showed no significant difference in abundance ($p > 0.17$).

Figure 4.7. *The response of members of Bifidobacteriaceae to a maternal prebiotic intervention. (A) Differences in the CLR transformed abundance of Bifidobacterium over the six timepoints (20, 28, 36-weeks, and 2, 4, and 6-months) of the study period. Negative abundance does not indicate a deficit in abundance, but rather an abundance less than the mean centre abundance. Colours compare between prebiotic (orange) and placebo (green) groups. Bifidobactericeae gen is a genus which were unresolved past the family level. (B) Proportional abundance of species within Bifidobacteriaceae according to enterotype and compared between prebiotic and placebo groups. Timepoints shown are baseline (20-weeks) and post intervention (28-weeks).*



4.5.4 Temporal overview of maternal and infant microbiome and SCFA concentrations

Changes to relative community abundance were observed over the pregnancy period with Bacteroidales expanding by 6% on average, and Oscillospirales reducing an average of 2.5%, from 28 to 36 weeks. The abundance of Bacteroidales dropped after birth to an average of 26%, and then expanded again by 4% at 6 months (Figure 4.8). Changes in abundance were seen when comparing directly between pre- (36w) and post- (2m) birth. After birth Clostridiales expanded by 1.5%, and Lachnospirales reduced by 2.6% on average. The infant microbiome also shifted over the study period and most noticeably after the introduction of solid food, which was on average at 5-months of age (Figure 4.9). From 4 to 6 months, Bacteroidales and Lachnospirales both expanded by 4% and 7% respectively while Clostridiales and Enterobacterales both contracted by 6.5% and 5% respectively. From 6 months to 12 months, Lachnospirales increase by 18.5% while Actinomycetales decrease by 20%.

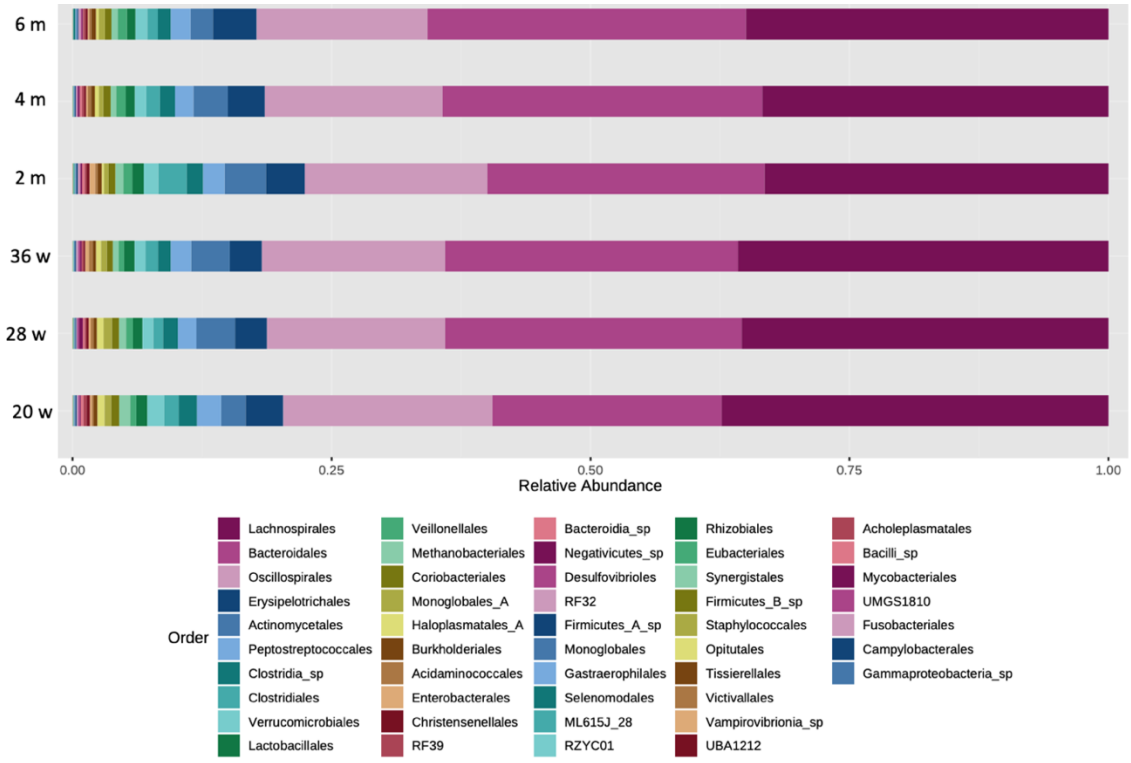


Figure 4.8. *Average relative abundance of the microbiome from all 71 mothers over the six timepoints of the study period. The legend lists the bacteria in descending order of abundance.*

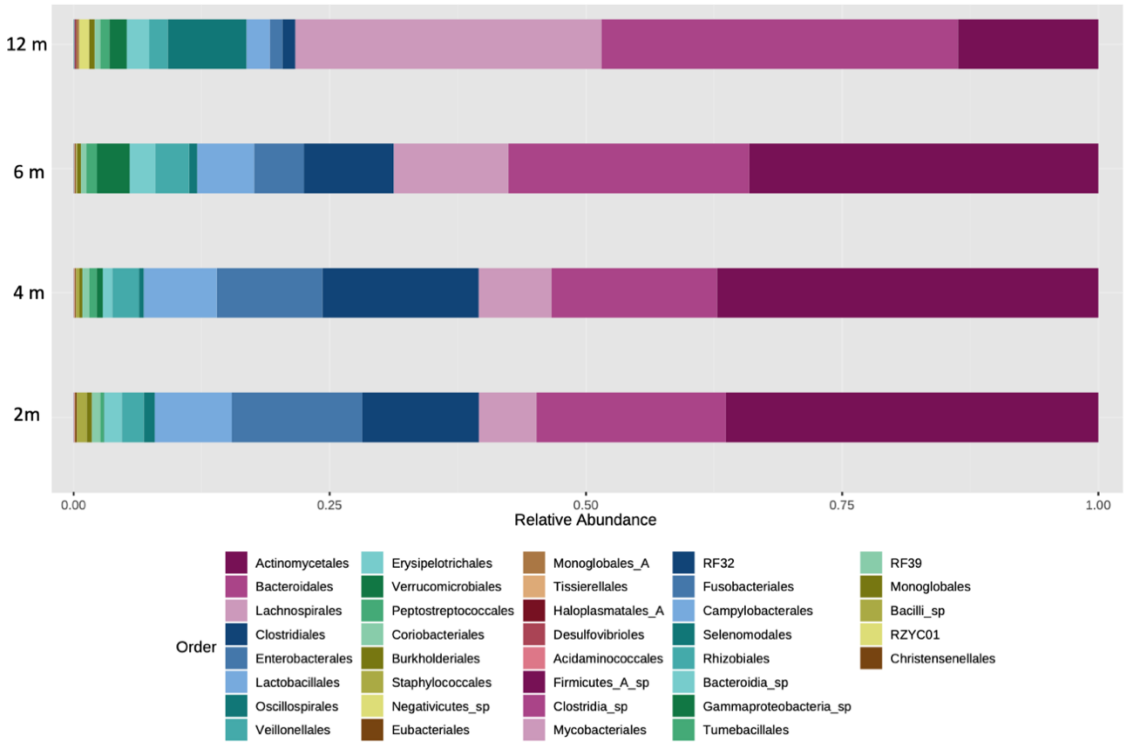


Figure 4.9. *Average relative abundance of the microbiome from all 75 infants over the four timepoints of the study period. The legend lists the bacteria in descending order of abundance.*

The average rarefied alpha diversity dropped consistently at each timepoint during the pregnancy period, with a significant drop in PD and Fisher diversity from 20- to the 36-weeks (padj<0.036), PD also dropped significantly from 2- to 6-months (padj=0.006). Estimates of richness and diversity for infants also increase significantly over the study period from 4 to 6 months (padj>0.001), and from 6 months to 1 year (padj>0.001) (Figure 4.10).

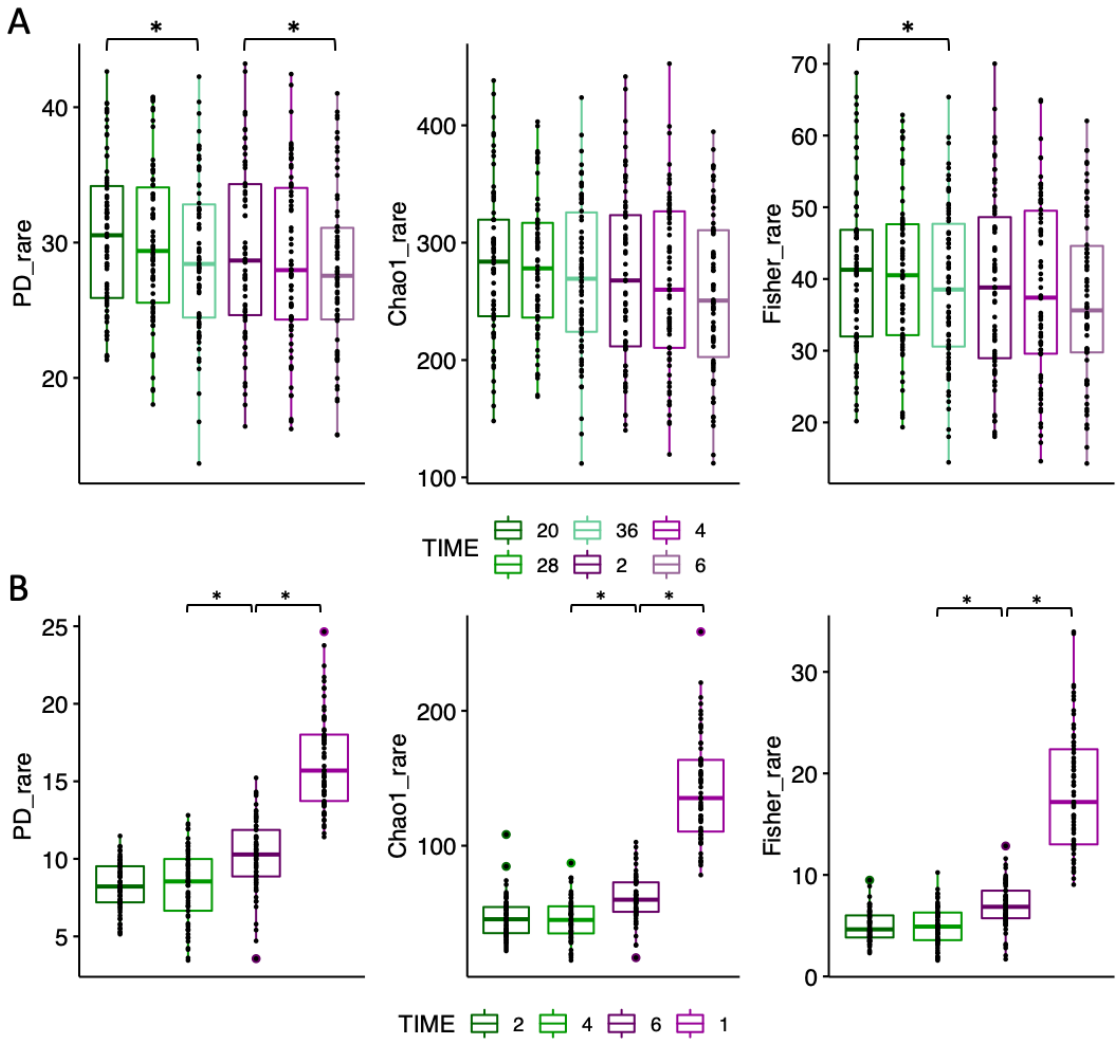


Figure 4.10. *Richness and diversity estimate for both maternal (A) and infant (B) samples over the study period. Significant differences (p > 0.05) between timepoints are indicated by a * over brackets.*

During the pregnancy period, the average acetate, propionate, and butyrate concentrations were highest at the 36-week timepoint. Both acetate (padj=0.02) and butyrate (padj=0.002) dropped significantly from the 36-week timepoint to the 2-month timepoint, and remained significantly lower at the 4-month (padj<0.01) and 6-month (acetate, padj=0.02; butyrate, padj=0.01) timepoints. Propionate was also significantly lower at 4-months after birth compared to 36-weeks' gestation (Figure 4.11).

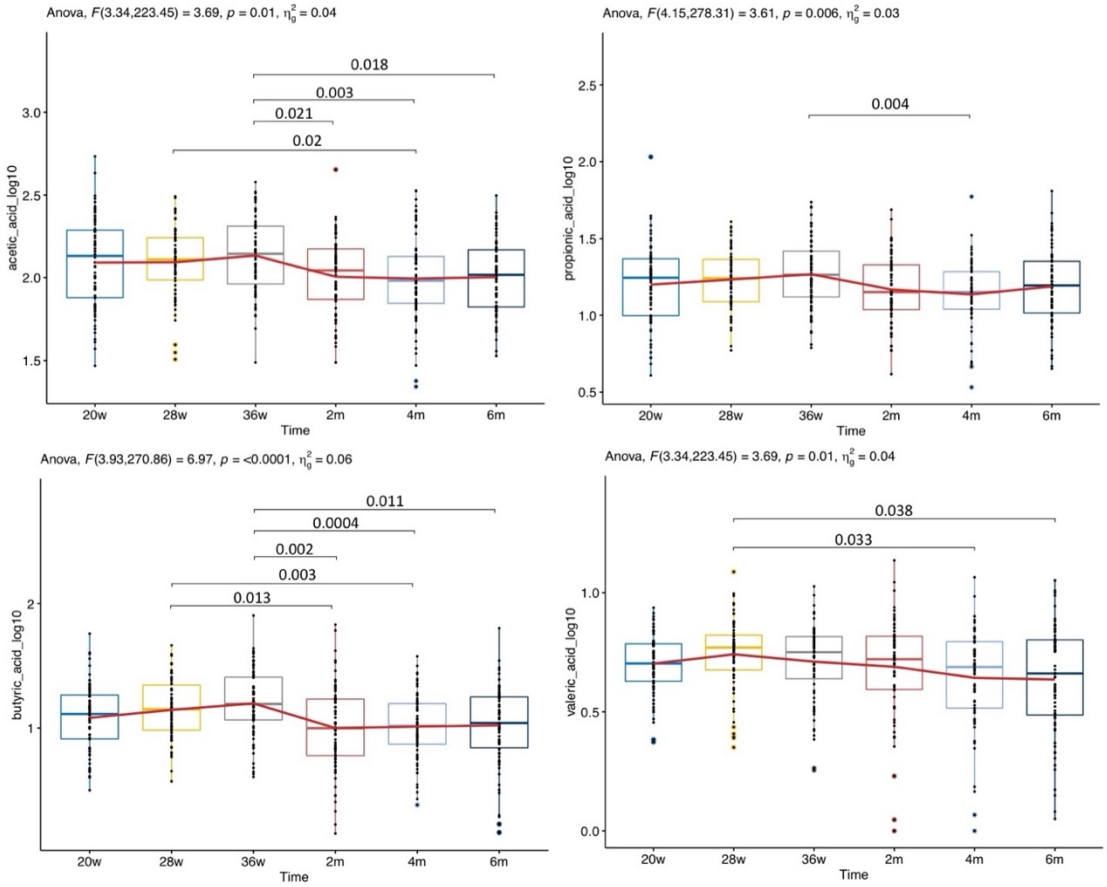


Figure 4.11. *One-way repeated measures ANOVA showing significant differences in SCFA concentration in maternal samples over time.*

From 2- to 4-months of age, butyric and propionic acid remained relatively similar, ranging from an average of 2.5±5.0 to 2.7±3.3 µmol g^{-1} and 7.1±7.2 to 6.8±8.2 µmol g^{-1} respectively. From 4- to 6-months of age, and 6 to 12-months of age both butyric acid and propionic acid significantly increase to 5.22±4.9 and then 20.47±13.2 µmol g^{-1}, and 19.1±14.5 and then 32.6±13.5 µmol g-1 respectively. Acetic acid remained relatively consistent over the first 6-months of life increasing from 298.1±84.4 µmol g-1 at 2-months to 320.1±99.2 µmol g^{-1} at 6-months, but then dropped from 6- to 12-months to an average of 269.9±77.1 µmol g^{-1}.

4.5.5 Other factors associated with microbiome composition

The maternal microbiome displayed significant differences in microbial beta diversity according to stool form over all timepoints (PERMANOVA p = 0.001; Pseudo-F = 2.17). A pairwise test comparing firm (1-2), normal (3-4), and loose (5-7) stool groups showed a significant difference accross all three groups (p > 0.01). The association between maternal stool form, diversity, and SCFA concentration was determined by comparing the average Fisher, Chao1, phylogenetic diversity, and total SCFA

concentrations within Bristol stool form scale (BSFS) groups (Figure 4.13). All diversity measures dropped significantly in the firm stool group compared to the loose stool group (rm-ANOVA $p<0.03$), while acetate, butyrate, and propionate all significantly increased (rm-ANOVA $p<0.001$). The average Bristol Stool form at baseline was nearly identical between the two groups (placebo 4.0±1.0, prebiotic 3.9±1.3), and remained similar during pregnancy and just after birth. At the 4-month timepoint the Bristol stool form was significantly lower in the placebo group (3.6±1.0) compared to the prebiotic group (4.3±1.1) (ANOVA padj = 0.02). While not significant, the average Bristol stool form remained lower in the placebo group at 6 months. Infant samples were grouped according timepoint and the proportion of stool samples in each stool form group was determined. No infant stool samples were firmer than a 3 on the BSFS, and at 2 months 75% of all stools were type 6. At 12 months, 79% of stools were of type 4 or 5 (Figure 4.14).

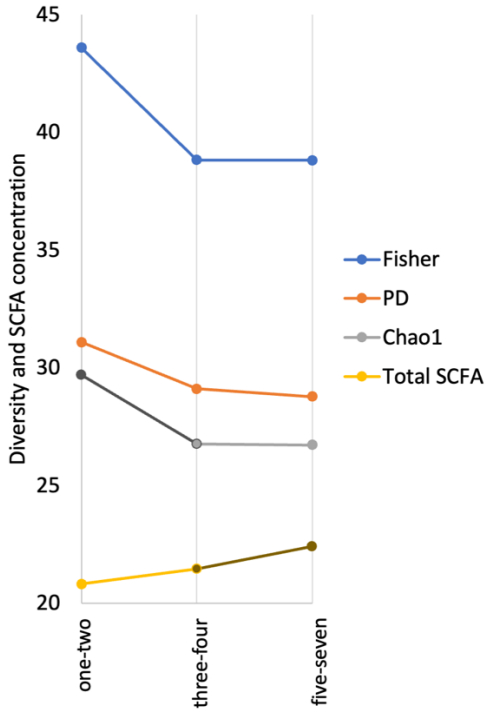


Figure 4.12. *Maternal microbiome diversity and total SCFA concentration according to Bristol stool form group. To fit the 4 measures visually on the same Y axis the phylogenetic diversity (PD) has been divided by 10, and Chao1 diversity and total (log 10) SCFA concentration have been multiplied by 10.*

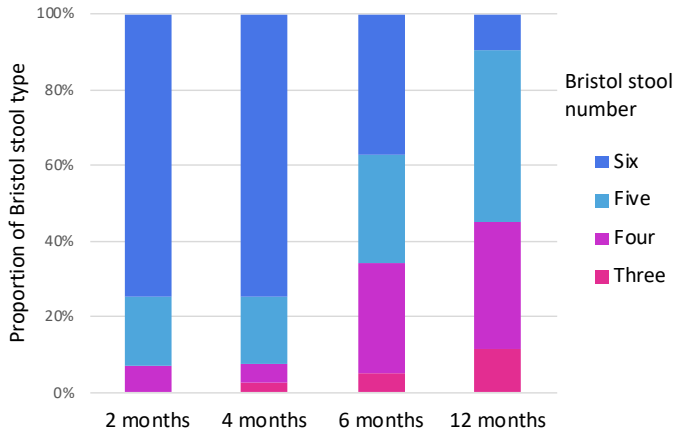


Figure 4.13. *Changes to Bristol stool form over the first year of life.*

Patterns search was also used to find groups of bacterial genera in maternal stool with abundance that corelated with stool firmness at baseline (20-week timepoint). Acutalibacteraceae gen, *Ruthenibacterium*, *Akkermansia*, and *Oscillibacter* were all significantly corelated with increased stool firmness ($p < 0.042$), while not significant, *Haemophilus*, Butyricicoccaceae gen, Christensenellales gen, and *Faecalibacillus* were all positively correlated with decreasing stool firmness (Figure 4.15).

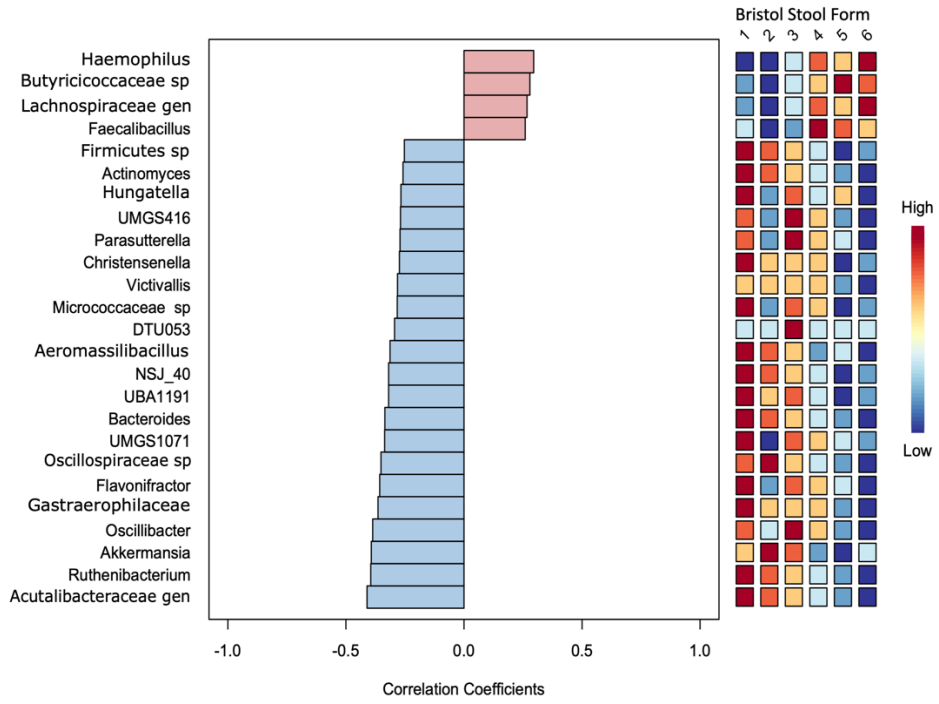


Figure 4.14. *Bacterial genera from 71 maternal microbiomes at baseline (20-weeks pregnancy), that corelate with Bristol stool form. The heatmap shows bacterial abundance, according to stool form which ranges from firm (1) to lose (6).*

A distance-based redundancy analysis was used to illustrate the variability explained by the confounding factors (identified using DistLM) at various timepoints for maternal and infant samples. Age and ethnicity appeared to have a similar direction of impact on maternal samples when visualised in 2D space during pregnancy, but not after birth. Through the study period, the consumption of pre- and pro- biotics appeared to have an opposing impact on the maternal sample distribution (Figure 4.16). Birth mode and gender were found to have a similar direction of influence on the infant microbiome, and distinct clusters by gender were observed at 2- and 4-months of age (Figure 4.17). The number of caesarean births differed significantly by gender (Chi-Squared statistic 5.63, $p = 0.02$), with boys more commonly being born caesarean than girls (25 boys and 15 females.). At 2-months of age, the abundance of Bacteroidota was significantly higher in vaginally born infants compared to caesarean born infants (FDR > 0.001), and at the genus level *Phocaeicola* and *Bacteroides* were significantly more abundant in the vaginally born infants (FDR > 0.04). At 4-months of age, the abundance of Bacteroidota was still significantly higher in vaginally born infants (FDR > 0.001) and additionally, the abundance of Firmicutes (Bacilli) was significantly higher in caesarean born infants (FDR = 0.03). At 4-months, there were no statistically significant differences in the abundance of bacterial genera, however at the family level Enterococcaceae were significantly higher in caesarean born infants (FDR = 0.048). By 6-months of age the grouping by gender and birth mode was no longer observed, and by 12-months prebiotic formula and antibiotic use were contributing more strongly, but in opposite directions on infant sample distribution (Figure 15). Duration of breastfeeding also explained a significant proportion of the infant microbiome variability at 2, 4, and 6-month timepoints. Overall, the total variation explained by factors included in the DistLM was

much lower than the fitted variation represented in the dbRDA plots, indicating that a large proportion of the variation in the microbiome is not explained by these factors.

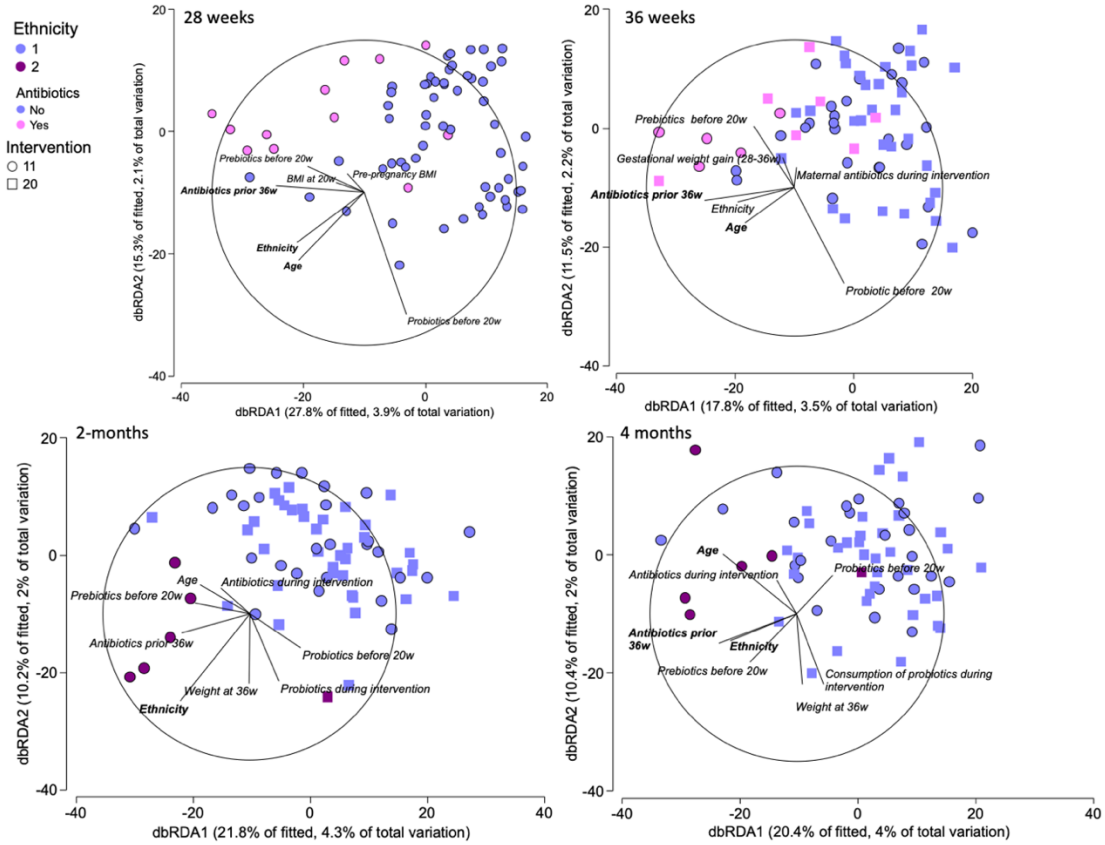


Figure 4.15. *Maternal samples plotted within a dbRDA at 4 timepoints. Vectors overlaying each plot are factors that contributed more than 1.5% to the overall variation explained (determined using DistLM), and bolded vectors contributed significantly to the variation explained.*

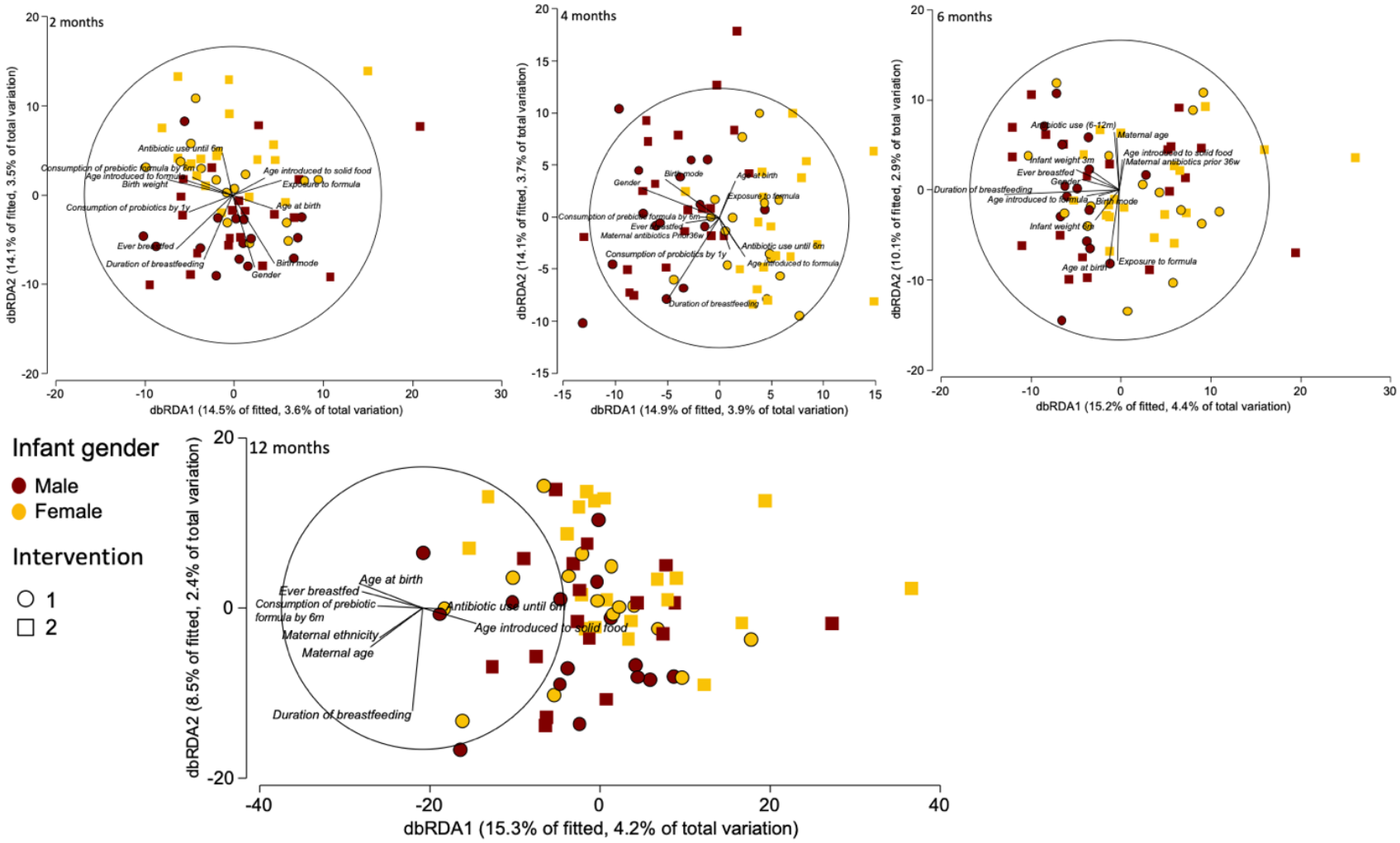


Figure 4.16. *Infant samples plotted using a dbRDA at each of the 4 timepoints. Vectors overlaying each plot are factors that contributed more than 1.5% to the overall variation explained (determined using DistLM), and bolded vectors contributed significantly to the variation explained.*

To explore the influence of breastmilk on the developing microbiome, at 4-months of age the infants were grouped into breastfeeding duration categories (0-2m n=8, and 2-4m n=63), and the average proportion of genera were plotted in each group. In infants breastfed for at least 2 months, *Bifidobacterium* and *Bacteroides* made up a larger proportion of the microbiome than in infants breastfed less than 2-months (Sup Figure 4.3). However, only 8 infants were breastfed less than 2-months so the lack of *Bifidobacterium* spp may also be due to under sampling.

4.6 Discussion

The breakdown of dietary fibre in the large intestine is one of the most important drivers of gut microbiota composition, and prebiotic dietary supplements have been shown to have significant positive effects on the microbial community of both infants and adults. In this randomised controlled trial, the structure and function of both the maternal and infant gut microbiome were assessed for the effects of a maternal prebiotic supplement. From 20 weeks of pregnancy, until 6 months after birth, mothers consumed either a prebiotic fibre mixture of fructo-oligosaccharides and galacto-oligosaccharides, or a placebo (maltodextrin). Despite the strong inter-individual variations observed between women over the study period, the prebiotic supplement made a significant impact to both the gut microbiome composition and SCFA concentrations. The maternal prebiotic supplementation also made a significant impact on the infant microbiome which was observed over the first year of life.

4.6.1 Microbiome response to prebiotic supplementation during pregnancy and during lactation

Supporting the gut microbial community via prebiotic supplementation is an emerging strategy that supports the growth of commensal microbes and assists the microbiome resist dysbiosis. In this randomised controlled trial, the prebiotic supplement significantly impacted the maternal microbial beta diversity over the intervention period. The same analysis showed beta diversity was not impacted by an interaction between the intervention and time, indicating no long-term cumulative effect of the prebiotic. Rather the prebiotic appeared to stimulate a shift in the microbial community that may persist as long as the environmental conditions remain. Similarly, in overweight adults a sustained shift to the bacterial population (cells/g of faeces) after a prebiotic supplement containing GOS was observed after 6- and 12-weeks when compared to a placebo group (Vulevic *et al.*, 2013). Although, most of the trials assessing prebiotic interventions in adults only last for 1 - 4 weeks (Sawicki *et al.*, 2017), so it may not be possible to support this hypothesis with other long-term studies.

One of the most predictable responses to a prebiotic supplement with FOS and GOS is an increase in the abundance of Bifidobacteria. A recent meta-analysis summarised 20 randomized controlled trials, 15 of which were examining FOS and GOS prebiotic types, and all were able to demonstrate an increase in this genus (Sawicki *et al.*, 2017). Bifidobacteria in particular are able to utilize FOS via beta-fructofuranosidase, and are suspected to prefer shorter chain oligosaccharides based on growth patterns with various oligosaccharides (Ose *et al.*, 2018). Bifidobacteria were found to be enriched in faecal samples at both 28 and 36-weeks' gestation, demonstrating that the prebiotic could elevate the abundance of this important genus during the second half of pregnancy. It was also found that during the pregnancy period, the prebiotic supplement increased the stool concentrations of acetate. Unlike other abundant gut residents such as Bacteroides and Clostridia, Bifidobacterium ferment primarily carbohydrates, and produce lactate and acetate (Ose *et al.*, 2018).Therefore, the increased acetate in the prebiotic group during pregnancy may be due in part to the increase in *Bifidobacterium*. An increase in both *Bifidobacterium* and acetate have been found in the faeces of pregnant mice receiving a GOS/inulin prebiotic supplement compared to mice fed a control diet during pregnancy (Brosseau *et al.*, 2021); however, in an earlier study the abundance of *Bifidobacterium* increased in healthy adults after supplementation with either FOS or GOS, but acetate concentrations did not increase (Liu et al. 2017).

After birth in the prebiotic group, the abundance of Bifidobacteria remained significantly higher than baseline except at 2-months postpartum, where the abundance was reduced to a level that was comparable to the abundance in the placebo group. Similarly, acetate concentrations were at their lowest point during the study period at 2-months in the prebiotic group. In another study, the microbiome of 47 healthy women was compared across pregnancy and the postpartum period: three of their timepoints (23-28w and 33-38w pregnancy, and 6w postpartum) match closely with 3 of our timepoints. They observed an increase in the relative abundance of Actinobacteria and *Bifidobacterium* beginning at 33-38w pregnancy that peaked at 6w postpartum (Qin *et al.*, 2021). In line with this study, a similar, but not significant pattern of increased *Bifidobacterium* was observed in the placebo group at 36-weeks; thus, an increase in this genus in the immediate post-partum period may be a characteristic fluctuation in the microbiome that occurs after pregnancy, or during the breastfeeding period. The early postpartum period is often a particularly difficult time in a mother's life. Lactation (by increasing energy demands), sleep disturbance and depravation (by impacting endocrine and nervous systems), depression and infection can all impact the composition of the microbiome (Mutic *et al.*, 2017), which may have contributed to the observed shifts in Bifidobacteria and acetate concentrations in this trial. However, it is also important to note that although the abundance of *Bifidobacterium* dropped in the prebiotic group at the 2-months timepoint, the abundance here was similar (mean difference of |0.32| in CLR) to that of the placebo group. This observation indicates that the prebiotic, and

potentially fibre in general, was most likely not driving the abundance of *Bifidobacterium* at 2-months postpartum.

Over the entire intervention period, several other bacterial taxa were identified with significantly higher abundance in the maternal prebiotic group compared to the placebo group. These enriched ASVs included several Bacilli, Bacteroidia, and Clostridia. One feature enriched in the prebiotic group, which is known to respond to supplementation with FOS was *Parabacteroides merdae* (ASV 26). This taxa has previously been shown to grow equally well with any oligosaccharide supplement, producing both acetate and propionate (Ose *et al.*, 2018), indicating this bacterium may have been able to utilise efficiently both FOS and GOS in this intervention. At 28 and 36-weeks' gestation specifically, bacterial families and genera including Atopobiaceae and *Lancefieldella*, Barnesiellaceae and *Barnesiella*, Enterobacteriaceae, and Eubacteriaceae were all enriched in the prebiotic group. While previously none of these taxa have been described explicitly responding to prebiotic consumption, some strains of Enterobacteriaceae have been shown to grow well on FOS or GOS growth mediums (Hoeflinger *et al.*, 2015). Also, the abundance of *Barnesiella* (Bacteroidota) has been found to be higher in lean male athletes (Bielik *et al.*, 2020), and lower in obese adults (Chiu *et al.*, 2014) which may indicate this genus is enriched in healthy weight individuals. It is expected that gut microbes' will engage in complex competitive and commensal cross-feeding habits (Seth and Taga, 2014). Therefore, the enrichment of Bifidobacteria, which produce mainly lactic acid and acetate (Liu *et al.*, 2017), may also increase the abundance of other taxa. For example, the highly prevalent Enterobacteriaceae are the most taxonomically diverse family of bacteria within the gut (Martinson *et al.*, 2019), and thus we might speculate that some strains may benefit from cross feeding, as this family displays no growth in absence of oligosaccharides (Hoeflinger *et al.*, 2015).

While the metabolic activity of some microbes can be anticipated, the eventual end-products that result from the entire microbial community are more difficult to predict. Furthermore, the microbiome does not function in isolation from the host, and the health of an individual also plays a role in the resulting microbial end-products (Ose *et al.*, 2018). The additional demands on nutrient acquisition, as well as immune system remodelling that happen during the pregnancy period are expected to occur in association with the gut microbiome. For example, the diversity of the gut microbiome has been shown to drop during pregnancy, and butyrate producing bacteria in particular are lost while lactic acid-producing members expand in abundance (Mesa *et al.*, 2020). In this study, the three predominant SCFA's (acetate, propionate, and butyrate) were higher in both groups during the antepartum period compared to the postpartum period, which may be related to the fundamental metabolic changes necessary to support the developing foetus (Ziętek, Celewicz and Szczuko, 2021). However, both acetate and

alpha diversity showed directional shifts over the pregnancy period, that were different according to the randomised groups. In the prebiotic group, acetate concentrations increased towards later pregnancy, while alpha diversity decreased, which was opposite to what was observed in the placebo group. Acetate has the lowest pKa of all SCFA, and is both the strongest (Zhou, Cao and Zhou, 2013), and most prevalent acid in the gut. It therefore contributes substantially to lowering the gut pH, which would pose a selective pressure on the microbial community that could result in a decrease in alpha diversity (Simpson and Campbell, 2015).

Butyrate concentrations also increased significantly during pregnancy in the prebiotic group, but remained at a relatively similar concentration in the placebo group. Butyrate displays anti-inflammatory effects in the gut, and is mainly produced by species within Clostridia, from acetyl-CoA which results from the breakdown of carbohydrates or lactate (Yadav *et al.*, 2018). Differential abundance testing identified 3 different genera from the family Lachnospiraceae that were enriched at 28-weeks in the prebiotic group. At this taxonomic level, it may be difficult to speculate about the relationship between these taxa and the butyric acid concentrations, especially as none of these taxa corelated with butyric acid concentrations. In the gut, ~95% of the butyrate that is produced is transported across the epithelium (Pryde *et al.*, 2002), and is utilised as an energy source preferentially by colonocytes (Roediger, 1980). Due to the rapid uptake of butyrate, two potential explanations for increased concentrations could be 1) decreased transit time which can impair the efficiency of butyrate uptake, or 2) butyrate production has increased. FOS has been shown to increase the production of butyrate through cross-feeding of lactate produced by *Bifidobacterium* (Macfarlane, Steed and Macfarlane, 2008), consequently, the prebiotic in this trial may also contribute to butyrate production, which would support gut health during pregnancy.

The concentration of propionate did not appear to be affected by the prebiotic supplement, and in both groups, was higher during pregnancy compared to after birth. Propionate resulting from carbohydrates is produced primarily via the succinate pathway, which is mainly found in Bacteroidetes (Louis and Flint, 2017), and which did not differ in abundance between the groups. While GOS and inulin supplementation has been shown to increase propionate concentrations in pregnant rats (Brosseau *et al.*, 2021), in our study supplementation with GOS and FOS did not have this affect. Propionate is involved in host metabolic and immune processes that are particularly relevant during the pregnancy period (Louis and Flint, 2017; Ladyman and Brooks, 2021; Ziętek, Celewicz and Szczuko, 2021). Therefore, propionate may be closely regulated during this time.

The microbiome composition is also expected to respond to the changes associated with the progression of pregnancy. Alpha-diversity (discussed previously) and enterotypes

were used to describe broad changes that occurred over the pregnancy period, and that differ according to the randomised groups. Enterotypes are identified by the variation in abundance between Bacteroides (B), Prevotella (P), and Ruminococcus/Firmicutes (F) (Arumugam *et al.*, 2011). Prevotella are considered mostly anti-inflammatory, while Bacteroides are considered more inflammatory, and the role of Ruminococcus is less clear (Klimenko *et al.*, 2018). Enterotypes are a useful way to stratify samples based on community composition and may be particularly relevant in describing response to dietary intervention (Costea *at al*., 2017), however as with any discrete grouping method, applied to continuous data, some community structures will be on the cusp of one or more groups, and therefore less effectively clustered (Knights *et al*., 2014). During pregnancy, enterotypes have been shown to remain relatively stable up until 37-weeks, when the proportion of F dominated enterotypes drops, and both P and B dominated enterotypes expand (Yang *et al.*, 2020). In our study, the proportion of women with F dominated enterotype was similar in both the prebiotic (63%) and placebo (65%) groups at baseline. With the progression of pregnancy, the proportion of women with F enterotypes fell as expected, however at 36-weeks, only the prebiotic group experienced a significant shift to a P or B enterotype. This shift may mean that the prebiotic potentially promoted this shift in enterotype that is known to occur during the later stages of pregnancy. While it would have been interesting to determine if the shift from an F enterotype was more likely to result in a P or B enterotype in the prebiotic group, the subdivision of mothers by intervention, timepoint, and then enterotype resulted in groups too small to assess further statistically (P-type, 4 placebo, 8 prebiotic; and B-type, 11 placebo, 14 prebiotic).

Lastly, the response to the prebiotic, in terms of an increase in Bifidobacteriaceae, appeared to differ according to enterotype. Within the prebiotic group, Bifidobacterium spp increased predominantly in the F dominated enterotype group, and Bifidobacterium infantis increased in the F and B dominated enterotype groups from baseline to 28 and 36 weeks. An individualised response to prebiotic fibre (including GOS) according to the baseline microbiome composition has been demonstrated previously in adults (Holmes et al., 2022). Dietary changes have also been observed to produce inconsistent responses (Bedu-Ferrari, 2022), and these observations are thought to occur due to differences in bacterial membership, and enzyme capacity. In line with this evidence, one potential explanation for the favourable response to the prebiotic supplement in the F dominated enterotype group is that the microbial environment of this group is less resistant to change. Also, there was a larger abundance of Bifidobacteriaceae at baseline within the F dominated enterotype which together may have allowed for the proliferation of Bifidobacteria species. It is also important to recall that a significantly larger number of mothers shifted to a P of B dominated enterotype in the prebiotic group at 36 weeks, which demonstrates the complex interplay between pregnancy, and baseline microbiome composition towards the response to the prebiotic supplement.

4.6.2 Detectable shifts to the infant microbiome over time due to the maternal prebiotic supplement

Unlike the maternal microbiome, the infant microbiome displays very little inter-individual differences prior to the introduction of solid foods and the cessation of breastfeeding. It is also much more susceptible to perturbation during the early stages of development, and impacts to the assembly of the microbiome are predicted to have more long-lasting affects compared to impacts that occur after the community is established. The maternal prebiotic supplement was given to pregnant mothers to support their microbiome during pregnancy, with the anticipation that the benefits would also be received by the infant. In this study, over the first year of life the infant microbiome increased in richness, and alpha-diversity, and beta diversity was significantly impacted by the maternal intervention. Some bacterial features were also found to differ significantly between the randomised groups. At the ASV-level, *Bacteroides* spp (ASVs 20, 46, and 586) were enriched at 2 or 4-months, and *Bacteroides caccae* (ASV 44), was enriched consistently at 2, 4, and 6-months. *B. caccae* is more abundant in infants born vaginally compared to caesarean (Stewart *et al.*, 2018), and has been shown to have similar growth patterns to *Bifidobacteria* spp when grown with FOS in culture (Ose *et al.*, 2018). Both of these strict anaerobic bacteria also colonise the infant gut within the first week of life (Yang *et al.*, 2021), and due to their similar metabolic abilities *Bacteroides* may occupy a similar niche to *Bifidobacteria*, and may be associated with a healthy infant microbiome.

Five ASVs classified as Negativicutes (ASVs 9, 91, 143, 151, 556), three of which were of the genera *Veillonella,* were enriched in the placebo group over the first year of life. This result is also supported by the enrichment of the Negativicutes class at 4-months, as well as Megasphaeraceae at 2-months in the placebo group. This class has been identified as a known propionate producer in the adult human gut (Reichardt *et al.*, 2014), but is not as commonly reported on in the infant microbiome; and in this study there was no difference in propionate concentrations between the randomised groups. In previous studies, compared to infants with feeding intolerance, both the Negativicutes class and *Veillonella* were significantly enriched in healthy infants (Hu *et al.*, 2021), and compared to vaginally born breastfed infants the Negativicutes class and *Veillonella* were significantly enriched in caesarean born, mixed fed (breastmilk and formula) infants (Liu *et al.*, 2019). Therefore, it is unclear what role this feature may play when enriched in the early infant microbiome. Although interestingly, Selenomodaceae (Negativicutes) were enriched in mothers within the placebo group at 36-weeks, indicating that potentially they are not as competitive in an environment that is supplemented by prebiotics. In the prebiotic group *Aerococcus* (Closteridia) was also enriched in both pregnant mothers and infants. While potential seeding of the infant microbiome from material gut microbiota has been assessed from stool samples using

metagenomic methods (Ferretti *et al.*, 2018; Ravi *et al.*, 2018), strain level variation in bacterial species requires that taxa are assess at this level. In this study, we did not identify any differentially abundant taxa at the strain level that were shared between mothers and infants, and because seeding was not a defined outcome of this study, no additional analyses of shared taxa were performed.

A longitudinal analysis of the infant microbiome including alpha diversity and SCFA concentrations demonstrated a dynamic community that coincided with a changing metabolic environment. Acetate was the predominate SCFA in the infant profile over time, and the highest concentrations were observed between 2 and 6-months of age, and the lowest concentrations were observed at 12-months. This is in agreement with other literature, showing acetate concentrations peaks at 3-months, and then reduce to lower concentrations by 12-months during maturation of the infant microbiome (Ziętek, Celewicz and Szczuko, 2021). Acetate and lactate are predominantly produced from the fermentation of breast milk (Bridgman *et al.*, 2017), and are therefore at elevated levels prior to the cessation of breastmilk feeding. On average, mothers in this study breastfed for 8.5(± 4.0)-months, which would explain the drop in acetic acid at 12-months. From 6 – 12-months, the concentration of both butyrate and propionate significantly increased in all infants. This increase coincided with a significant increase in richness and diversity of the microbiome over the same time period. As solid foods are first introduced, the microbial populations are exposed to more dietary diversity, and evolve to take advantage of the new energy sources. For example, a number of members of the Lachnospiraceae family produce both butyrate and propionate (Louis and Flint, 2017), and have been shown to increase in infants after the cessation of breastfeeding (Galazzo *et al.*, 2020). In this current study as well, the relative abundance of Lachnospirales increased from 4 to 6-months (7%) and from 6 to12-months (18.5%). As solid food intake increases breastfeeding is generally reduced, and with that, the transfer of bioactive compounds in breastmilk that regulate colonisation also reduce (Galazzo *et al.*, 2020). Additionally, microbial exposure from the environment (including pets, family environment, and foods) that begins at birth will be encountered more frequently as an infant ages, and contribute more heavily to the microbiome composition (Tamburini *et al.*, 2016). The continual expansion and diversification of the gut community is a necessary progression that will ideally lead to a relatively stable adult microbiome. Indeed, after accelerated maturation of the microbiome has begun (~6 months of age), reduced diversity in the infant microbiome is associated with higher incidence of allergy (Galazzo *et al.*, 2020), and increased diversity after ~6 months is more often associated with healthy outcomes (Matamoros *et al.*, 2013; Jakobsson *et al.*, 2014; Yassour *et al.*, 2016). However, it may be that the low diversity structure of initial colonisers) and a select group of bacteria, who perform evolutionarily evolved processes that benefit the development of the immune system and brain, may need to predominate in the gut for a certain period to establish host benefit. While exposure to microbes in the environment

is protective and necessary, potentially a rapid premature colonisation of more adult associated microbes facilitated by nutrients that support their growth, rather than a slower migration of microbes facilitated by partial breastfeeding may disrupt this critical process. The growth rates of potentially harmful or beneficial bacterial strains may also be an important consideration as there are lower levels of competition in the developing gut microbiome.

4.6.3 Stool form and other factors that are associated with maternal and infant microbiome composition over time

Birth mode has been suggested as one of the strongest determinants of microbiome composition in the first few months of life (Galazzo *et al.*, 2020). In this study, 56% of the infants were born via caesarean, and the other 44% were born vaginally. Using a DistLM, our results indicate that birth mode had an impact on the microbiome that could be detected up to 4-months of age. Furthermore, a significantly higher abundance of *Bacteroides* and *Phocaeicola* were found in vaginally born infants compared to caesarean born infants at 2-months of age. *Bacteroides* in particular have been shown to be strongly impacted by birth mode (Galazzo *et al.*, 2020), and have been detected at significantly higher abundances in the first week of life in vaginally born infants (Bäckhed *et al.*, 2015). The number of caesarean births in this study differed significantly by gender, with boys more frequently born via caesarean than girls. A higher birthweight is thought to increase the chances of caesarean delivery, however, only 3/25 male caesarean births were overweight (< 4.0 kg), and 2/15 female caesarean births were overweight, indicating birthweight alone cannot explain this finding. It is also important to point out that infants in the prebiotic group weighed significantly more than infants in the placebo group at 6-months, however, there was no significant difference in the number of caesarean births or overweight births between the prebiotic and placebo groups. A population study in Scotland has also found that after correcting for birthweight, males were still more likely to be born via caesarean section (Smith, 2000); and a more recent study found that after correcting for birthweight male babies have poorer intrapartum outcomes including higher rates of caesarean delivery and foetal distress (Dunn *et al.*, 2015). This is particularly relevant to the larger outcomes of this study – which are to assess the allergic outcomes of infants for a benefit attributed to the maternal prebiotic supplementation – as boys have an increased incidence of asthma (Fuseini and Newcomb, 2017), and food allergy (Kelly and Gangur, 2009) than girls. Supporting the development of the gut microbiome may be particularly important after caesarean delivery for promoting later health outcomes.

Breastfeeding has been shown to restore the microbiome of caesarean delivered infants, in part by increasing the abundance of *Bifidobacterium* (Guo *et al.*, 2020). Breastfeeding is well understood to provide optimal nutrition to the developing infant, supporting

growth, and providing both essential nutrients as well as commensal microorganisms to the gut microbiome (Chong, Bloomfield and O'Sullivan, 2018). Breastfed infants experience fewer incidence of diarrhoea (Guo *et al.*, 2020), and have lower incidence of atopic dermatitis in childhood (Friedman and Zeiger, 2005) compared to formula fed infants. At 4-months of age, we also found the proportion of *Bifidobacterium* to be higher in infants who were breastfed for at least 2-months, compared to those who were not. However, only 8 infants were breastfed less than 2-months so the lack of *Bifidobacterium* spp may also be due to under sampling. As the composition of breastmilk is influenced by diet (Bravi *et al.*, 2016), the prebiotic may also benefit infants who are breastfed by mothers in the prebiotic group, however, with the highly variable duration of breastfeeding (± 4-months), it was beyond the scope of this study.

Maternal and infants stools were categorised by stool form, which is a broad description of the water content of stool, and is associated with host health and transit time (Vandeputte *et al.*, 2016). Loose stools have a higher water content, and are associated with faster transit time, and impaired absorption of nutrients across the gut epithelium (Vandeputte *et al.*, 2016). Therefore, looser bowel movements may be more likely to flush opportunistic bacteria from the gut such as *Haemophilus*, which became more abundant in maternal stool with decreasing firmness at study baseline. On the other hand, longer transit times can increase stool firmness, and allow more time for water and nutrients to be taken up from the gut lumen. In this study Acutalibacteraceae gen, *Ruthenibacterium*, *Akkermansia*, and *Oscillibacter* became significantly more abundant with increasing stool firmness at study baseline. *Akkermansia* in particular has previously been shown to increase with stool firmness (Vandeputte *et al.*, 2016), and as a mucin degrader (Tailford *et al.*, 2015), that does not appear to benefit from the availability of oligosaccharides (Ose *et al.*, 2018), is likely to be more abundant in the mucosa. Bacteria that thrive in close proximity to the host epithelium may be therefore more likely to expelled with increasing stool firmness.

Across the entire study period, loose maternal stool also hosted a significantly lower bacterial diversity, and contained significantly higher SCFA concentrations. The opposite scenario, high diversity and low SCFA concentrations was observed in firm stools. Species richness has been shown previously to increase with increasing stool firmness (Vandeputte *et al.*, 2016), but to our knowledge, the relationship between stool form, alpha diversity and SCFA concentrations has only previously been assessed in vitro (Tottey *et al.*, 2017), and in chapter 2 and 3 of this thesis. In chapter 2, stool samples from adult women collected at two timepoints showed the same negative relationship between total SCFA concentrations and alpha-diversity, as well as associations between beta-diversity and stool form.

Stool form may also be impacted by the prebiotic supplement, as dietary fibres that resist digestion can increase faecal bulking in the large intestine (Simpson and Campbell, 2015). A systematic review of cereal fibre interventions also found these fibres significantly increase stool weight, frequency, and transit time (de Vries, Miller and Verbeke, 2015), which would improve stool form. Although, an improvement to stool form is expected to be more pronounced in individuals who consistently have more abnormal stool (very firm or very loose), compared to those with more normal stool. At baseline, the average stool form was both normal, and nearly identical for mothers in each of the intervention groups, and it remained similar between the two groups during pregnancy. At 4-months, stool form was significantly firmer in the placebo group compared to the prebiotic group, indicating a potential impact on stool form during the latter part of the intervention. One explanation may be that the gut was considerably disturbed over the pre- to postnatal period (as was evident in the significant drop in SCFA concentration, and shifts in the proportion of Clostridiales and Lachnospirales), and the prebiotic provided a buffer to this change, meaning mothers in the prebiotic group did not experienced a significant increase in stool firmness.

4.6.4 Limitations

In this intervention, the observed effect of the prebiotic in supporting the microbiome is limited by certain factors. The SCFA concentration and bacterial assemblages are expected to be more pronounced in individuals who are lacking in daily fibre consumption compared to individuals who may already be consuming adequate dietary fibre. Dietary patters are also expected to have a major influence on the structure and function of the microbiome. Diets with dramatic dietary differences such as vegetarian or vegan compared to omnivorous diets (Tomova *et al.*, 2019), plant based vs animal based (David *et al.*, 2014) or carbohydrate restricted compared to high-carbohydrate diets (Brinkworth *et al.*, 2009) have a profound effect on gut microbial communities. In a large study such as this, it may be difficult to collect and analyse a complete dietary breakdown. However, the benefit of conducting whole diet interventions to better understand the relationship between host health and disease cannot be underestimated; and closer collaboration between dietitians and microbiome researchers may aid in this pursuit. Additionally, one potential benefit of the maternal prebiotic on the infant microbiome that was not fully investigated with this dataset was differences in seeding patterns. If vertical transmission is possible, then both the metabolic and microbial environment of the mother will determine the very first step in the establishment of the gut microbiome. By assessing the meconium, the maternal impact on the infant microbiome, prior to the action of the environmental factors that influence the colonisation process, may become clearer. Lastly, because generalist microbes have the ability to utilise many different enzymes, they can change the metabolic profile of the gut without necessarily changing in absolute abundance. Therefore, assessing changes to

the total metabolomic environment may be another way to better capture changes induced by the prebiotics.

4.7 Conclusion

Over the study period which involved both the pre- and postpartum period, mothers displayed strong interpersonal differences and the prebiotic intervention made a significant impact to both the gut microbiome composition and SCFA concentrations. As anticipated, *Bifidobacterium* responded in the prebiotic group with increased abundance, and acetate concentrations significantly increase as well. For the first time in humans this study also shows a significant difference in the composition of the infant microbiome associated with a maternal prebiotic supplement. *Bacteroides caccae* was consistently enriched until 6-months of age within the prebiotic group, indicating potentially better growth conditions for this important early life commensal. Selenomodaceae (Negativicutes) were enriched in the maternal microbiome of mothers within the placebo group, and Negativecutes including *Veillonella* were significantly enriched in the microbiome of their infants. In the prebiotic group *Aerococcus* (Closteridia) was enriched in the gut microbiome of both pregnant mothers and infants. As an important driver of microbiota composition, the prebiotic appeared to have significant positive effects on the microbial community of both infants and adults. The results have implications for the impact of maternal diet on infant gut health, as they indicate that it is possible to intervene on the development of the infant microbiome in utero by modulation of the maternal gut microbiome. The long-term health of these infants will enable a better understanding of how this intervention may prevent childhood allergic disease, and how the early microbiome structure may play into the initial disease mechanism.

4.8 References

Anderson, M. J., Gorley, R. N. and Clarke, K. R. (2008) 'PERMANOVA+ for PRIMER: Guide to Software and Statistical Methods'. Plymouth, UK: PRIMER-E.

Arumugam, M. *et al*. (2011) 'Enterotypes of the human gut microbiome', *Nature*. Nature Publishing Group, 473(7346), pp. 174–180. doi: 10.1038/nature09944.

Bäckhed, F. *et al*. (2015) 'Dynamics and stabilization of the human gut microbiome during the first year of life', *Cell Host and Microbe*, 17(5), pp. 690–703. doi: 10.1016/j.chom.2015.04.004.

Ballard, O. and Morrow, A. L. (2013) 'Human Milk Composition: Nutrients and Bioactive Factors', *Pediatric clinics of North America*. NIH Public Access, 60(1), p. 49. doi: 10.1016/J.PCL.2012.10.002.

Beaumont, M. *et al*. (2017) 'Quantity and source of dietary protein influence metabolite production by gut microbiota and rectal mucosa gene expression: A randomized, parallel, double-blind trial in overweight humans', *American Journal of Clinical Nutrition*, 106(4), pp. 1005–1019. doi: 10.3945/ajcn.117.158816.

Bedu-Ferrari, C., Biscarrat, P., Langella, P., & Cherbuy, C. (2022) 'Prebiotics and the Human Gut Microbiota: From Breakdown Mechanisms to the Impact on Metabolic Health', *Nutrients* 14(10), p. 2096. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/nu14102096

Bielik, V. *et al*. (2020) 'Gut Microbiota Diversity in Lean Athletes Is Associated with Positive Energy Balance', *Human Nutrition: Research Article Ann Nutr Metab*, 76, pp. 242–250. doi: 10.1159/000509833.

Bravi, F. *et al*. (2016) 'Impact of maternal nutrition on breast-milk composition: a systematic review', *The American Journal of Clinical Nutrition*. Oxford Academic, 104(3), pp. 646–662. doi: 10.3945/AJCN.115.120881.

Bridgman, S. L. *et al*. (2017) 'Fecal Short-Chain Fatty Acid Variations by Breastfeeding Status in Infants at 4 Months: Differences in Relative versus Absolute Concentrations', *Frontiers in Nutrition*. Frontiers, 4, p. 11. doi: 10.3389/fnut.2017.00011.

Brinkworth, G. D. *et al*. (2009) 'Comparative effects of very low-carbohydrate, high-fat and high-carbohydrate, low-fat weight-loss diets on bowel habit and faecal short-chain fatty acids and bacterial populations', *The British journal of nutrition*. Br J Nutr, 101(10), pp. 1493–1502. doi: 10.1017/S0007114508094658.

Brosseau, C. *et al*. (2021) 'Prebiotic Supplementation During Pregnancy Modifies the Gut Microbiota and Increases Metabolites in Amniotic Fluid, Driving a Tolerogenic Environment In Utero', *Frontiers in Immunology*. Frontiers Media S.A., 12, p. 2857. doi: 10.3389/fimmu.2021.712614.

Cacho, N. T. and Lawrence, R. M. (2017) 'Innate Immunity and Breast Milk.', *Frontiers in immunology*. Frontiers Media SA, 8, p. 584. doi: 10.3389/fimmu.2017.00584.

Callahan, B. J. *et al*. (2016) 'DADA2: High-resolution sample inference from Illumina amplicon data', *Nature Methods*. Nature Publishing Group, 13(7), pp. 581–583. doi: 10.1038/nmeth.3869.

Caporaso, J. G. *et al*. (2011) 'Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 108(1), pp. 4516–4522. doi: 10.1073/pnas.1000080107.

Chiu, C. M. *et al*. (2014) 'Systematic analysis of the association between gut flora and obesity through high-throughput sequencing and bioinformatics approaches', *BioMed Research International*. Hindawi Limited. doi: 10.1155/2014/906168.

Chong, C., Bloomfield, F. and O'Sullivan, J. (2018) 'Factors Affecting Gastrointestinal Microbiome Development in Neonates', *Nutrients*. Multidisciplinary Digital Publishing Institute (MDPI), 10(3), p. 274. doi: 10.3390/nu10030274.

Christodoulides, S. *et al*. (2016) 'Systematic review with meta-analysis: effect of fibre supplementation on chronic idiopathic constipation in adults', *Alimentary Pharmacology & Therapeutics*. John Wiley & Sons, Ltd, 44(2), pp. 103–116. doi: 10.1111/APT.13662.

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

Costea, P. I. *et al*. (2017) 'Enterotypes in the landscape of gut microbial community composition', *Nature Microbiology 3*(1), pp. 8–16.

David, L. A. *et al*. (2014) 'Diet rapidly and reproducibly alters the human gut microbiome', *Nature*. Nature Publishing Group, 505(7484), pp. 559–563. doi: 10.1038/nature12820.

Dhariwal, A. *et al*. (2017) 'MicrobiomeAnalyst: A web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data', *Nucleic Acids Research*. Oxford University Press, 45(W1), pp. W180–W188. doi: 10.1093/nar/gkx295.

Dominguez-Bello, M. G. *et al*. (2016) 'Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer', *Nature Medicine*. NIH Public Access, 22(3), pp. 250–253. doi: 10.1038/nm.4039.

Dunn, L. *et al*. (2015) 'Gender specific intrapartum and neonatal outcomes for term babies', *European Journal of Obstetrics and Gynecology and Reproductive Biology*. Elsevier Ireland Ltd, 185, pp. 19–22. doi: 10.1016/J.EJOGRB.2014.11.035.

Fabiano, V. *et al*. (2021) 'Term infant formulas influencing gut microbiota: An overview', *Nutrients*. Multidisciplinary Digital Publishing Institute, p. 4200. doi: 10.3390/nu13124200.

Faria, A. M. C., Reis, B. S. and Mucida, D. (2017) 'Tissue adaptation: Implications for gut immunity and tolerance', *Journal of Experimental Medicine*. The Rockefeller University Press, pp. 1211–1226. doi: 10.1084/jem.20162014.

Ferretti, P. *et al*. (2018) 'Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome.', *Cell host & microbe*. Elsevier, 24(1), pp. 133-145.e5. doi: 10.1016/j.chom.2018.06.005.

Friedman, N. J. and Zeiger, R. S. (2005) 'The role of breast-feeding in the development of allergies and asthma', *Journal of Allergy and Clinical Immunology*, 115(6), pp. 1238–1248. doi: 10.1016/j.jaci.2005.01.069.

Fuseini, H. and Newcomb, D. C. (2017) 'Mechanisms driving gender differences in asthma', *Current allergy and asthma reports*. NIH Public Access, 17(3), p. 19. doi: 10.1007/S11882-017-0686-1.

Galazzo, G. *et al*. (2020) 'Development of the Microbiota and Associations With Birth Mode, Diet, and Atopic Disorders in a Longitudinal Analysis of Stool Samples, Collected From Infancy Through Early Childhood', *Gastroenterology*. W.B. Saunders, 158(6), pp. 1584–1596. doi: 10.1053/J.GASTRO.2020.01.024.

Gray, L. E. K. *et al*. (2017) 'The maternal diet, gut bacteria, and bacterial metabolites during pregnancy influence offspring asthma', *Frontiers in Immunology*. Frontiers Media SA. doi: 10.3389/fimmu.2017.00365.

Gritz, E. C. and Bhandari, V. (2015) 'The Human Neonatal Gut Microbiome: A Brief Review', *Frontiers in Pediatrics*. Frontiers Media SA, 3, p. 17. doi: 10.3389/fped.2015.00017.

Guo, C. *et al*. (2020) 'Breastfeeding restored the gut microbiota in caesarean section infants and lowered the infection risk in early life', *BMC Pediatrics*. BioMed Central Ltd, 20(1), pp. 1–6. doi: 10.1186/s12887-020-02433-x.

Hoeflinger, J. L. *et al*. (2015) 'In vitro impact of human milk oligosaccharides on Enterobacteriaceae growth', *Journal of Agricultural and Food Chemistry*. American Chemical Society, 63(12), pp. 3295–3302. doi: 10.1021/JF505721P/SUPPL_FILE/JF505721P_SI_001.PDF.

Holmes, Z. C. *et al*. (2022) 'Microbiota responses to different prebiotics are conserved within individuals and associated with habitual fiber intake', *Microbiome*, *10*(1), pp. 1–16. https://doi.org/10.1186/s40168-022-01307-x

Hu, X. *et al*. (2021) 'Altered gut microbiota is associated with feeding intolerance in preterm infants', *Turkish Journal of Pediatrics*. Turkish National Pediatric Society,

63(2), pp. 206–217. doi: 10.24953/TURKJPED.2021.02.004.

Jakobsson, H. E. *et al*. (2014) 'Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by Caesarean section', *Gut*. BMJ Publishing Group, 63(4), pp. 559–566. doi: 10.1136/gutjnl-2012-303249.

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

Jones, J. *et al*. (2021) 'Fecal sample collection methods and time of day impact microbiome composition and short chain fatty acid concentrations', *Scientific Reports*. Nature Publishing Group, 11(1), pp. 1–13. doi: 10.1038/s41598-021-93031-z.

Kelly, C. and Gangur, V. (2009) 'Sex Disparity in Food Allergy: Evidence from the PubMed Database', *Journal of Allergy*. Hindawi Publishing Corporation. doi: 10.1155/2009/159845.

Klimenko, N. S. *et al*. (2018) 'Microbiome Responses to an Uncontrolled Short-Term Diet Intervention in the Frame of the Citizen Science Project', *Nutrients*. MDPI AG, 10(5), p. 576. doi: 10.3390/nu10050576.

Knights, D., Ward, T. L., McKinlay, C. E., Miller, H., Gonzalez, A., McDonald, D., & Knight, R. (2014) 'Rethinking enterotypes', *Cell Host and Microbe* 16(4). pp. 433–437. https://doi.org/10.1016/j.chom.2014.09.013

Ladyman, S. R. and Brooks, V. L. (2021) 'Central actions of insulin during pregnancy and lactation', *Journal of Neuroendocrinology*. John Wiley & Sons, Ltd, p. e12946. doi: 10.1111/jne.12946.

Lin, H. and Peddada, S. Das (2020) 'Analysis of compositions of microbiomes with bias correction', *Nature Communications 2020 11:1*. Nature Publishing Group, 11(1), pp. 1–11. doi: 10.1038/s41467-020-17041-7.

Liu, F. *et al*. (2017) 'Fructooligosaccharide (FOS) and Galactooligosaccharide (GOS) Increase Bifidobacterium but Reduce Butyrate Producing Bacteria with Adverse Glycemic Metabolism in healthy young population', *Scientific Reports 2017 7:1*. Nature Publishing Group, 7(1), pp. 1–12. doi: 10.1038/s41598-017-10722-2.

Liu, Y. *et al*. (2019) 'The perturbation of infant gut microbiota caused by cesarean delivery is partially restored by exclusive breastfeeding', *Frontiers in Microbiology*. Frontiers Media S.A., 10(MAR), p. 598. doi: 10.3389/fmicb.2019.00598.

Louis, P. and Flint, H. J. (2017) 'Formation of propionate and butyrate by the human colonic microbiota', *Environmental Microbiology*. Blackwell Publishing Ltd, pp. 29–41. doi: 10.1111/1462-2920.13589.

Macfarlane, G. T., Steed, H. and Macfarlane, S. (2008) 'Bacterial metabolism and health-related effects of galacto-oligosaccharides and other prebiotics', *Journal of Applied Microbiology*, 104(2), pp. 305–344. doi: 10.1111/j.1365-2672.2007.03520.x.

Martin, C. and Li, J. (2017) 'Medicine is not health care, food is health care: plant metabolic engineering, diet and human health', *New Phytologist*, 216(3), pp. 699–719. doi: 10.1111/nph.14730.

Martin, M. (2011) 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet.journal*. EMBnet Stichting, 17(1), p. 10. doi: 10.14806/ej.17.1.200.

Martínez, I. *et al*. (2018) 'Experimental evaluation of the importance of colonization history in early-life gut microbiota assembly', *eLife*, 7. doi: 10.7554/eLife.36521.

Martinson, J. N. V. *et al*. (2019) 'Rethinking gut microbiome residency and the Enterobacteriaceae in healthy human adults', *The ISME Journal 2019 13:9*. Nature Publishing Group, 13(9), pp. 2306–2318. doi: 10.1038/s41396-019-0435-7.

Matamoros, S. *et al*. (2013) 'Development of intestinal microbiota in infants and its impact on health', *Trends in Microbiology*, 21(4), pp. 167–173. doi: 10.1016/j.tim.2012.12.001.

Mayengbam, S. *et al*. (2019) 'Impact of dietary fiber supplementation on modulating microbiota–host–metabolic axes in obesity', *Journal of Nutritional Biochemistry*. Elsevier Inc., 64, pp. 228–236. doi: 10.1016/J.JNUTBIO.2018.11.003.

McKnight, D. T. *et al*. (2019) 'microDecon: A highly accurate read-subtraction tool for the post-sequencing removal of contamination in metabarcoding studies', *Environmental DNA*. John Wiley & Sons, Ltd, 1(1), pp. 14–25. doi: 10.1002/edn3.11.

Mesa, M. D. *et al*. (2020) 'The evolving microbiome from pregnancy to early infancy: A comprehensive review', *Nutrients*. MDPI AG. doi: 10.3390/nu12010133.

Mínguez Pérez, M. and Benages Martínez, A. (2009) 'The Bristol scale - a useful system to assess stool form?', *Revista espanola de enfermedades digestivas*, 101(5), pp. 305–11. doi: 10.4321/S1130-01082009000500001.

Moles, L. and Otaegui, D. (2020) 'The impact of diet on microbiota evolution and human health. Is diet an adequate tool for microbiota modulation?', *Nutrients*. MDPI AG. doi: 10.3390/nu12061654.

Moore Simas, T. A. *et al*. (2013) 'Institute of medicine 2009 gestational weight gain guideline knowledge: Survey of obstetrics/gynecology and family medicine residents of the united states', *Birth*. NIH Public Access, 40(4), pp. 237–246. doi: 10.1111/birt.12061.

Moro, G. *et al*. (2006) 'A mixture of prebiotic oligosaccharides reduces the incidence of

atopic dermatitis during the first six months of age.', *Archives of disease in childhood*. BMJ Publishing Group, 91(10), pp. 814–9. doi: 10.1136/adc.2006.098251.

Mu, C. *et al*. (2015) 'Metabolomic analysis reveals distinct profiles in the plasma and urine of rats fed a high-protein diet', *Amino Acids*. Springer Vienna, 47(6), pp. 1225–1238. doi: 10.1007/s00726-015-1949-6.

Mutic, A. D. *et al*. (2017) 'The Postpartum Maternal and Newborn Microbiomes', *MCN The American Journal of Maternal/Child Nursing*. NIH Public Access, 42(6), pp. 326–331. doi: 10.1097/NMC.0000000000000374.

Ose, R. *et al*. (2018) 'The ability of human intestinal anaerobes to metabolize different oligosaccharides: Novel means for microbiota modulation?', *Anaerobe*. Academic Press, 51, pp. 110–119. doi: 10.1016/j.anaerobe.2018.04.018.

Pacifici, G. M. (2006) 'Placental transfer of antibiotics administered to the mother: A review', *International Journal of Clinical Pharmacology and Therapeutics*. Int J Clin Pharmacol Ther, pp. 57–63. doi: 10.5414/CPP44057.

Palmer, D. J. *et al*. (2022) 'Study Protocol for a Randomised Controlled Trial Investigating the Effects of Maternal Prebiotic Fibre Dietary Supplementation from Mid-Pregnancy to Six Months' Post-Partum on Child Allergic Disease Outcomes', *Nutrients*, 14(13), p. 2753. doi: 10.3390/nu14132753.

Pryde, S. E. *et al*. (2002) 'The microbiology of butyrate formation in the human colon', *FEMS Microbiology Letters*. John Wiley & Sons, Ltd, 217(2), pp. 133–139. doi: 10.1111/J.1574-6968.2002.TB11467.X.

Qin, S. *et al*. (2021) 'Distribution characteristics of intestinal microbiota during pregnancy and postpartum in healthy women', *Journal of Maternal-Fetal and Neonatal Medicine*. Taylor and Francis Ltd. doi: 10.1080/14767058.2020.1812571/SUPPL_FILE/IJMF_A_1812571_SM8908.DOCX.

Ravi, A. *et al*. (2018) 'Comparison of reduced metagenome and 16S rRNA gene sequencing for determination of genetic diversity and mother-child overlap of the gut associated microbiota', *Journal of Microbiological Methods*. Elsevier, 149, pp. 44–52. doi: 10.1016/j.mimet.2018.02.016.

Reichardt, N. *et al*. (2014) 'Phylogenetic distribution of three pathways for propionate production within the human gut microbiota', *The ISME Journal 2014 8:6*. Nature Publishing Group, 8(6), pp. 1323–1335. doi: 10.1038/ismej.2014.14.

Ríos-Covián, D. *et al*. (2016) 'Intestinal short chain fatty acids and their link with diet and human health', *Frontiers in Microbiology*. Frontiers Media S.A., 7(185), p. 185. doi: 10.3389/fmicb.2016.00185.

Roediger, W. E. W. (1980) 'Role of anaerobic bacteria in the metabolic welfare of the colonic mucosa in man', *Gut*. Gut, 21(9), pp. 793–798. doi: 10.1136/GUT.21.9.793.

Sawicki, C. M. *et al*. (2017) 'Dietary fiber and the human gut microbiota: Application of evidence mapping methodology', *Nutrients*. Multidisciplinary Digital Publishing Institute (MDPI), 9(125). doi: 10.3390/nu9020125.

Scholtens, P. A. M. J. *et al*. (2008) 'Fecal Secretory Immunoglobulin A Is Increased in Healthy Infants Who Receive a Formula with Short-Chain Galacto-Oligosaccharides and Long-Chain Fructo-Oligosaccharides', *The Journal of Nutrition*, 138(6), pp. 1141–1147. doi: 10.1093/jn/138.6.1141.

Scott, K. P., Duncan, S. H. and Flint, H. J. (2008) 'Dietary fibre and the gut microbiota', *Nutrition Bulletin*. John Wiley & Sons, Ltd, 33(3), pp. 201–211. doi: 10.1111/J.1467-3010.2008.00706.X.

Seth, E. C. and Taga, M. E. (2014) 'Nutrient cross-feeding in the microbial world', *Frontiers in Microbiology*. Frontiers Research Foundation, 5(JULY), p. 350. doi: 10.3389/FMICB.2014.00350/BIBTEX.

Simpson, H. L. and Campbell, B. J. (2015) 'Review article: dietary fibre–microbiota interactions', *Alimentary Pharmacology & Therapeutics*. John Wiley & Sons, Ltd, 42(2), pp. 158–179. doi: 10.1111/APT.13248.

Smith, G. C. S. (2000) 'A population study of birth weight and the risk of caesarean section: Scotland 1980–1996', *BJOG: An International Journal of Obstetrics and Gynaecology*. John Wiley & Sons, Ltd, 107(6), pp. 740–744. doi: 10.1111/j.1471-0528.2000.tb13334.x.

Stensballe, L. G. *et al*. (2013) 'Use of antibiotics during pregnancy increases the risk of asthma in early childhood', *Journal of Pediatrics*. Mosby, 162(4), pp. 832-838.e3. doi: 10.1016/j.jpeds.2012.09.049.

Stewart, C. J. *et al*. (2018) 'Temporal development of the gut microbiome in early childhood from the TEDDY study', *Nature*, pp. 583–588. doi: 10.1038/s41586-018-0617-x.

Stinson, L. F. *et al*. (2019) 'The not-so-sterile womb: Evidence that the human fetus is exposed to bacteria prior to birth', *Frontiers in Microbiology*. Frontiers Media S.A., 10(JUN), p. 1124. doi: 10.3389/FMICB.2019.01124/BIBTEX.

Tailford, L. E. *et al*. (2015) 'Mucin glycan foraging in the human gut microbiome', *Frontiers in Genetics*. Frontiers, 0(FEB), p. 81. doi: 10.3389/FGENE.2015.00081.

Tamburini, S. *et al*. (2016) 'The microbiome in early life: Implications for health outcomes', *Nature Medicine*, pp. 713–722. doi: 10.1038/nm.4142.

Tomova, A. *et al*. (2019) 'The effects of vegetarian and vegan diets on gut microbiota', *Frontiers in Nutrition*. Frontiers Media S.A., 6, p. 47. doi: 10.3389/FNUT.2019.00047/BIBTEX.

Tottey, W. *et al*. (2017) 'Colonic transit time is a driven force of the gut microbiota composition and metabolism: In vitro evidence', *Journal of Neurogastroenterology and Motility*. Journal of Neurogastroenterology and Motility, 23(1), pp. 124–134. doi: 10.5056/jnm16042.

Turck, D. *et al*. (2013) 'Breastfeeding: health benefits for child and mother', *Archives de Pediatrie*. Elsevier Masson SAS, 20(S2), pp. S29-48. doi: 10.1016/S0929-693X(13)72251-6.

Turner, S. *et al*. (1999) 'Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis', *Journal of Eukaryotic Microbiology*. John Wiley & Sons, Ltd (10.1111), 46(4), pp. 327–338. doi: 10.1111/j.1550-7408.1999.tb04612.x.

Vandeputte, D. *et al*. (2016) 'Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates', *Gut*. BMJ Publishing Group, 65(1), pp. 57–62. doi: 10.1136/gutjnl-2015-309618.

de Vries, J., Miller, P. E. and Verbeke, K. (2015) 'Effects of cereal fiber on bowel function: A systematic review of intervention trials', *World journal of gastroenterology*, 21(29), pp. 8952–8963. doi: 10.3748/wjg.v21.i29.8952.

Vulevic, J. *et al*. (2013) 'A Mixture of trans-Galactooligosaccharides Reduces Markers of Metabolic Syndrome and Modulates the Fecal Microbiota and Immune Function of Overweight Adults', *The Journal of Nutrition*. Oxford Academic, 143(3), pp. 324–331. doi: 10.3945/JN.112.166132.

Yadav, M. *et al*. (2018) 'A review of metabolic potential of human gut microbiome in human nutrition', *Archives of Microbiology*. Springer Berlin Heidelberg, 200(203), pp. 203–217. doi: 10.1007/s00203-017-1459-x.

Yan, R. *et al*. (2021) 'Dietary Fibre Intervention for Gut Microbiota, Sleep, and Mental Health in Adults with Irritable Bowel Syndrome: A Scoping Review', *Nutrients 2021, Vol. 13, Page 2159*. Multidisciplinary Digital Publishing Institute, 13(7), p. 2159. doi: 10.3390/NU13072159.

Yang, H. *et al*. (2020) 'Systematic analysis of gut microbiota in pregnant women and its correlations with individual heterogeneity', *npj Biofilms and Microbiomes*, 6(1). doi: 10.1038/s41522-020-00142-y.

Yang, L. *et al*. (2021) 'Recent advances of intestinal microbiota transmission from mother to infant', *Journal of Functional Foods*. Elsevier, 87, p. 104719. doi: 10.1016/J.JFF.2021.104719.

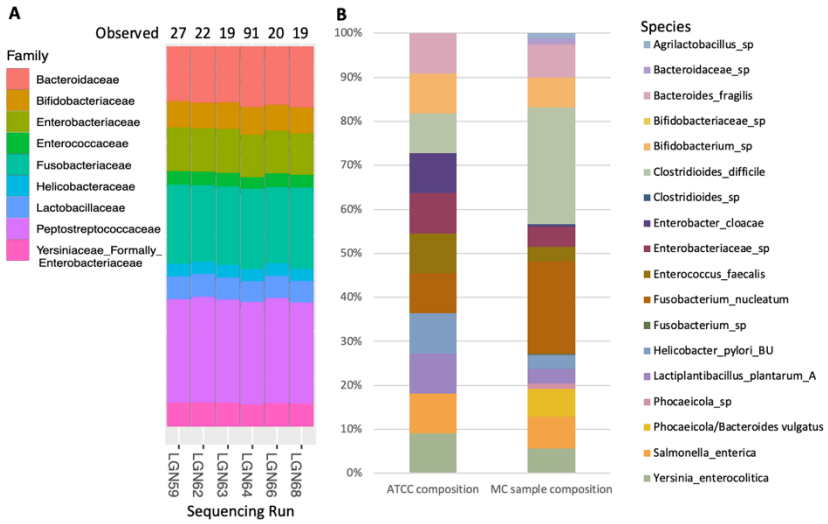
Yassour, M. *et al*. (2016) 'Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability.', *Science translational medicine*. NIH Public Access, 8(343), p. 343ra81. doi: 10.1126/scitranslmed.aad0917.

Zhou, Z., Cao, X. and Zhou, J. Y. H. (2013) 'Effect of resistant starch structure on short-chain fatty acids production by human gut microbiota fermentation in vitro', *Starch/Staerke*, 65(5–6), pp. 509–516. doi: 10.1002/star.201200166.

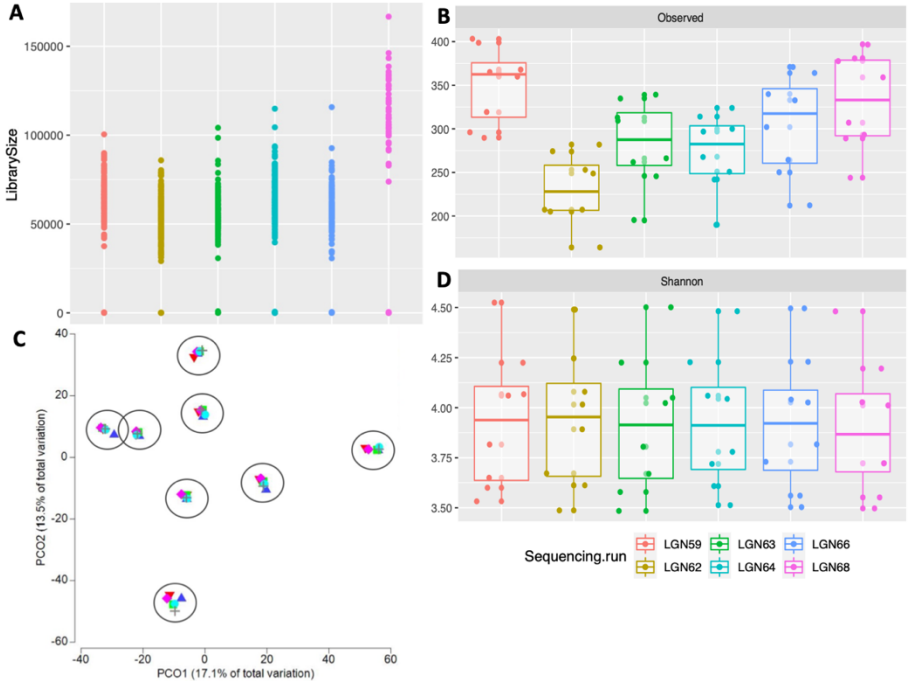
Ziętek, M., Celewicz, Z. and Szczuko, M. (2021) 'Short-chain fatty acids, maternal microbiota and metabolism in pregnancy', *Nutrients*. Multidisciplinary Digital Publishing Institute, p. 1244. doi: 10.3390/nu13041244.

4.9 Supplementary Material

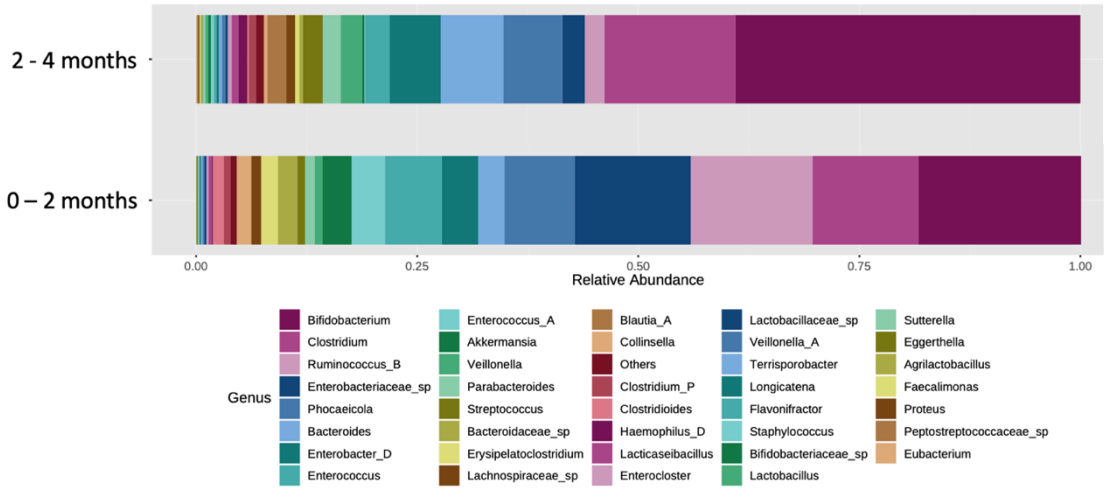
4.9.1 Supplementary figures



Supplamentary Figure 4.1. *Bacterial families recovered within the mock community sample across sequencing runs (A), and the proportion of reads assigned at the species level in the mock community sample compared to the known composition of the mock community (B).*



Supplementary Figure 4.2. *Library size of all samples and negative controls in each sequencing run (A). Richness (B), Shannon alpha-diversity (C), and beta-diversity of 8 replicate samples that were used as control samples in each sequencing run.*



Supplementary F *Distribution densities of maternal (A) and infant (B) microbiome samples according to sequencing batch number. Infant samples are also shown according to age, as 1-year old samples were all sequenced on run LGN 64****igure 4.3.*** *Relative abundance of bacterial genera in the microbiome community of infants breastfed for less than 2-months, and between 2- and 4-months.*

7.9.2 Supplementary tables

***Supplementary Table 4.1**. Read quality based on the proportion of reads with 99.9% base-call accuracy (Q30), and number of clusters passed filter (PF). The read count per sample, and the proportion of reads is shown before and after filtering steps. The number of ASVs are a count of the number of ASVs that appear in that sequencing run across all samples.*

Library	LGN59	LGN62	LGN63	LGN64	LGN66	LGN68	Average
Q30	89	83	90	90	90	91	89
PF	90	90	90	90	90	91	90
Reads PF	17.5	13.5	16	15	17.8	17.9	16.3
raw reads / sample (SD/1000)	99,804 (18.4)	72,318 (16.3)	86,619 (14.5)	89,814 (17.3)	90,619 (17.1)	171,488 (25.8)	101,777
% retained post filter	95.2	88.0	94.0	94.5	94.9	94.3	93.5
removal	70.0	85.5	69.7	77.6	68.3	69.7	73.5
range of reads per sample post filter	35,038 - 88,670	20,820 - 85,142	27,261 - 100,693	24,875 - 105,377	27,783 - 104,355	58,466 - 150,377	32,373 - 105,769
Number of ASVs	1,311	507	1,282	846	1,293	1,162	1,067

– Chapter 5 –
General discussion and conclusions

5.1 General Discussion

The ability to enjoy robust long-term health may be strongly impacted by the early development of the microbiome, and this thesis provides several new lines of evidence which support this working hypothesis. By looking at the microbiome composition of infants and children we discovered for the first time, shifts in gut microbiota associated with the earliest signs of autism, as well as evidence for a benefit to infants of mothers receiving a maternal prebiotic (GOS and FOS) supplement. Viewing human health through the microbiome lens has led to a deeper understanding of how our bodies are affected by environmental stressors, but this paradigm has many inherent limitations and challenges. Throughout this thesis, the manual, bioinformatic, and analytical biases have been carefully considered, mitigated, and discussed, resulting in robust microbiome data, which is presented along with detailed methodology that is in line with the current shift to improve data reproducibility.

While the inter-individual variation always exceeds the level of bias introduced by sample processing (contamination, extraction efficiency, subsampling), it is difficult to know how these factors might affect fine scale comparisons of the microbiome, such as differential abundance and beta-diversity, or the concentrations of profiled metabolites. To describe microbiome shifts associated with disease progression or therapeutic intervention, it was vital to ensure that methodological biases be eliminated as much as possible. The bias associated with stool subsampling was shown to be more pronounced in some individuals, and the effect of different commercial DNA stabilisation tubes on the microbiome composition was pronounced in some taxa. In line with this finding, larger volumes of frozen-only, homogenised stool were used for all further investigations in this thesis. Although, these factors may be more pronounced when comparing the microbiome in the same individuals over time, as an additive affect which may blur the intra-individual profile. Adding to this concern, alpha-diversity and SCFA concentrations were found to change considerably from the first to second bowel movement of the day, which could be partially mitigated by specifying a collection time (e.g. the first bowel movement of the day) or collecting all samples for a 24h period.

Throughout this thesis, alpha-diversity was not a good indicator of community association to ASD, although a difference in microbiome composition between children who were diagnosed with non-autistic autism spectrum disorder (NAASD) and children with ASD was indicated. Comparisons of alpha-diversity between randomised groups was also not as useful in monitoring the impact of a prebiotic supplement. It was,

however, valuable in describing changes that occurred over time. For example, between consecutive bowel movements and most notably, during the first year of life. A significant increase in both alpha-diversity and SCFA concentrations were notable, and changes were observed especially after 4-months of age, when infants begin transitioning from exclusive breast or formula feeding. Alpha-diversity was also strongly associated with stool form and SCFA in adult women and young children, indicating its usefulness in describing microbiome disruption during gastrointestinal distress (e.g constipation). Both alpha- and beta-diversity rely on sequencing depth, and the number of rare species (i.e. singletons or doubletons) in the dataset, with large numbers of both reads and rare species inflating diversity estimates. Both of these parameters can be manipulated by both sequencing strategy (i.e. number or samples per library) which directly impacts read depth, and prevalence filtering removes sparse species. As such, alpha-diversity in particular may not be appropriate to describe fine scale changes that are more susceptible to the noise of filtering parameters.

Differential abundance was however particularly useful in describing small-scale changes in microbiome composition. Significant differences between important bacterial strains was observed between both participants (both mothers and infants) who received either a prebiotic supplement or a placebo, as well as between autistic children and children with no developmental concerns. Most notably, Saccharomyces sp (ASV 21) was enriched in both ASD and non-autistic autism spectrum disorder (NAASD) groups compared to those in the NDC group. Interestingly, the large difference in microbiome composition associated with autism severity was driven by a loss in Actinobacteria in children who were diagnosed with non-autistic autism spectrum disorder (NAASD) compared to children with ASD. Furthermore, these differences were also observed in the predicted function profile, with NAASD children having elevated tetracycline biosynthesis. This finding highlighted the value of using differential abundance and predicted functional profiling. Although, this thesis also recognises the limitation of functional profiling when it is used without detailed participant characteristics. Without any knowledge of prior antibiotic use in these children, it was impossible to further speculate to why this difference was observed. Concerning differential abundance associated with prebiotic supplementation, for the first time we provide evidence that maternal prebiotic supplementation can impact the structure of the infant microbiome in the first year of life. Most notably, a strain identified as *Bacteroides caccae* was enriched from 2 to 4 and 6-months of age in the prebiotic group, and three strains identified as *Veillonella* were enriched at either 2, 4, or 6-months of age in the placebo group. There was also evidence suggesting members within the Negativicutes class were at a disadvantage in both mothers and infants who received the prebiotic supplement.

Large scale differences in overall community composition were also observed over time in the infant and maternal microbiome, as well as between young children showing early

behavioural signs of ASD who were receiving differing pre-emptive behavioural treatment (i.e. iBASIS-VIPP vs UCC). At the level of phylum, significantly tighter clustering of samples from the iBASIS-VIPP treatment group were observed, which indicates a more similar composition among children receiving this treatment. Also, at the level of genus significant differences in the microbiome beta-diversity was shown between iBASIS-VIPP and UCC treatment groups. While it is unclear how parent-infant interaction, in regards to the intervention, could impact the structure of the microbiome, the differences observed at multiple taxonomic levels between the two treatment groups indicate that there may be an unanticipated effect on the microbiome. It is not possible to comment on the benefit of one microbiome composition over the other, however, the bacterial structure appears to be less constrained in the UCC treatment group as compared to iBASIS-VIPP group, where the composition of the microbiome is more similar. The gut-brain-microbiota axis is the plausible route by which any differences in behaviours afforded by pre-emptive behavioural treatment might impact the microbiome structure; and with bi-directional communication between the brain and gut, perturbation or intervention at either point may achieve changes observed in both the gut and brain. If modifying the gut microbiome can affect brain function, and the way an individual interacts with the environment can in turn affect the gut microbiome, then potentially interventions which aim to effect neurological development and behaviour could have knock on effects which are detectable in the microbiome, especially at a young age when the microbiome is most susceptible to change.

Due to the integrated role the gut microbiota plays in human health and human immune system function, the therapeutic potential of supporting the microbiome is a growing research area. Prebiotics have very few mild side effects (e.g constipation and bloating), and are largely considered by scientists, regulators, and the general public as safe. Furthermore, because *Bifidobacterium* is a known target of GOS and FOS supplementation, and acetate is known to be a primary product of the fermentation strategy of this genus, an increase in this genus and acetate may be a useful indicator of microbiome response. During pregnancy, the prebiotic supplement made a significant impact on the composition of the maternal microbiome by increasing the abundance of *Bifidobacterium* and increased metabolic output of acetic acid, especially in the first 8-weeks of the intervention. Although interestingly, the abundance of *Bifidobacterium* appeared to be strongly influenced by the behavioural or physiological changes of post-pregnancy in both randomisation groups. In particular, the abundance of this genus dropped dramatically from the 36-weeks to 2-months in the prebiotic group, and over the same time-period in the placebo group, the abundance increased, resulting in both groups having nearly identical abundance of *Bifidobacterium* at 2-months. This observation indicates that there may be novel and important structural changes that happen to the maternal microbiome during the early breastfeeding period that warrant further investigation. One of the most impressive accomplishments of the SYMBA

study (Palmer *et al*., 2022)(the larger dataset from which the chapter 4 data came) is the vast number of biometric and biological samples that were collected from each of the > 600 women and infants in this study. Therefore, within this study it is likely that collaboration and further investigation using already collected data will reveal further insight on this question.

5.2 Limitations

While the abundance of particular taxa has been a valuable tool in the assessment of microbiome change throughout this thesis, understanding the implications of such changes are not always as straightforward as previously discussed with *Bifidobacterium*. ASD is complex, has a unique presentation in each individual, and our understanding of the development of this disorder in relation to brain development is not well understood, and the same is true of the gut microbiome. It is not surprising then that there is inconsistency in the literature which aims to characterise the gut microbiome of autistic children. Repeated non-parametric Mann-Whitney U tests were used to throughout this thesis, and are also commonly used in the microbiome literature (Xia and Sun, 2017), to detect taxa with significantly different abundance between groups. Although, repeated testing can result in a high number of false positives, especially when the groups being compared have a small number of samples, or when proportions of taxa are tested. False discovery rate corrections can eliminate the risk of inflated type 1 errors, although these corrections can be overly conservative (Knight *et al*., 2018). When the microbiome profiles of children were scrutinised for taxon abundance differences in chapter 3, several of the results were no longer significant after FDR correction. The results, both corrected and uncorrected were presented under the assumption that it was more beneficial to detect potentially associated microbes that could generate further hypothesis, rather than take the overly cautious conclusion of no differences. To further improve the reliability of these comparisons, we used: CLR transformed counts rather than proportions of taxa, improved sample collection methods developed in chapter 2, randomisation during sample processing and robust bioinformatic and taxonomic classification methods using positive and negative controls as a guide. To further remove other sources of grouping bias in this area of research, better characterization of diet and antibiotic use should be a priority, as well as consideration for the impact of age differences between infants and children. With more robust indicator species entering the catalogue of taxa associated with ASD, repeated characterisation across studies would accumulate more meaningful data.

Diet is also expected to have a major driving influence on the composition of the microbiome throughout life. Long-term dietary trends have a profound effect on the overall composition and function of the gut microbiome, while daily food choices can also cause fluctuations in microbiota composition. Understanding both short-term fibre

consumption, along with long-term diet type (e.g. vegetarian) would improve the interpretation of prebiotic supplement responses. Although, dietary fibre intake is expected to be more challenging to quantify, because fibre includes all carbohydrates (of at least 3 carbons) that resist digestion in the small intestine. For instance, individuals with regular low-fibre consumption likely respond more strongly to a prebiotic supplementation than those with high-fibre intake. Therefore, individuals with a typically high-fibre intake who struggle to maintain their diet during later stages of pregnancy, and individuals who typically maintain a low-fibre diet will demonstrate temporal shifts that will inflate the effectiveness of a prebiotic supplementation.

The safety and potential benefit of prebiotics has allowed for the first ever double-blind placebo-controlled RCT of the effectiveness of maternal prebiotic supplementation during foetal development, the most rapid and important developmental period of life. This study has made a significant contribution to the understanding of the early development of the microbiome and the impact of the pregnancy period. However, because of the inherent difference in long-term dietary trends between individuals (e.g. vegan, vegetarian, omnivore, carnivore), a brief dietary description would have been a valuable way of both explaining and correcting for these known differences. A maternal dietary category factor could have been used as a covariate when testing for differences in alpha diversity or SCFA concentrations between prebiotic and placebo groups. Furthermore, a more comprehensive assessment of short-term diet may be required during pregnancy when dietary shifts may progress over time. This information would be especially beneficial in trials of maternal prebiotic supplements, as it might help to explain individuals who first respond to the intervention (for example with an increase in *Bifidobacterium* and acetic acid), and then later appear not to respond. Likewise, dietary patterns in children with a high risk of developing autism is an important consideration, as autistic children often have very narrow food preferences (Yap *et al*., 2021). Therefore, a detailed 3-day diet diary may be particularly valuable to assist in interpreting microbiome shifts associated with autism.

Likewise, antibiotics can negatively impact the microbiome composition, and autistic children have a higher incidence of antibiotic use than neurotypical children (Krajmalnik-Brown *et al*., 2015). Also, antibiotics given during pregnancy can impact the infant microbiome colonization process (L. Yang *et al*., 2021). Information regarding antibiotic use in the young children at high risk of developing autism and the infants affected by prophylactic antibiotic used during pregnancy was not collected as part of the larger study protocols in chapter 3 and 4 respectively. Therefore, any difference attributed to the particular antibiotic use just mentioned cannot be identified. The smaller cohort of 30 children described in chapter 3 were potentially more susceptible to other confounding factors additional to antibiotic use. However, due to the pre-emptive recruitment method, it was impossible to select even numbers of children in

each diagnostic group, and most children fell within in the NAASD diagnosis group. In future, it would be advantageous to prospectively gather a larger cohort of children, collect both antibiotic and dietary information, and use sibling or age matched controls as a comparison group.

Lastly, while a large effort was made to rigorously conduct the metabarcoding analyses of the microbiome, and SCFA profiling from stool, the conclusions that can be drawn from this type of data have limitations. The microbiome composition refers to the number of sequenced microbial reads (which depends on a number of factors including differences in cell lysis, gene copy number, primer bias, the concentration of amplicon added to the Miseq, and cluster formation), and cannot represent the actual abundance of the microbiota in stool. While mean centring transformations reduce some of the bias resulting from describing reads in terms of retaliative abundance (proportions), the biases associated with PCR have been completely eliminated with whole-genome shotgun sequencing which does not rely on amplified DNA, although metagenome assembly is computationally much more difficult (Mas-Lloret *et al*., 2020).

Identifying microbial taxa using short gene regions – such as the V4 region of the 16S rRNA gene employed throughout this thesis – are also less accurate than taxonomic assignments resulting from shotgun sequencing, or long read sequencing such as PacBio. Taxonomic assignments using one or two hypervariable regions are often ambiguous at the species level, with multiple taxa sharing equal identity to the short reads. Throughout this thesis, to limit the use of unresolved species names, ASVs with lower taxonomic identities were discussed, and later (in chapter 3 and 4), taxa with multiple species ids were given the suffex 'spp' to identify them as unassigned due to ambiguity. With long read sequencing however, it is possible to sequence all nine hypervariable regions of the 16S gene, and make more accurate species assignments (Mas-Lloret *et al*., 2020). Metagenomic data also provides enhanced sequencing depth and taxonomic resolution, which would have been a valuable addition to chapter 4. With more accurate species identification and resolution *Bifidobacterium* could have been interrogated at the species level, potentially showing a strain dependent response to the prebiotic supplement.

Another benefit of whole genome sequencing is more confident functional profiling (van der Walt *et al.*, 2017). Functional predictions can be made using amplicon sequencing, as was done in chapter 3 of this thesis, however, these predictions rely on both the phylogenetic relationships between microbial taxa which can become blurry, and the pre-defined gene content of those near-neighbour species being included in the database used. Functional data resulting from whole genome sequencing can include a description of metabolic pathways or enzymes utilized by the microbial community. A combination of this functional and microbial abundance data can provide a clearer

picture of the microbiome response to an intervention, or to disease. For example, in Chapter 4, whole genome sequencing data would have allowed for differential abundance testing at the species level, as well as a description of the metabolic differences between the prebiotic and placebo groups. This would have been a notably powerful method for describing the microbiome response to a maternal prebiotic supplement; and potentially beneficial in explaining the enriched abundance of Negativicutes in the placebo group

5.3 Future directions

An assessment of stool form has been an invaluable factor in helping explain the shifts in microbiome composition throughout this thesis. Across all non-infant study participants (pregnant and non-pregnant women as well as young children), beta-diversity was strongly associated with stool form, and where numbers would allow, beta-diversity grouping were significantly impacted by stool form. Furthermore, both PCO plots and correlation coefficient plots (chapter 2) displayed clear clustering of microbiome profiles and microbiome-SCFA merged profiles respectively according to stool form. Additionally, the importance of butyrate formation, and its rapid uptake and utilisation by the host has been discussed through this thesis, and by incorporating stool form into the analysis, a pattern has emerged. Butyric acid shows an interesting inverse relationship with bacterial alpha-diversity which is associated with stool form. In pregnant women, firm stool hosted less bacterial diversity, and higher concentrations of both acetate and butyrate. In young children showing early signs of autism, both alpha-diversity, and acetate were also increased in firm stool, however butyric acid was depleted in firm stool. While the microbiome/metabolite patterns are expected to be somewhat different in children. This finding may potentially highlight a normal association between community diversity and butyric acid production or uptake that is impaired in the microbiome of these young children. Identifying stool form during sample processing was an easy way to obtain this additional information and gain further insight into this question.

The number of research articles pertaining to the human microbiome contained within PubMed has continued to increase each year since the beginning of the 21st Century (Prados-Bo and Casino, 2021). This expansion in research has been due in part to the reduced cost of sequencing and metabolomic technology (Ranjan *et al*., 2016), as well as better data integration and analysis methods (Na *et al*., 2021). However, while driving forward this exciting research area, mitigating the intrinsic biases in metabarcoding experiments has not received the same passionate interest. As we understand more about the ancient and vital role of symbiosis between microbes and humans, researchers have begun to develop microbiome-targeted products designed to improve and maintain quality of life. This research in turn has captured the attention of the commercial and

public audience, which has further fuelled both the interest and funding in this research area (Prados-Bo and Casino, 2021). However, when experiments are designed to support a hypothesis, especially in a field with very little methodological consensus, it may be quite easy to show support for an idea and at the same time, insure the experiment is difficult to both refute or replicate. This phenomenon is most obviously demonstrated by the stark lack of positive and negative controls (Hornung, Zwittink and Kuijper, 2019), as well as poor reporting in the methods section in microbiome research (Mirzayi *et al*., 2021). In this thesis, mock communities were use in each data chapter to identify a) families that were undetectable due to primer choice b) taxa that could not be resolved to lower phylogenetic ID, c) identify under represented microbial families, d) and to identify proportional changes to dominant reads across sequencing runs. Negative controls including both extraction and PCR non-template controls were also used in each chapter as input for bioinformatic sample decontamination.

Also, in this thesis I describe both structural and functional changes displayed by the gut microbiome that occur prior to a formal diagnosis of ASD, and profiled changes in the infant gut microbiome, which in the near future, will be analysed for any association with allergic disease. The microbial inhabitants play a complex role in the development of both the immune system and brain, and potentially a microbiome-based perspective might facilitate more overlap in our thinking between neurological and immune system disorders. Environmental exposure – including exposure to microbial metabolites – elicits dysregulated immune function and improper neuroimmune signalling in young autistic children. These include pro-inflammatory cytokine production (Masi *et al*., 2017), immune dysregulation of the BBB (Garcia-Gutierrez, Narbad and Rodríguez, 2020), and immune mediated mitochondria disturbance (Frye *et al*., 2015, 2017). Likewise, infection not only elicits a response by the immune system, but also includes neuronal symptoms such as itching and gastrointestinal motility, illustrating how the immune and nervous system respond to the environment in concert. Cytokines stimulate and enhance the excitability of peripheral nerves, and in this excited state, nerves will respond to non-noxious stimuli, releasing neuropeptides which are involved in the immune response (Kabata and Artis, 2019). Interestingly, communication between these two systems may be even more direct, due to the close anatomical proximity between neurons and immune mast cells that colocalise in the gut (Kabata and Artis, 2019). Therefore, it would be interesting to investigate periods of severe inflammation caused by gastrointestinal destress flair ups or allergies for their association with neurological issues such as anxiety or depression. Additionally, because autoimmune disease is a risk factor associated with ASD, and both diseases are associated with the microbiome, it could be revealing to investigate both neurological, immune system, and microbiome development together in infants. For example, with the data already collected as part of this thesis, an association between neurological development and microbiome composition could also be assessed. This could be achieved by correlating continuous

score data, such as the Mullens Scale of Early Learning, with microbiome abundance data. Together with allergen sensitisation and infant allergic disease outcome data, this analysis may reveal subtle microbial trends that are potentially shared between both immune and neurological dysregulation.

While the primary aim of chapter 4 was to assess the effectiveness of a maternal prebiotic supplement on the structure of the developing infant microbiome, the primary aim of the larger SYMBA study is to assess the effectiveness of the maternal prebiotic supplement on infant allergic disease outcomes. Once this primary outcome of the SYMBA trial is published, it would be fascinating to look back at the development of the microbiome and look for early signs of divergence between allergic and non-allergic infant groups. However, like the small changes observed between the young children with high behavioural risk of developing ASD, these subtle changes will be difficult to detect amongst the dramatic changes occurring over time. It would be most important to consider all factors that influence both the microbiome and immune system. This would include additional data not analysed to date in the SYMBA study, including pet ownership, prophylactic antibiotic use, and vaccination records. The significant impact of the infant microbiome on the immune response to vaccination has been widely studied and reviewed (de Jong, Olin and Pulendran, 2020); however the effect of vaccination on the microbiome composition – despite being called for (Ruck, Odumade and Smolen, 2020) – has only been investigated as far as I am aware, in animals (Hasegawa *et al*., 2018; Shi *et al*., 2022). These studies show significant increases in alpha-diversity, shifts in the metabolite profile, and shifts in bacterial membership including reduced *Bifidobacterium* and increased *Megasphaera*. With the integrated role the gut microbiota plays in both the development, and response of the immune system, it would be naive to assume there would be no impact on the infant microbiome from vaccination, which would confound a novel retrospective assessment of this kind.

A final point regarding early developmental health recognised in this thesis, is that birth mode was found to significantly impact the microbiome composition of infants, and this effect attributed to birth mode was confounded by the higher proportion of boys that were born by caesarean section. The effect of male gender on pregnancy outcomes including increased risk of caesarean birth has been observed previously (Antonakou and Papoutsis, 2016), but the disproportionate impact this phenomenon will have on the microbiome of boys has not been widely considered. Caesarean delivery is associated with a 23% increased odds of developing ASD (Curran *et al*., 2015), as well as an increased risk for developing allergies (Darabi *et al*., 2019). With the incidence of caesarean birth increasing by approximately 4% per year (Y. Yang *et al*., 2021), and the increasing prevalence of both ASD and allergy diagnosis, in future, it would be beneficial to collect information on birth mode when investigating the microbiome for association to diseases that impact boys and girls at different rates in early childhood.

Overall, this thesis has contributed to both our growing understanding of the early development of the microbiome and its necessity for long-term health, and to best-practices in generating microbiome data. From sample collection, to data analysis, each step in the workflow was personally carried out using advanced equipment, and continually updated protocols. By not outsourcing any of the laboratory or bioinformatic steps, every source of introduced bias was understood and controlled, or mitigated. The resulting microbiome profiles from each data chapter were incrementally more robust to align with the shift to improve data quality and reproducibility in microbiome research. Also, throughout each chapter alpha-diversity was found to be useful in describing largescale intra-individual changes over time, but poor for describing inter-individual differences. As a descriptive statistic, alpha-diversity may still be beneficial for describing community colonisation or depletion after antibiotics, however we advocate that in future, microbiome data should be analysed with less focus on shifts in alpha-diversity. We also found identifying stool form has been an invaluable method of capturing shifts in microbiome composition throughout this thesis. We have clearly established the significant association between stool form and beta-diversity, as well as stool form and SCFA concentrations. Going forward, not only could stool form data help to better describe GI symptoms, and dietary shifts reflected by the microbiome composition, but it may prove to be an adequate predictor of total gut transit time (particularly for very loose or firm stool), and assist in determining overall gut health. By collecting both stool form and transit time using a convenient blue dye 'blue poo' method (Asnicar *et al*., 2021) along with microbiome data it may be possible to further investigate the association between gut health and stool form.

5.4 References

Antonakou, A. and Papoutsis, D. (2016) 'The effect of fetal gender on the delivery outcome in primigravidae women with induced labours for all indications', *Journal of Clinical and Diagnostic Research*. JCDR Research & Publications Private Limited, 10(12), pp. QC22–QC25. doi: 10.7860/JCDR/2016/22099.9104.

Asnicar, F. *et al*. (2021) 'Blue poo: Impact of gut transit time on the gut microbiome using a novel marker', *Gut*. BMJ Publishing Group, 70(9), pp. 1665–1674. doi: 10.1136/gutjnl-2020-323877.

Curran, E. A. *et al*. (2015) 'Research review: Birth by caesarean section and development of autism spectrum disorder and attention-deficit/hyperactivity disorder: a systematic review and meta-analysis', *Journal of child psychology and psychiatry, and allied disciplines*. J Child Psychol Psychiatry, 56(5), pp. 500–508. doi: 10.1111/JCPP.12351.

Darabi, B. *et al*. (2019) 'The association between caesarean section and childhood asthma: An updated systematic review and meta-analysis', *Allergy, Asthma and Clinical*

Immunology. BioMed Central Ltd., 15(1), pp. 1–13. doi: 10.1186/S13223-019-0367-9/FIGURES/6.

Frye, R. E. *et al*. (2015) 'Gastrointestinal dysfunction in autism spectrum disorder: the role of the mitochondria and the enteric microbiome', *Microbial Ecology in Health & Disease*, 26. doi: 10.3402/mehd.v26.27458.

Frye, R. E. *et al*. (2017) 'Modulation of immunological pathways in autistic and neurotypical lymphoblastoid cell lines by the enteric microbiome metabolite propionic acid', *Frontiers in Immunology*. Frontiers Media S.A., 8(DEC). doi: 10.3389/fimmu.2017.01670.

Garcia-Gutierrez, E., Narbad, A. and Rodríguez, J. M. (2020) 'Autism Spectrum Disorder Associated With Gut Microbiota at Immune, Metabolomic, and Neuroactive Level', *Frontiers in Neuroscience*. Frontiers Media S.A., p. 1072. doi: 10.3389/fnins.2020.578666.

Hasegawa, Y. *et al*. (2018) 'Microbial structure and function in infant and juvenile rhesus macaques are primarily affected by age, not vaccination status', *Scientific Reports*. Nature Publishing Group, 8(1), pp. 1–11. doi: 10.1038/s41598-018-34019-0.

Hornung, B. V. H., Zwittink, R. D. and Kuijper, E. J. (2019) 'Issues and current standards of controls in microbiome research', *FEMS Microbiology Ecology*. Oxford University Press. doi: 10.1093/femsec/fiz045.

de Jong, S. E., Olin, A. and Pulendran, B. (2020) 'The Impact of the Microbiome on Immunity to Vaccination in Humans', *Cell Host and Microbe*. Cell Press, pp. 169–179. doi: 10.1016/j.chom.2020.06.014.

Kabata, H. and Artis, D. (2019) 'Neuro-immune crosstalk and allergic inflammation', *Journal of Clinical Investigation*. American Society for Clinical Investigation, 129(4), pp. 1475–1482. doi: 10.1172/JCI124609.

Knight, R. *et al*. (2018) 'Best practices for analysing microbiomes', *Nature Reviews Microbiology*. Nature Publishing Group, pp. 410–422. doi: 10.1038/s41579-018-0029-9.

Krajmalnik-Brown, R. *et al*. (2015) 'Gut bacteria in children with autism spectrum disorders: challenges and promise of studying how a complex community influences a complex disease', *Microbial Ecology in Health & Disease*, 26. doi: 10.3402/mehd.v26.26914.

Mas-Lloret, J. *et al*. (2020) 'Gut microbiome diversity detected by high-coverage 16S and shotgun sequencing of paired stool and colon sample', *Scientific Data*. Nature Publishing Group, 7(1), pp. 1–13. doi: 10.1038/s41597-020-0427-5.

Masi, A. *et al*. (2017) 'The Immune System, Cytokines, and Biomarkers in Autism Spectrum Disorder.', *Neuroscience bulletin*. Springer, 33(2), pp. 194–204. doi: 10.1007/s12264-017-0103-8.

Mirzayi, C. *et al*. (2021) 'Reporting guidelines for human microbiome research: the STORMS checklist', *Nature Medicine 2021 27:11*. Nature Publishing Group, 27(11), pp. 1885–1892. doi: 10.1038/s41591-021-01552-x.

Na, H. S. *et al*. (2021) 'Molecular subgroup of periodontitis revealed by integrated analysis of the microbiome and metabolome in a cross-sectional observational study', *Journal of Oral Microbiology*. Taylor and Francis Ltd., 13(1). doi: 10.1080/20002297.2021.1902707.

Palmer, D. J. *et al*. (2022) 'Study Protocol for a Randomised Controlled Trial Investigating the Effects of Maternal Prebiotic Fibre Dietary Supplementation from Mid-Pregnancy to Six Months' Post-Partum on Child Allergic Disease Outcomes', *Nutrients*, 14(13), p. 2753. doi: 10.3390/nu14132753.

Prados-Bo, A. and Casino, G. (2021) 'Microbiome research in general and business newspapers: How many microbiome articles are published and which study designs make the news the most?', *PLoS ONE*. Public Library of Science, 16(4 April). doi: 10.1371/journal.pone.0249835.

Ranjan, R. *et al*. (2016) 'Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing', *Biochemical and Biophysical Research Communications*. Academic Press, 469(4), pp. 967–977. doi: 10.1016/j.bbrc.2015.12.083.

Ruck, C. E., Odumade, O. A. and Smolen, K. K. (2020) 'Vaccine Interactions With the Infant Microbiome: Do They Define Health and Disease?', *Frontiers in Pediatrics*. Frontiers, 0, p. 811. doi: 10.3389/FPED.2020.565368.

Shi, J. *et al*. (2022) 'Effect of Killed PRRSV Vaccine on Gut Microbiota Diversity in Pigs', *Viruses 2022, Vol. 14, Page 1081*. Multidisciplinary Digital Publishing Institute, 14(5), p. 1081. doi: 10.3390/V14051081.

van der Walt, A. J. *et al*. (2017) 'Assembling metagenomes, one community at a time', *BMC Genomics*. BioMed Central Ltd., 18(1), pp. 1–13. doi: 10.1186/s12864-017-3918-9.

Xia, Y. and Sun, J. (2017) 'Hypothesis testing and statistical analysis of microbiome'. Chongqing Medical University, 4(3), pp. 138–148. doi: 10.1016/j.gendis.2017.06.001.

Yang, L. *et al*. (2021) 'Recent advances of intestinal microbiota transmission from mother to infant', *Journal of Functional Foods*. Elsevier, 87, p. 104719. doi: 10.1016/J.JFF.2021.104719.

Yang, Y. *et al*. (2021) 'Anesthesia, sex and miscarriage history may influence the association between cesarean delivery and autism spectrum disorder', *BMC Pediatrics*. BioMed Central Ltd, 21(1), pp. 1–10. doi: 10.1186/S12887-021-02518-1/FIGURES/4.

Yap, C. X. *et al*. (2021) 'Autism-related dietary preferences mediate autism-gut

microbiome associations', *Cell*. Elsevier, 184(24), pp. 5916-5931.e17. doi: 10.1016/j.cell.2021.10.015.

Appendices

Appendix 1. Copyright statements

The manuscript presented in chapter 2 is publicly available, and was published with Scientific Reports. This article was published under a Creative Commons Attribution 4.0 International License, which allows for maximum dissemination and re-use. The license agreement can be found here https://www.nature.com/srep/journal-policies/editorial-policies#license-agreement. The article presented in this thesis closely matches the peer-reviewed publicly available version. This article has been changed from its original format to align with the formatting of this thesis, including 1) main headings have been numbered and formatted in line with the thesis style 3) figure and table headings have been numbered and formatted in line with the thesis style, and 4) Figures and tables have been moved and resized from their original location to fit within the margins of this thesis.

To Whom it May Concern

I Jacquelyn M Jones contributed to the study concept, recruited all participants and performed all laboratory work, performed all data processing and analyses, and wrote and edited the resulting manuscript.

I as a co-author, endorse that this level of contribution by the candidate indicated above is appropriate, and give my permission for this manuscript to be included in this thesis.

Stacey N Reinke

Alishum Ali

Debra J Palmer

Claus T. Christophersen

The manuscript presented in chapter 3 is publicly available, and was published with Frontiers in Microbiology. This article was published under a Creative Commons Attribution 4.0 International License, which allows for maximum dissemination and re-use. The license agreement can be found here https://www.frontiersin.org/legal/copyright-statement. The article presented in this thesis closely matches the peer-reviewed publicly available version. This article has

been changed from its original format to align with the formatting of this thesis, including 1) main headings have been numbered and formatted in line with the thesis style 3) figure and table headings have been numbered and formatted in line with the thesis style, and 4) Figures and tables have been moved and resized from their original location to fit within the margins of this thesis.

To Whom it May Concern

I Jacquelyn M Jones contributed to the study concept, performed all sample processing and laboratory work, performed all data processing and analyses, and wrote and edited the resulting manuscript.

I as a co-author, endorse that this level of contribution by the candidate indicated above is appropriate, and give my permission for this manuscript to be included in this thesis.

Stacey N Reinke

Mahsa Mousavi-Derazmahalleh

Debra J Palmer

Claus T. Christophersen

Appendix 2. Storm checklists

Storm Checklist for the manuscript "Fecal sample collection methods and time of day impact microbiome and SCFA concentrations" presented in chapter 2.

No	Item	Recommendation	Item Source	Y/N/ NA
Abstract				
1.0	Structured or Unstructured Abstract	Abstract should include information on background, methods, results, and conclusions in structured or unstructured format.	STORMS	Yes
1.1	Study Design	State study design in abstract.	STORMS	No
1.2	Sequencing methods	State the strategy used for metagenomic classification.	STORMS	Yes
1.3	Specimens	Describe body site(s) studied.	STORMS	Yes
Introduction				

2.0	Background and Rationale	Summarize the underlying background, scientific evidence, or theory driving the current hypothesis as well as the study objectives.	STORMS	Yes
2.1	Hypotheses	State the pre-specified hypothesis. If the study is exploratory, state any pre-specified study objectives.	STORMS	Yes

Methods

3.0	Study Design	Describe the study design.	STORMS	Yes
3.1	Participants	State what the population of interest is, and the method by which participants are sampled from that population. Include relevant information on physiological state of the subjects or stage in the life history of disease under study when participants were sampled.	STORMS	Yes
3.2	Geographic location	State the geographic region(s) where participants were sampled from.	MIxS	No
3.3	Relevant Dates	State the start and end dates for recruitment, follow-up, and data collection.	STORMS	No
3.4	Eligibility criteria	List any criteria for inclusion and exclusion of recruited participants.	Modified STROBE	Yes
3.5	Antibiotics Usage	List what is known about antibiotics usage before or during sample collection.	STORMS	Yes
3.6	Analytic sample size	Explain how the final analytic sample size was calculated, including the number of cases and controls if relevant, and reasons for dropout at each stage of the study. This should include the number of individuals in whom microbiome sequencing was attempted and the number in whom microbiome sequencing was successful.	STORMS	Yes
3.7	Longitudinal Studies	For longitudinal studies, state how many follow-ups were conducted, describe sample size at follow-up by group or condition, and discuss any loss to follow-up.	STORMS	NA
3.8	Matching	For matched studies, give matching criteria.	Modified STROBE	NA
3.9	Ethics	State the name of the institutional review board that approved the study and protocols, protocol number and date of approval, and procedures for obtaining informed consent from participants.	STORMS	Yes
4.0	Laboratory methods	State the laboratory/center where laboratory work was done.	STORMS	Yes
4.1	Specimen collection	State the body site(s) sampled from and how specimens were collected.	MIxS	Yes
4.2	Shipping	Describe how samples were stored and shipped to the laboratory.	STORMS	Yes
4.3	Storage	Describe how the laboratory stored samples, including time between collection and storage and any preservation buffers or refrigeration used.	STORMS	Yes
4.4	DNA extraction	Provide DNA extraction method, including kit and version if relevant.	MIxS	Yes
4.5	Human DNA sequence depletion or microbial DNA enrichment	Describe whether human DNA sequence depletion or enrichment of microbial or viral DNA was performed.	STORMS	NA
4.6	Primer selection	Provide primer selection and DNA amplification methods as well as variable region sequenced (if applicable).	MIxS	Yes
4.7	Positive Controls	Describe any positive controls (mock communities) if used.	STORMS	Yes
4.8	Negative Controls	Describe any negative controls if used.	STORMS	Yes

4.9	Contaminant mitigation and identification	Provide any laboratory or computational methods used to control for or identify microbiome contamination from the environment, reagents, or laboratory.	STORMS	Yes
4.10	Replication	Describe any biological or technical replicates included in the sequencing, including which steps were replicated between them.	STORMS	Yes
4.11	Sequencing strategy	Major divisions of strategy, such as shotgun or amplicon sequencing.	MIxS	Yes
4.12	Sequencing methods	State whether experimental quantification was used (QMP/cell count based, spike-in based) or whether relative abundance methods were applied.	STORMS	Yes
4.13	Batch effects	Detail any blocking or randomization used in study design to avoid confounding of batches with exposures or outcomes. Discuss any likely sources of batch effects, if known.	STORMS	NA
4.14	Metatranscriptomics	Detail whether any mRNA enrichment was performed and whether/how retrotranscription was performed prior to sequencing. Provide size range of isolated transcripts. Describe whether the sequencing library was stranded or not. Provide details on sequencing methods and platforms.	STORMS	NA
4.15	Metaproteomics	Detail which protease was used for digestion. Provide details on proteomic methods and platforms (e.g. LC-MS/MS, instrument type, column type, mass range, resolution, scan speed, maximum injection time, isolation window, normalised collision energy, and resolution).	STORMS	NA
4.16	Metabolomics	Specify the analytic method used (such as nuclear magnetic resonance spectroscopy or mass spectrometry). For mass spectrometry, detail which fractions were obtained (polar and/or non polar) and how these were analyzed. Provide details on metabolomics methods and platforms (e.g. derivatization, instrument type, injection type, column type and instrument settings).	STORMS	NA
5.0	Data sources/ measurement	For each non-microbiome variable, including the health condition, intervention, or other variable of interest, state how it was defined, how it was measured or collected, and any transformations applied to the variable prior to analysis.	MIxS	NA
6.0	Research design for causal inference	Discuss any potential for confounding by variables that may influence both the outcome and exposure of interest. State any variables controlled for and the rationale for controlling for them.	STORMS	NA
6.1	Selection bias	Discuss potential for selection or survival bias.	STORMS	NA
7.0	Bioinformatic and Statistical Methods	Describe any transformations to quantitative variables used in analyses (e.g. use of percentages instead of counts, normalization, rarefaction, categorization).	STORMS	Yes
7.1	Quality Control	Describe any methods to identify or filter low quality reads or samples.	MIxS	Yes
7.2	Sequence analysis	Describe any taxonomic, functional profiling, or other sequence analysis performed.	MIxS	Yes
7.3	Statistical methods	Describe all statistical methods.	Modified STROBE	Yes
7.4	Longitudinal analysis	If the study is longitudinal, include a section that explicitly states what analysis methods were used (if any) to account for grouping of measurements by individual or patterns over time.	STORMS	NA
7.5	Subgroup analysis	Describe any methods used to examine subgroups and interactions.	STROBE	Yes

7.6	Missing data	Explain how missing data were addressed.	STROBE	Yes
7.7	Sensitivity analyses	Describe any sensitivity analyses.	STROBE	NA
7.8	Findings	State criteria used to select findings for reporting.	STORMS	Yes
7.9	Software	Cite all software (including read mapping software) and databases (including any used for taxonomic reference or annotating amplicons, if applicable) used. Include version numbers.	Modified STREGA	Yes
8.0	Reproducible research	Make a statement about whether and how others can reproduce the reported analysis.	STORMS	Yes
8.1	Raw data access	State where raw data may be accessed including demultiplexing information.	STORMS	Yes
8.2	Processed data access	State where processed data may be accessed.	STORMS	No
8.3	Participant data access	State where individual participant data such as demographics and other covariates may be accessed, and how they can be matched to the microbiome data.	STORMS	Yes
8.4	Source code access	State where code may be accessed.	STORMS	No
8.5	Full results	Provide full results of all analyses, in computer-readable format, in supplementary materials.	STORMS	No

Results

9.0	Descriptive data	Give characteristics of study participants (e.g. dietary, demographic, clinical, social) and information on exposures and potential confounders.	STROBE	Yes
10.0	Microbiome data	Report descriptive findings for microbiome analyses with all applicable outcomes and covariates.	STORMS	No
10.1	Taxonomy	Identify taxonomy using standardized taxon classifications that are sufficient to uniquely identify taxa.	STORMS	Yes
10.2	Differential abundance	Report results of differential abundance analysis by the variable of interest and (if applicable) by time, clearly indicating the direction of change and total number of taxa tested.	STORMS	Yes
10.3	Other data types	Report other data analyzed--e.g. metabolic function, functional potential, MAG assembly, and RNAseq.	STORMS	Yes
10.4	Other statistical analysis	Report any statistical data analysis not covered above.	STORMS	Yes

Discussion

11.0	Key results	Summarise key results with reference to study objectives	STROBE	No
12.0	Interpretation	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.	STROBE	No
13.0	Limitations	Discuss limitations of the study, taking into account sources of potential bias or imprecision.	STROBE	Yes
13.1	Bias	Discuss any potential for bias to influence study findings.	STORMS	Yes
13.2	Generalizability	Discuss the generalisability (external validity) of the study results	STROBE	Yes
14.0	Ongoing/future work	Describe potential future research or ongoing research based on the study's findings.	STORMS	Yes

Other information

15.0	Funding	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	STROBE	NA

15.1	Acknowledgements	Include acknowledgements of those who contributed to the research but did not meet critera for authorship.	STORMS	Yes
15.2	Conflicts of Interest	Include a conflicts of interest statement.	STORMS	Yes
16.0	Supplements	Indicate where supplements may be accessed and what materials they contain.	STORMS	Yes
17.0	Supplementary data	Provide supplementary data files of results with for all taxa and all outcome variables analyzed. Indicate the taxonomic level of all taxa.	STORMS	No

Storm Checklist for the manuscript "Changes to the gut microbiome in young children showing early behavioural signs of autism" presented in chapter 3.

No	Item	Recommendation	Item	Y/N /NA
Abstract				
1.0	Structured or Unstructured Abstract	Abstract should include information on background, methods, results, and conclusions in structured or unstructured format.	STORMS	Yes
1.1	Study Design	State study design in abstract.	STORMS	No
1.2	Sequencing methods	State the strategy used for metagenomic classification.	STORMS	No
1.3	Specimens	Describe body site(s) studied.	STORMS	Yes
Introduction				
2.0	Background and Rationale	Summarize the underlying background, scientific evidence, or theory driving the current hypothesis as well as the study objectives.	STORMS	Yes
2.1	Hypotheses	State the pre-specified hypothesis. If the study is exploratory, state any pre-specified study objectives.	STORMS	Yes
Methods				
3.0	Study Design	Describe the study design.	STORMS	Yes
3.1	Participants	State what the population of interest is, and the method by which participants are sampled from that population. Include relevant information on physiological state of the subjects or stage in the life history of disease under study when participants were sampled.	STORMS	Yes
3.2	Geographic location	State the geographic region(s) where participants were sampled from.	MIxS	Yes
3.3	Relevant Dates	State the start and end dates for recruitment, follow-up, and data collection.	STORMS	NA
3.4	Eligibility criteria	List any criteria for inclusion and exclusion of recruited participants.	Modified STROBE	NA
3.5	Antibiotics Usage	List what is known about antibiotics usage before or during sample collection.	STORMS	No
3.6	Analytic sample size	Explain how the final analytic sample size was calculated, including the number of cases and controls if relevant, and reasons for dropout at each stage of the study. This should include the number of individuals in whom microbiome sequencing was attempted and the number in whom microbiome sequencing was successful.	STORMS	Yes
3.7	Longitudinal Studies	For longitudinal studies, state how many follow-ups were conducted, describe sample size at follow-up by group or condition, and discuss any loss to follow-up.	STORMS	Yes
3.8	Matching	For matched studies, give matching criteria.	Modified STROBE	NA
3.9	Ethics	State the name of the institutional review board that approved the study and protocols, protocol number and date of approval, and procedures for obtaining informed consent from participants.	STORMS	Yes

4.0	Laboratory methods	State the laboratory/center where laboratory work was done.	STORMS	Yes
4.1	Specimen collection	State the body site(s) sampled from and how specimens were collected.	MIxS	Yes
4.2	Shipping	Describe how samples were stored and shipped to the laboratory.	STORMS	Yes
4.3	Storage	Describe how the laboratory stored samples, including time between collection and storage and any preservation buffers or refrigeration used.	STORMS	Yes
4.4	DNA extraction	Provide DNA extraction method, including kit and version if relevant.	MIxS	Yes
4.5	Human DNA sequence depletion or microbial DNA enrichment	Describe whether human DNA sequence depletion or enrichment of microbial or viral DNA was performed.	STORMS	Yes
4.6	Primer selection	Provide primer selection and DNA amplification methods as well as variable region sequenced (if applicable).	MIxS	Yes
4.7	Positive Controls	Describe any positive controls (mock communities) if used.	STORMS	Yes
4.8	Negative Controls	Describe any negative controls if used.	STORMS	Yes
4.9	Contaminant mitigation and identification	Provide any laboratory or computational methods used to control for or identify microbiome contamination from the environment, reagents, or laboratory.	STORMS	Yes
4.10	Replication	Describe any biological or technical replicates included in the sequencing, including which steps were replicated between them.	STORMS	Yes
4.11	Sequencing strategy	Major divisions of strategy, such as shotgun or amplicon sequencing.	MIxS	Yes
4.12	Sequencing methods	State whether experimental quantification was used (QMP/cell count based, spike-in based) or whether relative abundance methods were applied.	STORMS	Yes
4.13	Batch effects	Detail any blocking or randomization used in study design to avoid confounding of batches with exposures or outcomes. Discuss any likely sources of batch effects, if known.	STORMS	NA
4.14	Metatranscriptomics	Detail whether any mRNA enrichment was performed and whether/how retrotranscription was performed prior to sequencing. Provide size range of isolated transcripts. Describe whether the sequencing library was stranded or not. Provide details on sequencing methods and platforms.	STORMS	NA
4.15	Metaproteomics	Detail which protease was used for digestion. Provide details on proteomic methods and platforms (e.g. LC-MS/MS, instrument type, column type, mass range, resolution, scan speed, maximum injection time, isolation window, normalised collision energy, and resolution).	STORMS	NA
4.16	Metabolomics	Specify the analytic method used (such as nuclear magnetic resonance spectroscopy or mass spectrometry). For mass spectrometry, detail which fractions were obtained (polar and/or non polar) and how these were analyzed. Provide details on metabolomics methods and platforms (e.g. derivatization, instrument type, injection type, column type and instrument settings).	STORMS	NA
5.0	Data sources/ measurement	For each non-microbiome variable, including the health condition, intervention, or other variable of interest, state how it was defined, how it was measured	MIxS: host	Yes

		or collected, and any transformations applied to the variable prior to analysis.	disease status	
6.0	Research design for causal inference	Discuss any potential for confounding by variables that may influence both the outcome and exposure of interest. State any variables controlled for and the rationale for controlling for them.	STORMS	Yes
6.1	Selection bias	Discuss potential for selection or survival bias.	STORMS	NA
7.0	Bioinformatic and Statistical Methods	Describe any transformations to quantitative variables used in analyses (e.g. use of percentages instead of counts, normalization, rarefaction, categorization).	STORMS	Yes
7.1	Quality Control	Describe any methods to identify or filter low quality reads or samples.	MIxS: sequence quality check	Yes
7.2	Sequence analysis	Describe any taxonomic, functional profiling, or other sequence analysis performed.	MIxS: feature prediction ; similarity search method	Yes
7.3	Statistical methods	Describe all statistical methods.	Modified STROBE	Yes
7.4	Longitudinal analysis	If the study is longitudinal, include a section that explicitly states what analysis methods were used (if any) to account for grouping of measurements by individual or patterns over time.	STORMS	NA
7.5	Subgroup analysis	Describe any methods used to examine subgroups and interactions.	STROBE	NA
7.6	Missing data	Explain how missing data were addressed.	STROBE	Yes
7.7	Sensitivity analyses	Describe any sensitivity analyses.	STROBE	NA
7.8	Findings	State criteria used to select findings for reporting.	STORMS	Yes
7.9	Software	Cite all software (including read mapping software) and databases (including any used for taxonomic reference or annotating amplicons, if applicable) used. Include version numbers.	Modified STREGA	Yes
8.0	Reproducible research	Make a statement about whether and how others can reproduce the reported analysis.	STORMS	No
8.1	Raw data access	State where raw data may be accessed including demultiplexing information.	STORMS	Yes
8.2	Processed data access	State where processed data may be accessed.	STORMS	No
8.3	Participant data access	State where individual participant data such as demographics and other covariates may be accessed, and how they can be matched to the microbiome data.	STORMS	Yes
8.4	Source code access	State where code may be accessed.	STORMS	No
8.5	Full results	Provide full results of all analyses, in computer-readable format, in supplementary materials.	STORMS	No

Results

9.0	Descriptive data	Give characteristics of study participants (e.g. dietary, demographic, clinical, social) and information on exposures and potential confounders.	STROBE	Yes

10.0	Microbiome data	Report descriptive findings for microbiome analyses with all applicable outcomes and covariates.	STORMS	Yes
10.1	Taxonomy	Identify taxonomy using standardized taxon classifications that are sufficient to uniquely identify taxa.	STORMS	Yes
10.2	Differential abundance	Report results of differential abundance analysis by the variable of interest and (if applicable) by time, clearly indicating the direction of change and total number of taxa tested.	STORMS	Yes
10.3	Other data types	Report other data analyzed--e.g. metabolic function, functional potential, MAG assembly, and RNAseq.	STORMS	Yes
10.4	Other statistical analysis	Report any statistical data analysis not covered above.	STORMS	NA

Discussion

11.0	Key results	Summarise key results with reference to study objectives	STROBE	No
12.0	Interpretation	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.	STROBE	Yes
13.0	Limitations	Discuss limitations of the study, taking into account sources of potential bias or imprecision.	STROBE	Yes
13.1	Bias	Discuss any potential for bias to influence study findings.	STORMS	Yes
13.2	Generalizability	Discuss the generalisability (external validity) of the study results	STROBE	Yes
14.0	Ongoing/future work	Describe potential future research or ongoing research based on the study's findings.	STORMS	Yes

Other information

15.0	Funding	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	STROBE	NA
15.1	Acknowledgements	Include acknowledgements of those who contributed to the research but did not meet critera for authorship.	STORMS	Yes
15.2	Conflicts of Interest	Include a conflicts of interest statement.	STORMS	Yes
16.0	Supplements	Indicate where supplements may be accessed and what materials they contain.	STORMS	Yes
17.0	Supplementary data	Provide supplementary data files of results with for all taxa and all outcome variables analyzed. Indicate the taxonomic level of all taxa.	STORMS	No