SOLID FOOD INTRODUCTION AND THE INFANT GUT MICROBIOME

THE IMPACT OF SOLID FOOD INTRODUCTION ON MICROBIAL SUCCESSION IN THE INFANT GUT

By LUNA ELDAKIKY, B.SC.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Master of Sciences

McMaster University MASTER OF SCIENCE (2018) Hamilton, Ontario (Medical Sciences)

TITLE: The Impact of Solid Food Introduction on Microbial Succession in the Infant Gut

AUTHOR: Luna ElDakiky, B.Sc. (Honours)

SUPERVISOR: Professor Jennifer C. Stearns

NUMBER OF PAGES: CXII, 112

Abstract

Succession of the gut microbial community in infancy is marked by natural disturbances such as the introduction of solid food, the details and timing of which may be important factors in shaping the microbiome and, ultimately, human health. Previous gut microbiome studies focused on molecular profiles of bacterial communities generated through amplicon-based sequencing or metagenomics. Although providing a genus level profile of the microbiome, this approach does not functionally address microbial metabolism or interactions. This thesis explores the impact of solid food introduction on the developing gut microbiome of a breastfed, vaginally-born infant from the Baby & Mi prospective birth cohort. To investigate this, we designed a targeted culture strategy to complement molecular profiling and applied it to a longitudinal sample over a period of 25 days during the transition to solid foods. Upon observing community and species-level changes throughout the course of the study, we were prompted to explore strain-level and functional changes over that period. This was executed using bacterial functional phenotype screens, to measure bacterial strain metabolism of fiber substrates for genera of interest. Data show that our targeted culture strategy is more sensitive than 16S rRNA gene sequencing of stool alone at capturing bacteria of interest in the infant gut. Using this culture strategy, we show that alpha diversity is primarily driven by species evenness rather than richness following solid food introduction. Additionally, some phylotypes within the same genus show different patterns of change in relative abundance following solid food introduction.

To explore this further, we use cultured isolates from the longitudinal study and show that different isolates of the same species may utilize substrates differently (e.g. E*nterococcus faecalis*, *Bacteroides ovatus*) and others may utilize substrates in the same way (e.g. *Bifidobacterium* breve, *Paeniclostridium sordelli*). We also show that isolates of the same species acquired before solid food introduction may differ from ones acquired after (e.g. *Enterococcus faecalis*), or may remain the same (e.g. *Bifidobacterium breve*). Combining targeted-culturing with molecular profiling methods such as 16S rRNA gene sequencing and metagenomic approaches can provide us with more comprehensive and detailed functional data regarding how the gut microbiome responds to a disturbance, such as solid food introduction, during infancy.

Acknowledgements

I would like to, first and foremost, acknowledge my professor Dr. Jennifer Stearns for her constant support throughout the course of my thesis. Jen is an excellent mentor. Her door is open for her students at any time of the day. As a new Master's student to the Medical Sciences program, I was initially nervous about meeting the goals of my project. However, walking out of Jen's office every week, I felt much more confident and clear about my next move. Jen provided me with encouragement and unwavering support. She was my mentor every step of the way and made me strive to be best at what I do. Most importantly, Jen gave me the freedom and the resources to explore and pursue any direction that I desire when it comes to my research. Her vast knowledge and experience has helped me immensely at structuring my project, building my aims, and keeping up with my timeline. As Jen's first Master's student I can vouch that her vast knowledge, support, and guidance make her an excellent role model and mentor that can foster confidence, creativity, and persistence amongst her students. Thanks to her continuous help and support, I pushed myself to do things I never thought I was capable of. I would also like to thank the Baby&Mi team as well as my committee members Dr. Michael Surette and Dr. Lesley MacNeil for their valuable guidance and input which helped me drastically throughout the course of my thesis.

A big thank you to all my lab members as well as members of the Hynes lab for their support and valuable feedback on my work during the long hours of lab meeting presentations. During the two years of my masters Lucas Flett, Aman Patel, Clara Long, and Sara Dizzell were all my lab family and the people I spent the most time with. They

have significantly supported me in my project as well as emotionally when things got rough. Special thanks to Sara for her help in obtaining my samples and spending countless hours in the chamber to help me culture them and pick isolates. Additionally, I would like to thank 'big brother' Aman Patel for all his help during the execution of my project. Although Aman is an undergraduate student in our lab, he is always the one to go to if you need help. Thank you, Aman, for having the graduate students' backs at the lab and for being involved and knowing every aspect of my project. Your help has been invaluable to me and not enough coffees in the world can show that.

I would also like to thank Surette lab members Michelle Shah, for preforming the DNA extractions during the first aim of my study, as well as Laura Rossi, for helping in V3 16S rRNA gene sequencing and ID sequencing and running the data through the Sl1P pipeline.

Finally, I would like to thank my parents for their utmost support through the course of my thesis. Thank you, dad and Lillian, for being there for me when I needed someone to talk to when I was stressed. Thanks for your encouragement and for constantly believing in my potential. Special thanks to my mom who was there for me and was supportive despite being thousands of miles away. I extend my thanks to everyone who contributed to my work directly or indirectly, you all made my experience smoother, rewarding and much more memorable.

Table of Contents

List of figures and tables

List of abbreviations

AER	Aerobic condition
ANA	Anaerobic condition
AXOS	Arabinoxylan oligosaccharides
CMC	Carboxymethyl cellulose
FOS	Fructo-oligosaccharides
GALT	Gut-associated lymphoid tissue
GI	Gastrointestinal tract
HMO	Human milk oligosaccharides
HMP	Human microbiome project
IEC	Intestinal epithelial cells
LAB	Lactic acid-producing bacteria
MC	Microbial culture
ORF	Open reading frames
OTU	Operational taxonomic unit
PP	Plate pools
PUL	Polysaccharide utilization loci
RAPD	Random amplification of polymorphic DNA
rRNA	Ribosomal Ribonucleic acid
SCFA	Short-chain fatty acids
SNV	Single nucleotide variant
ST	Stool

Declaration of academic achievement

This study was primarily designed and executed by the author of this thesis, under the guidance and supervision of Dr. Jennifer C. Stearns. Aman Patel and Sara Dizzell significantly assisted with the lab work. Clara Long designed the functional substrate assays used in this study. Dr. Jennifer C. Stearns contributed considerably to the study's design, analysis and interpretation of the data.

1.0 Introduction

1.1 The human gut microbiome

1.1.1 Importance of the human gut microbiome

The human gut microbiome is composed of complex communities of microorganisms in close contact with their host. However, many gaps remain in our knowledge regarding how those communities form and function within the host. It is known that the gut microbiome plays an important role in human health and is responsible for a wide range of functions. Intestinal commensals are involved in host nutrition and metabolism. They synthesize important substances like vitamin K, vitamin B12, biotin, and various amino acids (Gorbach, 1996). Furthermore, the intestinal microbiome participates in the breakdown and fermentation of indigestible carbohydrates, like pectin and cellulose, into short-chain fatty acids (SCFAs) including propionate, butyrate, and acetate. SCFAs can then be broken down through β-oxidation to provide energy in the form of ATP to intestinal epithelial cells (IEC) (Bull & Plummer, 2014). The production of SCFAs, particularly butyrate, by the intestinal microbiome also contributes to IEC barrier integrity in the gut (Kelly *et al.,* 2015). In addition to providing an excellent source of energy to IECs, SCFA metabolism aids in lowering oxygen saturation in the lumen of the gut, contributing to a state of physiological hypoxia (Kelly *et al.,* 2015) which is important for maintaining the commensal microbiota, the majority of which are strict anaerobes.

In addition to host nutrition and metabolism, the intestinal microbes play a central role in protecting against foreign pathogens and shaping the gut's immune system.

Through competitive exclusion, commensal bacteria can outcompete incoming pathogens by consuming the available nutrients and occupying space in the lumen (Bernet *et al.,* 1994). Antimicrobial substances like bacteriocins and short-chain fatty acids that are secreted by both Gram positive and negative commensals further create an inhospitable environment for pathogens (Lievin, 2000). For example, Lactic acid-producing bacteria (LAB) such as *Enterococcus faecium*, *Lactococcus lactis*, and *Streptococcus mutans* synthesize bacteriocins that target and inhibit pathogens like *Salmonella enteritidis* (Bhardwaj *et al.,* 2010), *Clostridium difficile* (Rea *et al.,* 2007), and *Staphylococcus aureus* (Mota-Meira *et al.,* 2005), respectively. Commensal bacteria also play an active role in the maturation of important immune structures such as the gut-associated lymphoid tissue (GALT). Studies on germ-free mice reveal that in the absence of commensal microbes, GALT is severely under-developed (Houghteling & Walker, 2015). Peyer's patches are GALT present in the lamina propria of the distal gut that house a variety of immune cells including B and T lymphocytes, macrophages, and dendritic cells. Sufficient maturation of these immunological components insures proper sampling and presentation of various antigens within the lumen and the development of a homeostatic immune responses accordingly (Houghteling & Walker, 2015; Kamada *et al.,* 2013). As a result, the gut commensals, in conjunction with GALT, are involved in creating an immunological state of tolerance to various types of antigens (e.g. food) (Shen *et al.,* 2013).

1.2 Microbial ecology in the infant gut

1.2.1 Adult vs. infant gut microbiota

The adult gut microbiome is primarily shaped during infancy. Within the first year of life, the infant gut microbiome is characterized by an overall lower diversity, and a higher volatility than the adult gut microbiome, and is often predominated by members of the Actinobacteria and Proteobacteria phyla (Rodríguez *et al.,* 2015). The colonization patterns of the infant gut are highly dynamic with progressive changes in bacterial groups until stability is reached between one and three years of age (Rodríguez *et al.,* 2015; Bäckhed *et al.,* 2015; Yassour *et al.,* 2016). Eventually the infant gut microbiome begins to resemble that of adults in its diversity and stability. Changes in bacterial populations become slower and less chaotic and the Firmicutes and Bacteroidetes phyla begin to dominate in abundance (Rodríguez *et al.,* 2015). This pattern of colonization and succession is heavily influenced by a combination of genetic and environmental factors like mode of delivery, mode of feeding, antibiotic administration, and nutrition (Dominguez-Bello *et al.,* 2010; Penders *et al.,* 2006; Arrieta *et al.,* 2014; David *et al.,* 2013)

1.2.2 Establishment of the human gut microbiome

It was previously believed that colonization began immediately following birth and that the fetal gut was sterile *in utero*. However, analysis of meconium, a fetal excrement of swallowed amniotic fluid, reveals a distinctive low diversity microbiome within the samples (Ardissone *et al.,* 2014). This indicates that fetal exposure to microbes or DNA may indeed commence during gestation.

Following gestation, the mode of delivery plays a vital role in the colonization of the infant gut (Dominguez-Bello *et al.,* 2010; Penders *et al.,* 2006) with potential later health outcomes from interventions like Caesarean section (C-section) (Neu & Rushing, 2011; Chu *et al.,* 2017). Species of *Bacteroides* begin to appear in the infant gut around 10 days postpartum except when infants are born by C-section, in which case they appear many months later. Turroni et al. (2015) sequenced the V6-V8 hypervariable regions of the 16S rRNA gene from the stool of 11 infants, ranging in age between two and five months, and their mothers. Actinobacteria was dominant at approximately 88% in the infant gut, whereas Firmicutes were most abundant in the mothers. Bifidobacteriales was found to be the most abundant order in the infant gut microbiome at 80.6%, followed by Lactobacillale and Clostridiales (Turroni *et al.,* 2015).

Significant differences exist between the gut microbial profiles in breastfed and formula-fed infants. Breast milk contains human milk oligosaccharides (HMOs) that are fermented only by bifidobacteria (Coppa *et al.,* 2006). Compared with the formula-fed infants, breastfed infants have a higher abundance of species of bifidobacteria such as *B. longum*, as well as *Lactobacillus johnsonii, L. gasseri*, *L. paracasei* and *L. casei.* On the other hand, formula-fed infants have a less stable, more diverse gut microbial environment (Guaraldi & Salvatori, 2012), with lower counts of bifidobacteria and a much wider microbial spectrum containing species such as *Clostridium difficile, Citrobacter spp.* and *Enterobacter cloacae* (Backhed *et al.,* 2015). While *Bifidobacterium* appears to be the most predominant genus in breastfed infants prior to weaning, *Bacteroides* abundance noticeably increases following weaning and solid food

introduction (Wexler, 2007). This is due to the introduction of diverse of polysaccharides in solid food that can be fermented by *Bacteroides*.

1.2.3 Solid food introduction

Diet impacts the types of microbes present in the gut. In a study by David and colleagues (2013), ten adult participants were placed on either a plant-based diet (grains, vegetables, fruits) or an animal-based diet (eggs, cheese, meat) and monitored for 6 days. Participants on the animal-based diet showed an increase in bile-tolerant microbes such as Bacteroidetes (e.g. *Bacteroides*, *Alistipes*). In contrast, individuals consuming a plant-based diet had increased abundances of plant polysaccharide-metabolizing microbes such as Firmicutes (e.g. *Eubacterium rectale*, *Ruminococcus bromii)*. Those rapid shifts accompanying dietary changes suggest that different types of nutrients can favor the growth of certain microbes over others.

The solid food diet is introduced around 6 months of age in Canada and provides many new substrates to the gut bacteria such as insoluble fibre (e.g. cellulose, lignin, hemicelluloses), which can predominantly be found in whole cereals and vegetables (Agostoni *et al.,* 1995), and resistant starch, naturally present in cereal grains, unripe bananas, and rice (Fuentes-Zaragoza *et al.,* 2010). The introduction of solid food causes a major shift in the gut environment because the infant diet becomes supplemented with a complex selection of substrates aside from the ones found in breast milk which are broken down by a smaller subset of microbes (e.g. HMOs) (Arrieta *et al.,* 2014). Following solid food introduction, the infant gut microbiome shifts from being primarily predominated by milk-fermenting organisms (e.g. lactic acid bacteria like *Bifidobacterium*, *Lactobacillus*)

and enterobacteria to one with a higher bacterial species diversity, as members of the Firmicutes and Bacteroidetes, that can break down the new substrates (e.g. *Clostridium spp*. and *Bacteroides spp*.) (Bergstrom *et al.,* 2014; Koenig *et al.,* 2011). Additionally, *Bifidobacterium longum* and *Bifidobacterium breve*, that utilize HMOs, decrease as new substrates from solid foods favor the growth of *Bifidobacterium adolescentis and Bifidobacterium catenulatum* (Bergstrom *et al.,* 2014).

1.3 *Bifidobacterium* in the infant gut

1.3.1 Diversity

The *Bifidobacterium* genus is part of the Actinobacteria phylum, and in the order Bifidobacteriales (Zhang *et al.,* 2016). Members of this genus can be found in the gut of humans as well as mammals, insects and birds (Zhang *et al.,* 2016; Milani *et al.,* 2015). It is diverse, comprising approximately 39 species and 9 subspecies, which account for 75% of all the taxa within its order (Zhang *et al.,* 2016). The genus is divided into 7 clusters (*B. pseudolongum, B. bifidum, B. asteroides, B. pullorum , B. longum*, *B. boum* groups, and *B. adolescentis*) with genome sizes ranging from 1.73- 3.25Mb and approximately 1400- 2500 protein-encoding open reading frames (ORFs) (Zhang *et al.,* 2016; Milani *et al.,* 2015). 13% of all ORFs are predicted to encode enzymes associated with carbohydrate metabolism, 14% of these carbohydrate genes are are part of the *Bifidobacterium* core genome and hence are shared among all its sub(species) (Milani *et al.,* 2015). Those core genes are specialized in the utilization of either a specific or a wide set of host-and dietary-derived glycans (Milani *et al.,* 2015). *B. longum* is the dominant *Bifidobacterium* species in both the infant and the mother's gut microbiota at 56% and

38%, respectively. *B. bifidum* is the second most dominant bifidobacteria in the infant gut. Conversely, *B. adolescentis* appears to be the second most abundant in the mothers' stool **(**Turroni *et al.,* 2012).

1.3.2 Importance to the host

Bifidobacteria play an important role in host physiology and immunology. One of the most prominent features of bifidobacteria is its involvement in host carbohydrate metabolism. During infancy *B. longum* and *B. bifidum* predominate and aid in metabolizing human milk oligosaccharides (Sela *et al.,* 2008; O'Callaghan and van Sinderen, 2016). Following weaning, *B. longum* and *B. adolescentis* thrive due to their ability to break down diet-derived carbohydrates (Sela *et al.,* 2008; O'Callaghan and van Sinderen, 2016) Whereas, *B. bifidum* primarily shifts its metabolic profile to mucin degradation (Sela *et al.,* 2008; O'Callaghan and van Sinderen, 2016). Bifidobacteria participate in competitive exclusion of foreign pathogens through blocking pathogen binding sites on epithelial cells (O'Callaghan and van Sinderen, 2016; Tanaka *et al.,* 1983*).* Additionally, bifidobacteria are associated with reduced instances of bowel diseases such as colorectal cancer, inflammatory bowel disease, and ulcerative colitis (Pool-Zobel *et al.,* 1996; Venturi *et al.,* 1999).

1.3.3 Importance to the microbial ecology

The presence of bifidobacteria is important to the composition of the intestinal microbiota. Some studies highlight its involvement in cross-feeding processes in the gut. *Bifidobacterium longum* degrades arabinoxylan oligosaccharides (AXOS), a class of butyrogenic prebiotics, in the human gut converting it into acetate (Rivière *et al.,* 2015).

Eubacterium rectale, an acetate-utilizing microbe, subsequently converts the acetate byproduct by *B. longum* to butyrate, a short-chain fatty acid utilized by intestinal epithelial cells for energy (Rivière *et al.,* 2015). Moreover, *Bifidobacterium adolescentis* has been shown to break down inulin, fructo-oligosaccharides (FOS) and starch (Belenguer *et al.,* 2006; Ze *et al.,* 2012; Duncan *et al.,* 2004). Alternatively, some butyrate-producing bacteria in the gut, such as *Eubacterium rectale*, *Bacteroides thetaiotaomicron, Eubacterium hallii*, *and Anaerostipes caccae* are unable to metabolize those substances and instead utilize lactate and acetate byproducts generated by *B. adolescentis* (Belenguer *et al.,* 2006; Ze *et al.,* 2012; Duncan *et al.,* 2004). As a result, cross-feeding interactions between bifidobacteria and butyrate-producing microbes become important in controlling lactate accumulation and promoting butyrate production, both of which are vital to gut health (Duncan *et al.,* 2004).

1.4 *Bacteroides* in the infant gut

1.4.1 Diversity

The Bacteroidetes phylum is primarily comprised of *Bacteroides*, *Parabacteroides*, *Prevotella*, and *Alistipes* (Wexler & Goodman, 2017). These taxa encompass many of the microbes present in an adult gut environment (Wexler & Goodman, 2017).

The *Bacteroides* genus is limited to species belonging to the *Bacteroides fragilis* group, consisting of 10 species (*B. fragilis*, *B. caccae, B. thetaiotaomicron, B. distasonis*, *B. vulgatus*, *B. eggerrthii*, *B. merdae*, *B. stercoris*, *B. uniformis*, and *B. ovatus)* (Wexler, 2007). *Bacteroides* species (e.g. *B. thetaiotaomicron* and *B. ovatus*) are diverse

and can be isolated from humans as well as other mammalian hosts such as cows, goats, and pigs. (Atherly & Ziemer, 2014). The species share approximately 1085 common core proteins, many of which are involved in carbohydrate metabolism and transport (Karlsson *et al.,* 2011). Polysaccharide utilization loci (PULs) are clusters found to be unique in *Bacteroides* species (Martens *et al.,* 2008; Sonnenburg *et al.,* 2010). PULs contain a variety of components (e.g. carbohydrate esterases, glycoside hydrolases) that allow *Bacteroides* spp. to utilize a wide range of plant and host-derived glycans (e.g. inulin, mucin, levan-type fructans, etc.) (Martens *et al.,* 2008; Sonnenburg *et al.,* 2010).

1.4.2 importance to the host

Species of *Bacteroides* adapt well to changes in the host environment. Species such as *B. theaiotaomicron* and *B. fragilis,* that are present before weaning, switch their metabolic profiles to adapt to changes in nutrient availability within the host (Sonnenburg *et al.,* 2005). *B. thetaiotaomicron* contain enzymes that can break down HMO's in breastmilk (Wexler, 2007). Following weaning, such species expand their metabolic repertoire and are capable of fermenting indigestible plant-derived polysaccharides (*Sonnenburg et al., 2005*). Additionally, *Bacteroides* play a role in competitive exclusion of pathogens (Averlar *et al.,* 1999; Mossie *et al.,* 1981), aid in the activation of CD4+ T- cell responses (Mazmanian & Kasper, 2006), produce bacteriocidal lectins that assist in Gram positive bacterial killing (Averlar *et al.,* 1999; Mossie *et al.,* 1981), and assist in the maturation of GALT (Rhee *et al.,* 2004).

1.4.3 importance to the microbial ecology

While species of bifidobacteria like *B. longum* are efficient at breaking down sugars in breastmilk (e.g. lactose, human milk oligosaccharides), *Bacteroides* species can ferment more complex plant and host-derived glycans (i.e pectin, mucin) (Martens *et al.,* 2008; Sonnenburg *et al.,* 2010). For example, *B. thetaiotamicron* contains a variety of glycosyl hydrolases that can catalyze the breakdown of a wide range of substances, including mucins. Gut microbes that lack those enzymes, like *E. coli*, benefit from this by utilizing the byproducts of mucin breakdown (e.g. mucosal-derived monosaccharides) (Sonnenburg *et al.,* 2004). Additionally, *Bacteroides* appear to be crucial in the establishment of gut microbiome communities by helping create an anaerobic environment for themselves and other obligate anaerobes in the gut (Wexler and Goodman, 2017). For instance, *Bacteroides* encodes a cytochrome bd oxidase with a high affinity for oxygen *(*Baughn & Malamy, 2004; Meehan *et al.,* 2012). Cytochrome bd oxidase can bind nanomolar levels of O_2 in the gut lumen and catalyze its reduction into two water molecules (Giuffré *et al.,* 2014). This helps to reduce intestinal oxygen levels, making the gut environment a hospitable niche for obligate anaerobes (Baughn & Malamy, *2004;* Meehan *et al.,* 2012).

1.5 Current tools to study the gut microbiome

1.5.1 Molecular profiling

The Human Microbiome Project (HMP) is the National Institutes of Health intiative to study the effect of the human microbiome (e.g. mouth, esophagus, stomach, colon, and vagina) on states of health and disease (Turnbaugh *et al.,* 2007). Using

molecular profiling methods, such as 16S rRNA gene sequencing and shotgun metagenomics, it reports that the gastrointestinal tract (GI) houses the largest number of microbes in the human body, with the colon being the most densely populated site in the GI (Turnbaugh *et al.,* 2007). Such methods have been used widely to study the microbes in the human gut. The 16S rRNA gene (1500bp) is a gene present in all bacteria, making it a universal marker for identifying them (Janda & Abbott, 2007; Patel, 2001). It is composed of regions that are conserved between prokaryotic taxa, as well as 9 hypervariable regions (V1-V9), ranging in size from 30-100bp, that can be used as a phylogenetic tool to distinguish between microbial genera (Chakravorty *et al.,* 2007). Moreover, not every region can distinguish between all bacteria simultaneously and different regions appear to have different discriminatory power (Chakravorty *et al.,* 2007). Hypervariable regions V2 and V3 can discriminating best between 110 bacterial species except in the case of closely related Enterobacteriaceae (Chakravorty *et al.,* 2007). The next molecular profiling option is shotgun metagenomics, a genomic analysis method of the collective microbiome found in environmental samples (e.g. stool, soil, oral cavity) (Chistoserdova, 2013). All microbial genes are sequenced with this method, allowing the study of new genes of interest and offering better taxonomic and functional resolution than does the 16S rRNA gene method (Jovel *et al.,* 2016; Chistoserdova, 2013). Unlike 16S rRNA gene sequencing, metagenomics can be used to dissect populations of eukaryotic, archaeal, and viral microbes in addition to bacteria (Breitbart *et al.,* 2003; Jovel *et al.,* 2016). Metagenomics is much more expensive, however, and analysis of the resulting data requires more computational power than for 16S rRNA

sequencing (Jovel *et al.,* 2016). Therefore, 16S rRNA gene sequencing remains the most efficient option to profile gut microbial communities.

1.5.2 Microbial culture and enriched molecular profiling

Culture enriched-molecular profiling is a method that combines targeted culturing and 16S rRNA gene sequencing (Lau *et al.,* 2016). A study performed by Lau et al (2016) revealed that by using culture enriched-molecular profiling a higher OTU count was recovered than by using 16S rRNA gene sequencing of stool alone (Lau *et al.,* 2016; Lagier *et al.,* 2012). This is attributed to the fact that 'targeted' culturing, as the name suggests, comprises an assembly of selective and non-selective microbial culture conditions that are effective at targeting either a specific or a wide subset of microbes. Therefore, in addition to identifying more OTUs than sequencing alone, culture allows access to microbial isolates that we can study *in vitro* (Lau *et al.,* 2016). Culture-enriched molecular profiling is the use of sequencing (either the 16S rRNA gene or with shotgun sequencing) of material collected through microbial culture of a sample. The combination of molecular profiling with microbial culture offers high sensitivity and discrimination of viable from dead organisms within a sample. Culture has its drawbacks: (a) culture requires the use of fresh samples as they contain higher amounts of live microbes (Lau *et al.,* 2016) (b) culture is biased since nutrients in the media could be selecting for certain types of microbes, or simply missing others by selecting against a microbial partner needed for survival (Lau *et al.,* 2016). Therefore, profiles are often compared between direct sequencing of stool and culture enriched methods.

1.5.3 Tools for assessing strain dynamics

As discussed in the previous section, shotgun metagenomics offers higher taxonomical and functional resolution than 16S rRNA gene sequencing (Jovel *et al.,* 2016). This means that while 16S rRNA gene sequencing can distinguish microbes at the genus (and sometimes species) level, metagenomics can provide species and strain-level resolution (Jovel *et al.,* 2016; Chistoserdova, 2013). Costea et al (2017) applied metaSNV, a single nucleotide variant (SNV) analysis pipeline, to 676 shotgun metagenomics samples from the human oral cavity, showing that strain populations acquired from the tongue dorsum cluster differently than strains from the supragingival plaque. Another study by Sharon et al (2012) used metagenomics to track species and strain-level differences in 11 stool samples of premature infants. Results at the strain-level reveal that relative abundances of three *Staphylococcus epidermidis* strains and the phages that infect them vary during the first month of life (Sharon *et al.,* 2012). Moreover, results at the species level show that there is a shift from *Propionibacterium acnes* strains, dominating at earlier stages of colonization, to new strains of *Propionibacterium* and *Peptoniphilus* species which increased in abundance later during the first month (Sharon *et al.,* 2012). Shotgun metagenomics, therefore, is a method to track strain diversity and dynamics.

1.6 Research question

Succession of the gut microbial community in infancy is punctuated with natural disturbances such as the introduction of solid food and the cessation of breastfeeding, the details and timing of which may be important factors in the final shape of the microbiome

and, ultimately, human health. Previous gut microbiome studies have focused on molecular profiles of bacterial communities generated through amplicon based sequencing (16S rRNA gene) or metagenomics. Although providing a good overview of the microbiome, these approaches do not functionally address microbial metabolism or interactions. This thesis explores the impact of solid food introduction on the developing gut microbiome of a breastfed, vaginally-born infant from the Baby & Mi prospective birth cohort study. I investigated this research questions with the use of a combined approach involving culture and molecular profiling, to determine the effects of solid food on the bacterial community as a whole, and bacterial functional phenotype screens, to measure bacterial strain metabolism of fibre substrates for genera of interest.

1.6.1 Aims

Aim 1: Design a targeted bacterial culture strategy to capture bacterial isolates from the infant gut

Rationale: most studies aimed at exploring the diversity of the gut microbiome rely on the use of culture-independent molecular-based methods, such as 16S rRNA gene sequencing. Although this method provides some taxonomic information about the bacteria present in a sample the region sequenced often does not provide sufficient taxonomic resolution to identify strain or even species level assignments. A marker-based profiling method such as 16S rRNA gene sequencing can provide only limited information about bacterial function. In order to study the effect of solid food introduction on the functional changes in the gut microbiome we aimed to add a targeted culture-based strategy to molecular profiling in order to obtain bacterial isolates for further in vitro study. In addition

to the benefits of culture for securing bacterial isolates that can be studied in the lab, the addition of culture to a molecular profile strategy, termed culture-enriched 16S rRNA gene profiling, can increase the sensitivity of the method for lower abundance bacterial populations (Lau *et al.,* 2016). Here we aimed to test a culture strategy that would capture the dominant bacterial taxa within the gastrointestinal tract of infants in the Baby & Mi prospective birth cohort (Simioni *et al.,* 2016) at the time of introduction of solid foods (between 12 weeks and 1 year of age). The 50 most abundant families of bacteria (**Table 1**) were determined from Baby & Mi cohort 16S rRNA gene data at 5 months and 1 year of age (unpublished) and a targeted culture strategy was devised based on the bacterial culture protocols presented in Lau *et al.,* (2016).

Structural and functional attributes of members of those microbial families, shown in **Figure 1**, are important to determine which media can be used to target those organisms of interest. The characteristics include shape (cocci vs. bacilli), cell wall structure (Gram positive/negative) type of metabolism (Obligate/facultative, Aerobic/Anaerobic), and nutritional requirements (fastidious or otherwise).

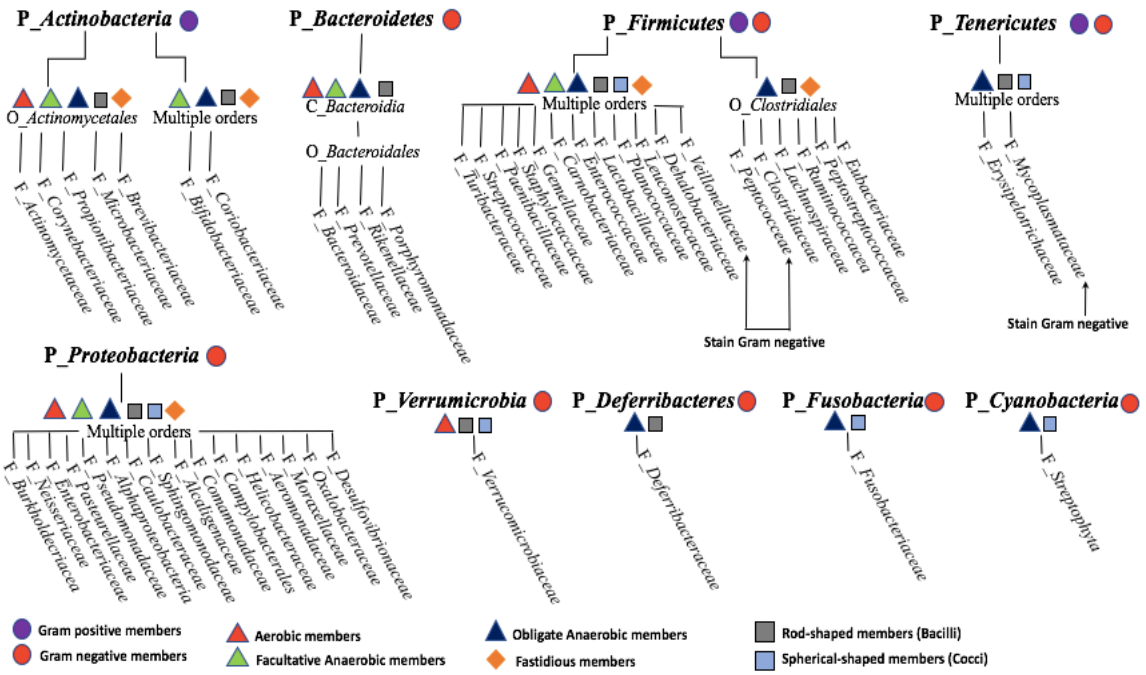


Figure 1: Microbial classes, orders and families of interest, listed in Table 1. Structural and Functional classifications (denoted as symbols) are placed at the class/order level. Multiple symbols at the order level indicate that members of the families listed within may express at least one or more of the listed attributes. (P_) refers to phylum, (C_) class, (O_) order, and (F_) family.

Table 1: List of family-level taxa chosen for this study from the Baby & Mi perspective birth cohort samples n=74, (between 5 months and 1 year) and their maximum relative abundance in any sample. (F_) refers to family, (C_) class, and (O_) order.

Bacterial Family	Max. relative abundance
F_Deferribacteraceae	3.10E-05
F_Microbacteriaceae	5.64E-04
F_Brevibacteriaceae	2.73E-05
F_Alcaligenaceae	5.18E-02
F_Peptococcaceae	7.86E-04
F_Verrucomicrobiaceae	2.56E-01
F_Desulfovibrionaceae	2.39E-03
F_Oxalobacteraceae	1.63E-04
F_Helicobacteraceae	8.40E-05
F_Caulobacteraceae	3.16E-04
F_Eubacteriaceae	3.99E-02
F_Dehalobacteriaceae	6.54E-05
F_Leuconostocaceae	4.15E-04
F_Sphingomonadaceae	1.09E-04
F_Aeromonadaceae	3.81E-04
F_Moraxellaceae	4.47E-04
F_Paenibacillaceae	3.47E-03
F_Planococcaceae	3.39E-05
F_Peptostreptococcaceae	1.23E-02
O_Streptophyta	1.81E-02
F_Fusobacteriaceae	4.01E-02
O_Campylobacterales	3.39E-03
F_Comamonadaceae	3.69E-02
C_Alphaproteobacteria	1.06E-02
F_Coriobacteriaceae	1.48E-01
F_Propionibacteriaceae	1.44E-03
F_Rikenellaceae	6.01E-02
F_Pseudomonadaceae	5.49E-03
F_Neisseriaceae	6.32E-04
F_Burkholderiaceae	8.71E-05
F_Turicibacteraceae	6.20E-03
F_Mycoplasmataceae	3.77E-02
F_Gemellaceae	5.14E-03

Bacterial Family	Max. relative abundance
F_Enterobacteriaceae	8.02E-01
F_Clostridiaceae	6.87E-01
F_Lactobacillaceae	7.82E-02
F_Enterococcaceae	4.81E-01
F_Streptococcaceae	1.23E-02
F_Porphyromonadaceae	3.07E-01
F_Bacteroidaceae	7.01E-01
F_Bifidobacteriaceae	9.47E-01
F_Actinomycetaceae	2.39E-02
F_Pasteurellaceae	2.01E-02
F_Erysipelotrichaceae	4.00E-01
F_Carnobacteriaceae	6.02E-03
F_Prevotellaceae	9.87E-03
F_Ruminococcacea	4.55E-01
F_Corynebacteriaceae	1.32E-03
F_Staphylococcaceae	2.36E-02

Aim 2: Explore the impact of solid food introduction on the infant gut microbiome using molecular and culture methods

Rationale: Solid food introduction has a sustained impact on the infant gut microbiome, shifting its community composition to resemble that of an adult's (Koenig *et al.,* 2011). This is attributed to the fact that breast milk contains substrates such as HMOs that can be utilized by a small subset of microorganisms (e.g. Bifidobacterium and Lactobacillus) (Bergstrom *et al.,* 2014; Koenig *et al.,* 2011). Conversely, foods supplement the infant's diet with complex substrates that promote characteristics typically found in an adult gut microbiome (e.g. increase in carbohydrate metabolism genes, fecal SCFAs, vitamin biosynthesis, and prevalence of the Bacteroidetes phylum) (Koenig *et al.,* 2011). The purpose of this aim is to apply the previously-designed targeted culture

strategy to a set of samples from one vaginally-born and exclusively breastfed infant, in order to investigate the impact of solid food introduction on overall microbial diversity and composition. Additionally, we will investigate the effect of solid food introduction on OTU dynamics in the infant gut.

Aim 3: Investigate strain diversity and fiber breakdown phenotypes of isolates from the infant gut during the introduction of solid food.

Rationale: Using the tools developed in Aim 1 and the strains and data collected in Aims 1 and 2 I wanted to explore the species and strain diversity and fibre breakdown phenotypes in two important bacterial genera in the infant gut: *Bifidobacterium* and *Bacteroides*. OTUs identified as part of the genus *Bifidobacterium* were abundant in the infant gut before 5 months of age in our study (Stearns *et al.,* 2017) and there were many *Bifidobacterium* OTUs with different patterns of change in relative abundance as solid food was introduced in the longitudinal study (Aim 2). *Bacteroides spp* abundance increases following solid food introduction (Koenig *et al.,* 2010; Bergstrom et al., 2014) and species in this genus has a flexible nutritional profile (Wexler, 2007).

2.0 Materials and methods

2.1 Choosing microbial culture conditions to test

To target the microbial taxa listed in **Table 1**, a supplementary figure presented in Lau et al. (2016) was used since it demonstrated family-level taxa abundances in 66 microbial culture conditions. Using this reference, I chose 14 aerobic and 28 anaerobic culture conditions (**Table 2**) to use in my first aim for culturing organisms of interest based on taxonomy. The 28 anaerobic culture conditions used in the first aim exclude BHI 7 medium, which was used exclusively in the second aim to culture *Bacteroides spp* (Eley *et al.*, 1984). A search of the literature revealed an additional medium used to target bifidobacteria (Beerens' medium (Beerens,1990)) and one to target species of *Baceroides* (BHI7 medium (Eley *et al.,* 1984)). Beerens' medium is composed of Columbia agar mixed with propionic acid. The low pH condition of the media promotes the growth of various *Bifidobacterium* species, such as *B. longum*, *B. adolescentis* and *B. breve* (Beerens, 1990). To target a wider range of *Bacteroides* species I tested the addition of menadione, yeast extract, and haemin to BHI (BHI7) (**Table 2**). The combination of those ingredients under anaerobic conditions has been shown to promote the growth of *Bacteroides fragilis* and *B. Destasonis* (Eley *et al.,* 1984).

Table 2: Microbial culture media used in this study.

Aerobic Media[c]	Anaerobic Media[c]	Media Type	Ingredients	Favors growth of
M9 Minimal Mucin Media(M9 muc)	M9 Minimal Mucin Media(M9 muc)	Minimal. Enrichment	Various salts, Water, Mucin	Mucin-metabolizing bacteria
Bacteroides Bile Esculin Agar (BBE)	***Bacteroides* Bile Esculin Agar (BBE)**	Selective, Differential	Bile, Hemin, Gentamicin, Esculin, Ferric ammonium citrate	*Bacteroides* Fragilis Group
Columbia Agar with Colistin, Nalidixic Acid and Blood (CNA)	**Columbia Agar with Colistin, Nalidixic Acid and Blood (CNA)**	Selective, Differential	Colistin, Nalidixic acid, Sheep blood	Gram positive Cocci
Brain Heart Infusion + Supplement set A[a] and B[b] (BHI 4)		Selective	Peptone and Dextrose components, Vitamin K, Hemin, Cysteine, Colistin, Nalidixic Acid	Gram positive bacteria
M9 Minimal Starch Medium(M9 star)	**M9 Minimal Starch Medium (M9 star)**	Minimal. Enrichment	Various salts, Water, Starch	Starch-metabolizing bacteria
Brain Heart Infusion + cellulose (BHI cell)	**Brain Heart Infusion + cellulose (BHI cell)**	Non-selective, Enrichment	Peptone and Dextrose components, Cellulose	Mainly cellulose metabolizing organisms, but also other bacteria
Columbia Blood Agar (CBA)	**Columbia Blood Agar (CBA)**	Non-selective, Enrichment	Vitamin K, Hemin, Sheep blood	Fastidious organisms
Bifidus Selective Medium Agar (BSM)	**Bifidus Selective Medium Agar (BSM)**	Selective	Peptone, Dextrose, Lactulose, Propionic acid	*Bifidobacterium longus, infantis*

de Man Rogosa Sharpe Agar (MRS) Agar	**de Man Rogosa Sharpe Agar (MRS) Agar**	Selective	Citrate, Acetate, Manganese, Magnesium	Fastidious *Lactobacillus spp.*
Kanamycin Vancomycin Laked Blood (KVLB)	**Kanamycin Vancomycin Laked Blood (KVLB)**	Selective	Laked blood, Vitamin K, Hemin, Kanamycin, Vancomycin	Gram negative, Anaerobic Bacilli (*Prevotella spp, Fusobacteria spp, Bacteroides spp*)
MacConkey Agar (MAC)	MacConkey Agar (MAC)	Selective, Differential	Peptones, Crystal violet, Bile salts, Neutral red indicator, Lactose	Gram negative and Enteric Bacilli (*Escherichia, Psuedomonas, Klebsiella*)
Chocolate Agar (CHOC)	Chocolate Agar (CHOC)	Non-selective, Enrichment	Protease peptone, Dextrose, Bovine Hemoglobin, NAD (V-Factor)	Fastidious pathogenic respiratory bacteria *(Haemophilus, Neisseria Meningitisis, Strep Pnuemo)*
Brain Heart Infusion + Supplement set B[b] (BHI 2)	**Brain Heart Infusion + Supplement set B[b] (BHI 2)**	Selective	Peptone and Dextrose components, Colistin, Nalidixic Acid	Gram positive bacteria
McKay Agar	**McKay Agar**	Selective, Differential	Sulfadizine, Hemin, Colistin, Oxolinic acid	Streptococcus Milleri Group (SMG)
	Mannitol Salt Agar (MSA)	Selective, Differential	7.5% NaCl, Beef extract, Peptones, Mannitol	*Staphylococci*
	Sterile Beerens' Agar (Beer st.)	Selective	Dextrose, Peptonized milk, Potassium	Lactic acid-producing bacteria (Lactobacillus,

			acetate, L-malic acid, Ethanol	Acetobacter, pediococcus)
	Non-Sterile Beerens' Agar (Beer n.s)[e]	Selective	Dextrose, Peptonized milk, Potassium acetate, L-malic acid, Ethanol	Lactic acid-producing bacteria (Lactobacillus, Acetobacter, pediococcus)
	Fastidious Anaerobe Agar (FAA)	Non-selective, Enrichment	Peptones, Vitamin K, Hemin, Glucose	Anaerobic organisms
	Cooked Meat Agar (Beef)	Non-selective, Enrichment	Peptones, Dextrose, Hemin, Vitamin K, Cooked beef heart	Anaerobic organisms, *Lachnospiraceae*
	Acintomycete Isolation Agar (AIA)	Non-selective, Enrichment	Nitrogen source, Glycerol, Sulphates	*Actinomycetes*
	Brain Heart Infusion Agar (BHI I)	Non-selective	Peptone and Dextrose components	Wide range of organisms
	Brain Heart Infusion + Supplement set A[a] (BHI 3)	Non-selective, Enrichment	Peptone and Dextrose components, Vitamin K, Hemin, Cysteine	Anaerobic organisms
	Brain Heart Infusion + Supplement A[a] + 1% Propionic acid (BHI 6)	Selective	Peptone and Dextrose components, Propionic acid	Propionic acid-tolerant bacteria
	Brain Heart Infusion + inulin (BHI inu)	Non-selective, Enrichment	Peptone and Dextrose components, Inulin	Mainly Inulin metabolizing organisms
	Brain Heart Infusion+ pectin (BHI Pec)	Non-selective, Enrichment	Peptone and Dextrose	Mainly Pectin metabolizing organisms

			components, Pectin	
	Brain Heart Infusion+ starch (BHI star)	Non-selective, Enrichment	Peptone and Dextrose components, Starch	Mainly Starch metabolizing organisms
	Brain Heart Infusion+ mucin (BHI muc)	Non-selective, Enrichment	Peptone and Dextrose components, Mucin	Mainly Mucin metabolizing organisms
	M9 Minimal Pectin Medium(M9 Pec)	Minimal. Enrichment	Various salts, Water, Pectin	Pectin-metabolizing bacteria
	Phenylethyl Alcohol Agar (PEA)	Selective	Peptones, Sheep blood, Phenylethyl alcohol	Gram positive, Obligate Anaerobes
	Brain Heart infusion 7 (BHI 7)[d]	Non-Selective Enrichment	Menadione, yeast extract, haemin	*Bacteroides fragilis* and *Destasonis*

[a]***Supplement Set A*:** 0.5 g/L L-cysteine hydrochloride hydrate, 10 mg/L hemin and 1 mg/L Vitamin K
[b]***Supplement Set B*:** 10 mg/L Colistin Sulphate, 5 mg/L Naladixic acid
[c]The 24 chosen media conditions are highlighted in red.
[d]BHI 7 was not a part of the pilot study but was chosen amongst the 24 media conditions to attempt to culture certain types of *Bacteroides* (based on a paper by Eley *et al.,* 1984)
[e]Non-sterile Beerens' medium was boiled but not autoclaved as it is suspected that sterilization damages components of the medium (based on a paper by Beerens, 1990).

2.2 Culturing infant stool

2.2.1 Pilot study culture

One fresh stool sample was collected from each of the three participants from the Baby&Mi prospective cohort study (Simioni *et al.,* 2016) when the infants were between five and six months of age. The methodology for culture-enriched molecular profiling was adapted from Lau et al (2016). 100 mg of each stool sample was aliquoted for

molecular profiling and frozen at -20°C. Approximately 1g of each sample was dissolved in 10 mL of BHI broth with 10µl L-cysteine (500 mg/mL), and a 1 in 10 serial dilution was performed subsequently. For the first two samples, 100µl of the 10^{-5} and 10^{-7} dilutions were plated separately on 42 microbial media conditions (28 anaerobic media, 14 aerobic media; **Table 2**). Anaerobic media were incubated for 5 days in an anaerobic chamber (Bactron IV Anaerobe Chamber). The bactertial colonies on the surface of each plate were then scraped using approximately 3 mL of BHI + L-cysteine media and 300µl was pooled and stored at -20°C prior to DNA extraction. Aerobic media was incubated for approximately 48 hours and then pooled similarly to the anaerobic plates. For sample 3, 100µl of the 10^{-3}, 10^{-5}, and 10^{-7} dilutions were plated separately on the same media types used for the first two samples. Plates were subsequently incubated and pooled following the same procedure listed above.

2.2.2 Longitudinal study culture

A total of 6 stool samples were collected from a breastfed infant over the course of 25 days. Two samples were acquired before and four were obtained after solid food introduction (**Figure 2**). The first and last samples were processed prior to freezing for microbial culture and the rest were frozen at -20°C prior to DNA isolation. Fresh samples underwent targeted microbial culture using the 24 media conditions chosen during the pilot phase of the study (**Table 2**) as per section **2.2.1**. Following incubation, individual colonies from each media condition were picked with a wooden toothpick into 96-well plates containing 150µl Brain-Heart Infusion Broth + L-cysteine per well. The 96-well plates,

containing the picked isolates, were incubated overnight and then frozen at -80°C in 10% skim milk as a cryoprotectant.

All anaerobic plates from the first fresh sample (except AIA, KVLB, and MSA) and all aerobic and anaerobic plates from the second fresh sample were scraped using approximately 3 mL of BHI+ L-cysteine Media and 300µl were pooled into genomic preparation tubes.

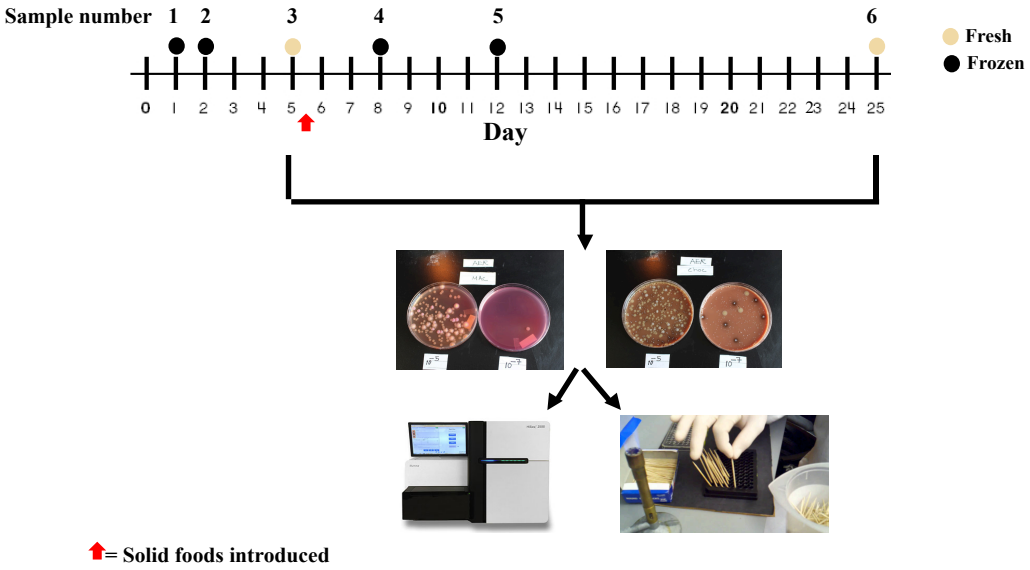


Figure 2: Timeline of the study: Solid circles represent fresh and frozen samples acquired from a breastfed, vaginally-born infant over a period of 25 days. Solid foods are introduced at day 5.

2.2.3 Obtaining bacterial isolates

Sequenced plate pool data, acquired from the pilot study samples (**Aim 1**), were used to determine the media conditions with the highest relative abundances of bifidobacteria. Based on the data acquired from 16S rRNA gene sequencing of the 24

plate pools (**Aim 2**), it was determined that anaerobic BSM was the plate pool with the highest relative abundance of bifidobacteria. 96-well plates with colonies obtained from this media condition were retrieved and isolates were recovered from frozen by transferring 3µl of frozen stock into the well of a 96 well plate containing 150µl BHI broth and L-cysteine.

Two fresh stool samples (one from before and one after solid food introduction) were plated on bifidogenic media Beeren's and Bifidus Selective Medium (BSM) as well as media to select for and differentiate the growth of *Bacteroides*: BHI7, *Bacteroides* Bile Esculin Agar (BBE) and Kanamycin Vancomycin-Laked Blood Agar (KVLB). Plates were grown anaerobically for 3-5 days (Baker-Ruskin anaerobic chamber, Concept) and approximately 5 representative colonies of the same size/color were individually picked into 96-well plates containing 150µl BHI broth and L-cysteine. The remaining agar plate contents were pooled in a genomic prep tube and sent for 16S rRNA gene sequencing of the V3 region (as per section **2.3**). The 96-well plates were grown anaerobically overnight and then frozen at -80C in 10% skim milk as a cryoprotectant.

2.3 DNA extraction and Illumina sequencing

DNA isolation from stool samples was conducted by M. Shah in the Surette laboratory, according to Lau *et al.,* 2016, as follows: 0.1g of stool or 500µl of pooled culture was aliquoted into genomic preparation tubes containing 800µl of 200 mM $NaPO_4$, pH 8 and 100µl guanidine thiocyanate-EDTA-*N*-lauroyl sarcosine. Samples were mechanically homogenized using 0.2g of 0.1mm and 2.8mm glass beads for pooled culture and stool, respectively (MoBio, Carlsbad, CA, USA). Enzymatic lysis was then

performed by incubating samples at 37 °C for 1 hour in 50 µl lysozyme (100 mg/ml), 50µl mutanolysin (10 U/µl), and 10µl RNase A (10 mg/ml). Samples are further incubated at 65 °C for 1 hour in 25µl 25 % sodium dodecyl sulfate (SDS), 25µl Proteinase K (20 mg/ml) and 75µl 5 M NaCl (Lau *et al.,* 2016). DNA was extracted by combining supernatants with phenol-chloroform-isoamyl alcohol (25:24:1; Sigma, St. Louis, MO, USA) and then purified according to the manufacturer's instructions by a DNA Clean and Concentrator-25 columns (Zymo, Irvine, CA, USA) using a vacuum manifold (EveryPrep Universal Vaccuum Manifold, Life Technologies #K2111-01). DNA concentration was quantified using a spectrophotometer (Nanodrop 2000c Spectrophotometers, Fisher Scientific, #ND- 2000C) and then subsequently stored at -20°C. The V3 region of the 16S rRNA gene was amplified through PCR amplification with a modified protocol and primers for Illumina high-throughput sequencing as previously described (Bartram *et al.,* 2011). Approximately 200ng of template was combined in a 50µl reaction mixture containing 1.5µl of 50mM $MgCl_2$, 200µM dNTPs, 5µl of 1µM V3F_mod2 barcoded primer, 5µl of 1µM V3R primer, and Taq polymerase (1.25 units/ 50µl PCR) (Invitrogen), amplified with the following conditions: 94C for 2min, followed by 30 cycles of 94°C for 30s, 50°C for 30s, and 72°C for 30s, with a final extension of 72°C for 10min. Amplified V3 16S rRNA gene products were sequenced at the McMaster Genomics Facility on a Miseq illumina sequencer (v2 2x250bp kit)

Raw sequence reads from illumina sequencing were trimmed with cutadapt (v1.14; cutoff of 30 and minlength of 100) (Martin, 2011) and then paired with PANDAseq (Masella *et al.,* 2012). OTUs were grouped at 97% similarity (Abundant OTU, Ye, 2011).

OTUs were then compared to sequences in Greengenes reference database (4 February 2011) and assigned specific taxonomies (Whelan and Surette, 2017). Number of media conditions required to represent the complete microbial profiles from each sample are determined using Custom R studio (V0.99.903) Scripts (**Appendix 8.3; A1, A2**) and Microsoft Excel (2015). OTUs with fewer than 10 read total across all samples were excluded.

2.4 Genus level identification: ID sequencing

To obtain a crude cell lysis for DNA extraction the 96-well plates were incubated anaerobically overnight and then stamped into worm lysis buffer with Proteinase K (Applied Biosystems) for DNA isolation. To make 50mL of worm lysis buffer, Milli-Q water was combined with 1M of 2.5mL KCl (Sigma-Aldrich), 1M of 0.5mL Tris-Base (Sigma-Aldrich), 1M of 125µl $MgCl_2$, 225µl of Igepal (Sigma-Aldrich), and 225µl of Tween-20 (Sigma-Aldrich). 50µl of worm lysis buffer and 2.5µl of Proteinase K (Applied Biosystems) were aliquoted per well. Colonies were incubated in worm lysis buffer for 60 minutes at 56°C then the suspension was boiled at 95°C for 15 minutes to inactivate the proteinase K. The cell lysates were then frozen at -80°C. 2µl of each cell lysate (Pilot study) was used as template for PCR amplification of the 16S rRNA gene V3 region for illumina sequencing as described in **section 2.3**. A subset of PCR products were separated by gel electrophoresis on an Ethidium Bromide (0.1mg/mL)- stained 1% agarose gel to verify that no non-specific products were amplified and that the product was the correct size of approximately 300bp. Approximately 250 individually barcoded amplified

products were then pooled and loaded as 1% of the final Illumina sequencing MiSeq run and sequenced as described above.

To determine the identity of each isolate Illumina sequencing files were combined from three separate runs and demultiplexed. Primer and adapter sequences were trimmed from the sequencing reads using cutadapt (version 1.14 using quality cutoff of 30 and minimum length of 100) (Martin, 2011) then single variants obtained with DADA2 (version 1.6.0 with filter and trim parameters: truncLen=c(130,130), maxN=0, maxEE=c(2,2), truncQ=11) (Callahan *et al.,* 2016). Taxonomy was classified against SILVAver132. Mean sequencing depth for ID sequencing was determined to be approximately 4,444 reads/sample, with minimum and maximum counts of 0 to 134,988 reads, respectively .The amplicon sequence variant table (***ASV***), containing approximately 827 isolates, was then used to identify isolates belonging to the bifidobacteria and *Bacteroides* genera.

2.5 Random amplification of polymorphic DNA (RAPD)

2.5.1 RAPD primer optimization

A comparison was preformed between AP17(5′-CAGGCCCTTC-3′) (IDT), AP21(5′-CACACGCACACGGAAGAA-3′), AP22(5′-ACGCACACGCACAG AGAG-3′) (Selenska *et al.,* 1995), PER1(5’-AAGAGCCCG T-3’), and CORR1(5’-TGCTCTGCCC-3’) (Jarocki *et al.,* 2016) RAPD primers on 8 isolates suspected to be of the bifidobacteria genus. All primers were obtained from IDT (California, USA). RAPD PCR was performed in a 25µl reaction volume containing: 2µl of 10µM RAPD primers, 2.5 µl of 10x Taq polymerase buffer (Invitrogen), 200µM dNTPs (Invitrogen), 0.75µl of

50µM $MgCl_2$ (Invitrogen), 1µl of 10mg/mL BSA (Sigma-Aldrich), Taq polymerase (1.25 units/ 50 µl PCR) (Invitrogen), and 300ng of template. PCR protocol was as follows: 95°C for 5 minutes, then 45 cycles of 94°C for 1min, 36°C for 1min, and 72°C for 2min, with a final extension at 72°C for 10 minutes. 5µl of PCR product was run on 2% agarose gel at 80mV for approximately 60 min for band separation. The agarose gel was stained with Ethidium Bromide (0.1mg/ml) (Thermofischer).

RAPD primer banding patterns for AP17, AP21, and AP22 were compared based on the number and clarity of bands across all four isolates. A primer that produces a higher number of bands is more discriminative than a primer that produces one or two bands (Miteva *et al.,* 2001). Additionally, primers were assessed based on the clarity of the bands they produce, with a brighter band being more favorable.

We determined the following visual criteria for scoring RAPD primer bands: a score of **0** indicates no bands are present. **1** indicates two bands or less are present. **2** indicates three bands or more are present, with none being bright. **3** indicates three bands or more are present, with at least two or more being bright. The primer with the highest sum of scores across all four isolates was chosen for RAPD PCR in Aim 3.

2.5.2 RAPD optimization of AP21

7 suspected isolates of *Bifidobacterium* and 7 suspected *Bacteroides* isolates (based on genus level assignments with ID sequencing) were then assessed for differential RAPD banding patterns using the AP21 primer (**Figure 3**). 3µl of the crude cell lysate was used as template in a final reaction volume of 25µl containing: 5µl of 5x standard

buffer with $MgCl_2$ (NEB), 0.5µl of 10µM dNTPs (Invitrogen), 1µl 10mg/ml BSA (Sigma-Aldrich), 0.125 µl, 2µl of 10µM AP21 primer (IDT), and 0.125µl OneTaq (NEB) per one reaction. PCR protocol was: 94°C for 5 minutes, followed by 45 cycles of 94°C for 1 min, 36°C for 1min, and 68°C for 2 min, a final extension was performed at 68°C for 10 minutes. PCR products were separated by 2% agarose gel electrophoresis at 60mV and 90 minutes, where 5µl of PCR product was loaded. The agarose gel was stained with Ethidium Bromide (0.1mg/ml) (Thermofischer) to allow for DNA visualization under ultraviolet light.

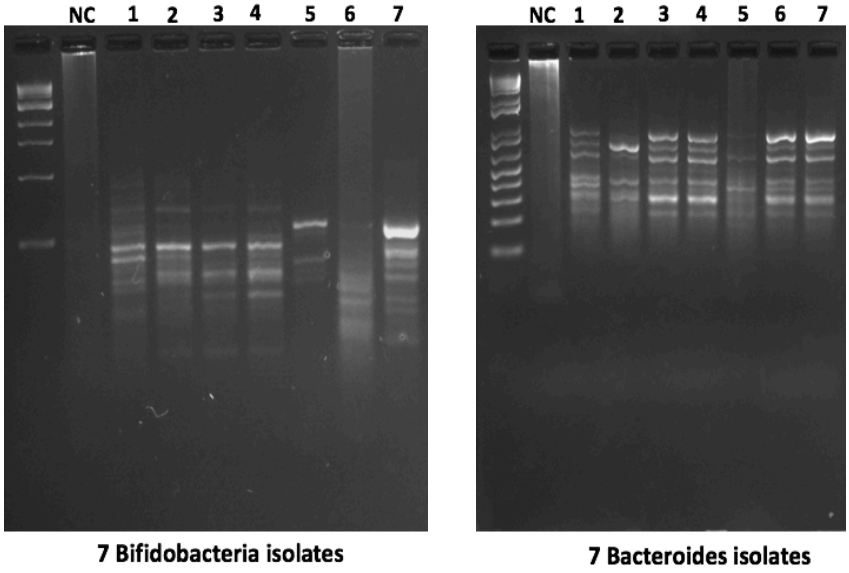


Figure 3: RAPD PCR Optimization preformed on seven bifidobacteria and seven *Bacteroides* isolates (identified by ID sequencing) using AP21 primer. NC refers to negative control. 1Kb ladder (Invitrogen) (left image) and 1Kb plus ladder (x) (Invitrogen) used for comparison.

2.5.3 RAPD scoring criteria

To compare banding patterns across all isolates of the same genus (96 bifidobacteria and 24 *Bacteroides* isolates) bands were aligned individually to the kb plus ladder (Invitrogen) and to one another. Bands on separate agarose gels were cropped and aligned to one another, following the alignment of their ladders. Using visual assessment, isolates with a similar banding pattern (e.g. same number of bands, same location relative to the ladder) were assigned a letter corresponding to that pattern. To account for observer bias, the bands were scored by two individuals on separate occasions. Subsequently, one isolate was chosen from each letter category and media condition to undergo full-length 16S rRNA gene sequencing (**Figure 20**, **21**)

2.6 Species-level assignment: full length 16S rRNA gene sequencing

48 isolates with different banding patterns were chosen as representatives and amplified with primers for the full length 16S rRNA gene. 120 isolates were retrieved from the -80°C freezer and re-streaked on BHI agar and incubated for 48 hours. Colonies were then picked into worm lysis buffer and proteinase K, and a DNA extraction was performed as per **section 2.4**. Colonies from each isolate were then frozen at -80C in 10% skim milk as a cryoprotectant. 16S rRNA gene was amplified with the 27F and 1492R primers (IDT, California, USA) (Frank *et al.,* 2008). 1µl of DNA from crude cell lysates was used as a template and combined in a 25µl reaction mixture containing 2.5µl 10x PCR buffer, 0.75µl 50mM $MgCl_2$, 200µM dNTPs, 1µl 10uM 27F primer, 1µl 10uM 1492R primer, and Taq polymerase (1.25 units/ 50 µl PCR) (Invitrogen), amplified with

the following conditions: 94C for 5min, followed by 30 cycles of 94°C for 1 min, 56°C for 1 min, and 72°C for 2 min, with a final extension of 72°C for 10min.

For 16S rRNA gene sequencing preformed on the 48 isolates obtained in aim 3, 3µl of template was combined in a 25µl reaction mixture containing 5µl 5x PCR buffer (with $MgCl_2$), 200µM dNTPs, 1µl 10uM 27F primer, 1µl 10uM 1492R primer, and Taq polymerase (1.25 units/ 50 µl PCR) (Invitrogen), amplified with the following conditions: 94C for 5min, followed by 30 cycles of 94°C for 1 min, 56°C for 1 min, and 68°C for 2 min, with a final extension of 68°C for 10min.

PCR products (1465bp) were separated by gel electrophoresis to ensure specificity. Products from 48 bacterial isolates were sent for Sanger sequencing in the forward direction alone in the McMaster Genomics Facility (Mobix Laboratory). I obtained approximately 700 bp of good quality sequence for each isolate that allowed me to identify each to the species-level. 20µl of each PCR product was then purified using the Monarch PCR clean-up kit (NEB, Massachusetts, USA) according to the manufacturer's instructions. The DNA was then eluted using 10µl PCR water, concentration was quantified using a spectrophotometer (Nanodrop 2000c Spectrophotometers, Fisher Scientific, #ND- 2000C), and then sent for Sanger sequencing (Mobix lab, Farncombe Institute, McMaster University). Genera and species of isolates were then identified by BLASTn alignment against NCBI's nucleotide collection (August, 2018).

2.7 Functional substrate assays

120 isolates were screened for fiber breakdown capacity. Bacterial stocks (**Section 2.6**) were recovered from frozen in BHI broth and grown anaerobically for two days. 3µl of this broth was stamped onto functional substrate assays for inulin ($C_{6n}H_{10n+2}O_{5n+1}$), starch ($C_6H_{10}O_5$), pectin ($C_6H_{10}O_7$), and cellulose ($(C_6H_{10}O_5)_n$) (all fibres acquired from Sigma, Missouri, USA) (**Figure 4, 5**) as follows: to make ½ BHI + 4g/L starch plates, half the amount of BHI broth required for a specific volume of media was used, as per manufacturer's instructions (BD). This was combined with 15g/L of agar (Thermo-Fisher), 4g/L Starch (Sigma), and dissolved in deionized water. The mixture was then autoclaved for 15-30 minutes, as per manufacturer's instructions. To visualize starch hydrolysis, agar plates were stained with Gram's iodine (2g KI, 1g I_2, and 300mL dH_2O), then excess iodine was dispensed in a beaker, and pictures were taken immediately to avoid halo fading. To make ½ BHI + 1.5g/L inulin plates, half the amount of BHI broth required for a specific volume of media was used, as per manufacturer's instructions (BD). This was combined with 15g/L of agar (Thermo-Fisher), 1.5g/L inulin (Sigma), 1.2mL Bromocresol Purple solution (0.16g Bromocresol purple (Sigma) in 10 mL 95% Ethanol), and dissolved in deionized water. The pH of the media was then adjusted to 7.0 ± 0.2 using HCl or NaOH and autoclaved for 15-30 minutes, as per manufacturer's instructions. To make ½ BHI + 10g/L pectin plates, follow the same procedure above for inulin (Sigma) using 10g/L pectin and 3.6mL of Bromocresol Purple solution. To make ½ BHI + 15g/L inulin plates. To make ½ BHI + 10g/L cellulose plates, follow the same procedure for previous media using 10g/L cellulose (Sigma) and 0.2g/L Congo Red

(Sigma). The pH of the media was then adjusted to 7.0 ± 0.2 using HCl or NaOH and autoclaved for 15-30 minutes, as per manufacturer's instructions. Since cellulose does not dissolve in liquid media, the solution was mixed frequently during pouring.

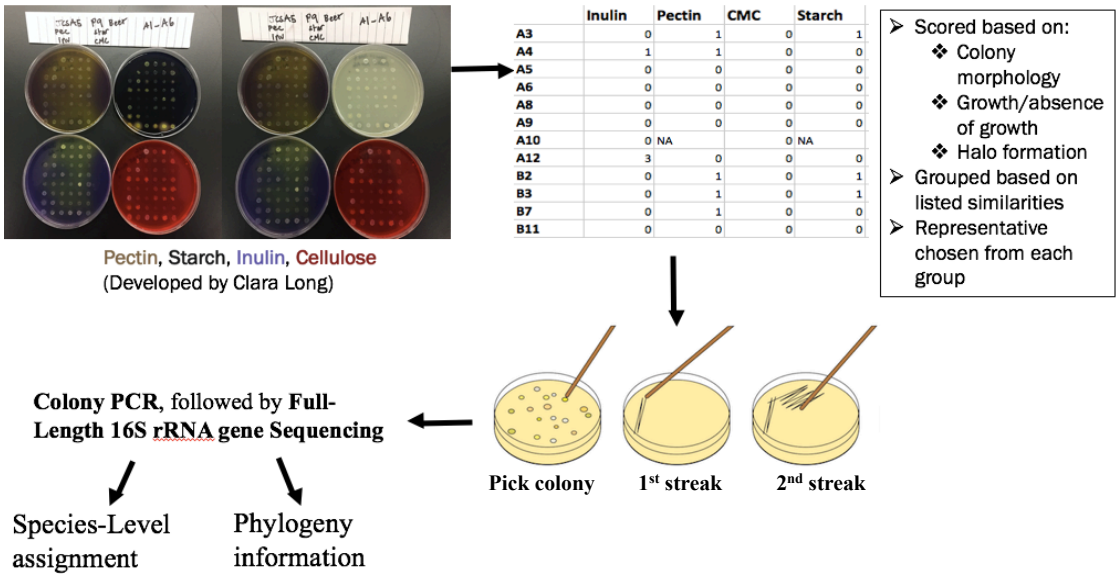


Figure 4: Functional substrate assays of pectin, starch, inulin and carboxymethyl cellulose used to derive a set of representative bifidobacteria and *Bacteroides spp* isolates

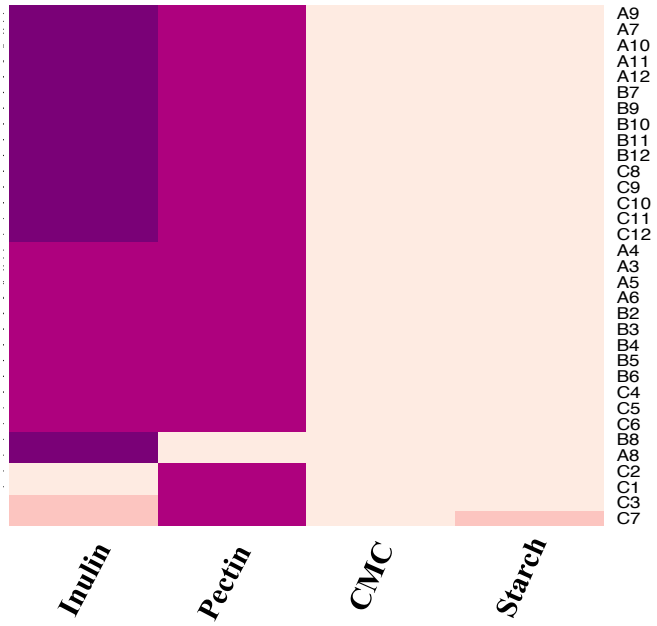


Figure 5: A subset of isolates obtained from plating fresh longitudinal samples on Beerens' Media. Isolates were stamped on Inulin, Pectin, CMC, and Starch plates. Isolates were allotted a score of 0-4 based on the degree of halo formation/discoloration around the colony (indicative of substrate utilization), where 0= no halo and 4= large, visible halo. Dark purple= 4, light purple=3, Dark beige= 1, light beige= 0.

2.8 Data analysis

All Illumina V3 rRNA gene sequencing data were run using the Sl1P pipeline (Whelan & Surette, 2017) prior to analysis, with the exception of ID sequencing data, which were run exclusively through DADA2 (Callahan *et al.,* 2016).

2.8.1 Determination of a representative subset of media to capture taxa of interest

R Code (**Appendix 8.3; A1**) was written to determine the combined number of OTUs across all stool and plate pool samples. Subsequently, we determined the number of overlapping OTUs between each stool sample and its respective plate pool. The data was used to construct a representative Venn diagram (**Figure 7**) (R packages: **Grid**,

VennDiagram). The number of OTUs shared across all samples was summarized for each media condition. Taxonomy assignments and OTU numbers were then used to determine which microbes at the genus level were: (a) present in stool and in the plate pools collectively; (b) overlapped between stool and plate pools; (c) were found exclusively in the plate pools; (d) were found exclusively in stool (not targeted by plate pools). All organisms of interest that were found in stool and in the plate pools were classified as abundant if existing at reads higher than 10 reads/OTU. Organisms of interest captured at lower than 10 reads/OTU were classified as rare. Otherwise, organisms were classified as absent (0 reads)

Relative frequencies of microbes in stool and in each culture condition were used to construct a taxonomical bar chart for visual comparison (**Figure 9**). Additionally, we constructed figures to compare the efficacy of all media conditions at capturing each individual family or genus of microbes found in our samples (**Figure 9**). The information provided by these figures was used to determine which conditions were most appropriate for culturing microbes of interest, both at the family and genus level (**Table 3** and **4**, respectively)

2.8.2 Longitudinal sample analysis

Microbial diversity within samples (α diversity) was calculated by the Smith method using the vegan package (Oksanen *et al.,* 2015) in R Studio. Microbial diversity between samples (β diversity) was estimated using the Bray-Curtis dissimilarity using the vegan package (Oksanen *et al.,* 2015) in R. Trends of individual microbial phylotypes

(OTUs) during the period of solid food introduction were visualized by plotting OTU abundance data over time in R (**Appendix 8.3, A2**).

2.8.3 Functional assay analysis

Fibre breakdown by bacterial isolates was scored visually based on the size of the halo and the amount of bacterial growth. Each colony was allotted a score between 0 and 3. A score of 0 indicates that no growth occurred on the medium. 1 indicates that the colony had a small diameter (approximately 1-3 mm) and minimum to no substrate utilization (little to no halo formation or discoloration). 2 indicates a larger colony diameter (approximately 4-5 mm) and moderate substrate utilization (smaller halo formation or slight discoloration around colony relative to colonies of the same size). A score of 3 also indicates that the colony ranges in size between 4-5 mm but is deemed to have a high substrate utilization (bigger halo formation or considerable discoloration around colony relative to colonies of the same size). To account for observer bias, the colonies were scored using the same criteria by two individuals on separate occasions.

3.0 Results

3.1 Success rate of culture for targeted organisms

As stated in **sections 2.2.1** of the methods, one fresh stool sample was obtained from each of the three infant participants enrolled in the Baby&Mi cohort study between the ages of 5 and 6 months. These three cross-sectional samples were then used to investigate the ability of the culture strategy to capture bacteria of interest. Stool samples were plated on 42 aerobic and anaerobic culture conditions that were designed to target 55 taxa of interest (**Table 1**). Both the plate pools and the stool samples underwent 16S rRNA gene sequencing of the V3 region (**section 2.3**). Data in **Figure 6** shows that majority of the taxa we predicted to be abundant were present in our samples. Furthermore, our targeted culture strategy appears to capture them more sensitively than 16S rRNA gene sequencing of stool alone. Only three taxa (Brevibacteriaceae, Microbacteriaceae, and Deferribacteriaceae) were not captured by targeted culture. Moreover, 18 taxa appear to be missing when using 16S rRNA gene sequencing of stool alone.

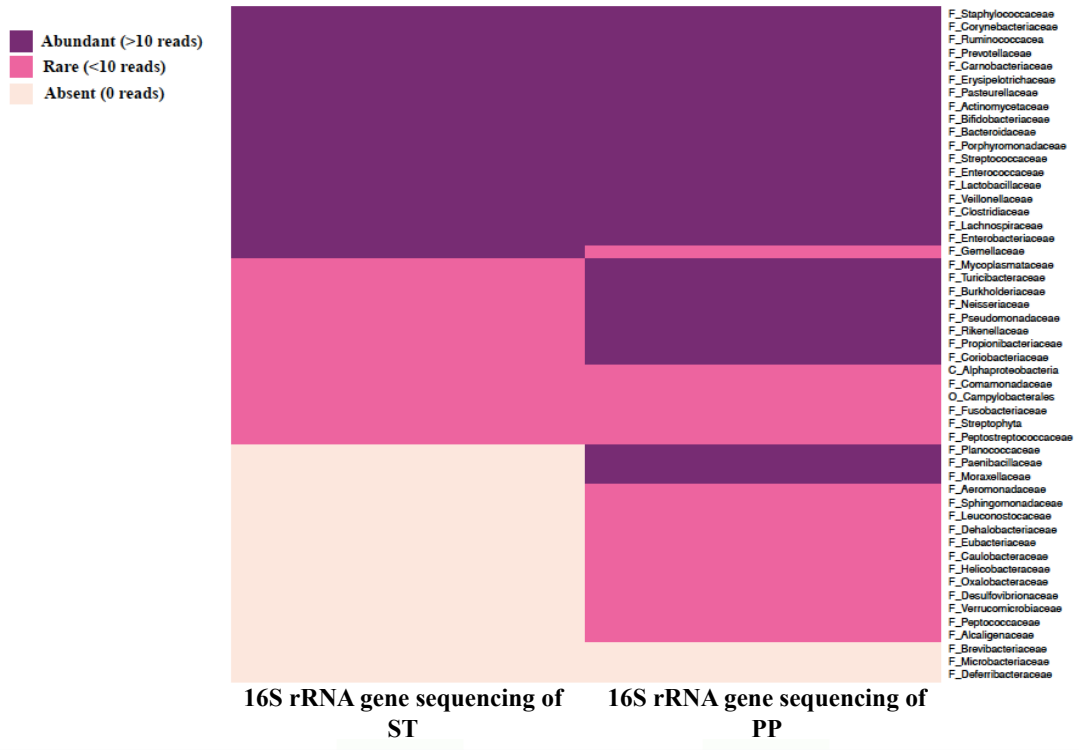


Figure 6: Heatmap comparing the abundance of taxa of interest (vertical axis) found using 16S rRNA gene sequencing data from plate pools (PP) of the first three pilot samples (right) and 16S rRNA gene sequencing of Stool (ST) (Left)

3.2 Conditions most appropriate for culture of microbial taxa of interest

According to data obtained from sequencing plate pools of the first three pilot samples, BHI2 and MRS appear to be the best conditions to capture the Bifidobacteriaceae family anaerobically (**Table 3**). Moreover, the genus bifidobacteria is targeted best by BSM and Beeren's media under anaerobic conditions, and is found to be captured aerobically using BHI4 medium (**Table 4**). The genus *Bacteroides* is captured anaerobically by a variety of media conditions including BHI inulin, pectin, starch, as well as CBA, PEA, and M9 pectin. Moreover, both KVLB and CNA appear to capture the highest abundance of *Bacteroides* under anaerobic conditions overall (**Table 4**). These observations can be also

made in **Figure 9** which shows the relative abundance of microbes, at both the family and generic levels, in all 42 media conditions and their respective stool samples.

In Table 3 and 4 taxa of interest differ in the type of media and conditions they grow on. Some taxa such as *Parabacteroides* primarily grow anaerobically on one type of media, such as BBE (**Table 4**). Others like Blautia grow under many conditions and are targeted by more than one type of media. Data indicates that Balutia can grow both aerobically on CNA and anaerobically on BSA medium (Refer to **Tables 3** and **4**).

Table 3: List of taxa chosen for this study, grouped at the family level (with unknown genera), and media conditions that best culture them across all three cross-sectional stool samples.

Family-level assignment of taxa of interest	**Media/Condition appropriate for culture**
f__Bifidobacteriaceae_g_unclassified	**BHI2**/ANA; **MRS**/ANA
f__Clostridiaceae_g_unclassified	**BHI Inulin**/ANA; **PEA**/ANA
f__Enterobacteriaceae_g_unclassified	**M9 Mucin**/AER
f__Erysipelotrichaceae_g_unclassified	**Beerens' ns**/ANA
f__Lachnospiraceae_g_unclassified	**McKay**/ANA; **M9 Starch**/AER/ANA
f__Ruminococcaceae_g_unclassified	**M9 Starch**/ANA; **CNA**/AER
f__Veillonellaceae_g_unclassified	**CBA**/ANA
f__Lactobacillaceae_g_unclassified	**BHI2**/ ANA; **Beef**/ANA

Table 4: List of taxa chosen for this study, grouped at the genus level, and media conditions that best culture them across all three cross-sectional stool samples.

Genus-level assignment of taxa of interest	**Media /Condition appropriate for culture**
g__bifidobacteria	**BHI4**/AER; **CNA**/AER/ANA; **Beer**/ANA; **BSM**/ANA
g__Akkermansia	**Beef**/ ANA
g__Bacteroides	**KVLB**/ANA; **CNA**/ANA
g__Blautia	**MSA**/ANA; **CNA**/AER
g__Clostridium (Clostridiaceae)	**BHI2**/ANA; **MRS**/ANA
g__Clostridium (Erysipelotrichaceae)	**BHI2**/ANA
g__Collinsella	**Choc**/AER
g__Enterococcus	**BHI2**/AER; MSA/ANA
g__Faecalibacterium	**CNA**/AER; **BHI cellulose**/ANA
g__Klebsiella	**MRS**/AER
g__Lachnospira	**BHI2**/AER

g__Lactobacillus	**MRS**/ANA/AER
g__Megasphaera	**CBA**/ANA
g__Oscillospira	**BHI inulin**/ANA
g__Parabacteroides	**BBE**/ANA
g__Prevotella	**CBA**/ANA
g__Ruminococcus (Lachnospiraceae)	**PEA**/ANA
g__Ruminococcus (Ruminococcaceae)	**M9 Starch**/AER/ANA
g__Streptococcus	**McKay**/ANA; **BHI4**/AER
g__Veillonella	**AIA**/ANA; **MRS**/ANA

3.3 Culture adds sensitivity to molecular methods

Mean sequencing depth for V3 16S rRNA gene sequencing was determined to be 128,434 reads/sample, with minimum and maximum counts of 12,536 and 223,064 reads, respectively. A total of 193 OTUs were found across all samples after filtering for >10 reads/OTU. 58 OTUs were obtained from sequencing stool while 187 were obtained from microbial culture. **Table 5** shows sequence depth of cross-sectional stool samples (reads) and mean sequence depth of each of their plate pools (reads/sample). Sequence depths of stool samples 1, 2, and 3 are comparable to the average sequence depths of their respective plate pools (**Table 5**). This is indicative that the higher OTU count found in plate pools relative to stool is not a result of a higher sequence depth. Therefore, we can reasonably assume that culture adds sensitivity to molecular methods.

Table 5: Sequence depth of cross-sectional stool samples (reads) and mean sequence depth of each of their respective plate pools (reads/sample)

Sample	Sequence depth of stool sample (reads)	Mean sequence depth of plate pools (reads/sample)
1	144,567	129,626
2	106,204	117,197
3	72,483	117,611.5

The mean overlap was calculated by averaging the overlap of the three samples to their respective plate pools (e.g. 100%, 89%, and 68% overlap between stool and microbial culture in sample 1, 2, and 3, respectively). Therefore, microbial culture captured 85.60% ± 9.52% of the diversity within participants' stool samples (**Figure 7, left**). Moreover, approximately 137 OTUs were unique to microbial culture. This confirms that culture enriched molecular profiling (Lau *et al.,* 2016) captures a higher diversity from the samples (193 OTUs) than sequencing stool alone (58 OTUs). We speculate that the 14.5% of stool OTUs that were missed by sequencing microbial culture (e.g. some OTUs of *Rothia*, *Haemophilus* and *Streptococcus*) may have consisted of dead cells (Lau *et al.,* 2016). This observation can be further supported by combining all samples, resulting in an overlap between plate pool and stool OTUs at 94.5% (**Figure 7, right)**, as opposed to at 85.6% (**Figure 7, left)**. This means that the same OTUs that were not picked up by any medium in one sample, were sufficiently targeted by media in another. These findings demonstrate that our microbial culture conditions are representative of the OTU diversity found in stool.

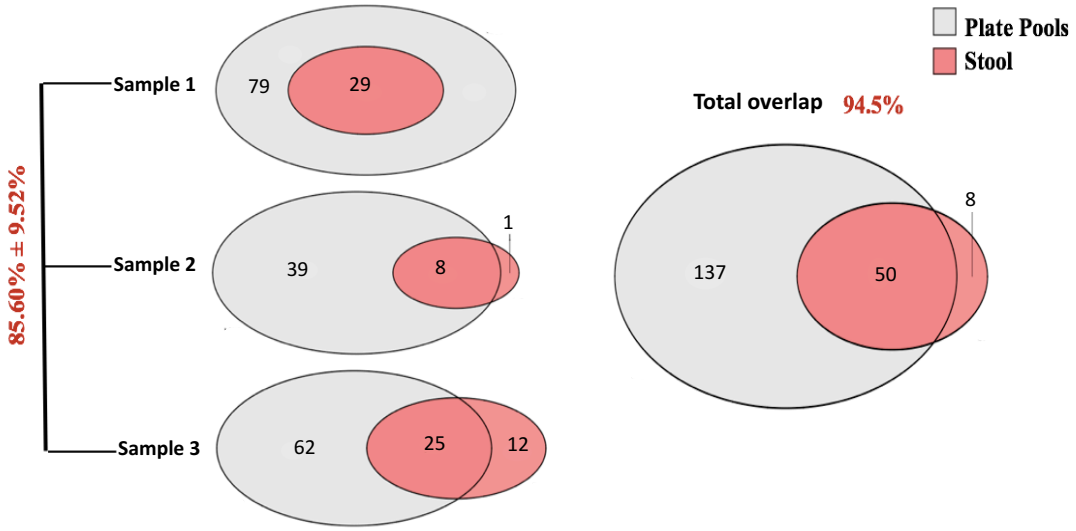


Figure 7: Degree of overlap between OTUs in Stool (St) and Microbial Culture (MC) in samples 1, 2, and 3 (Left). Degree of overlap between OTUs in all three stool samples combined and all pooled culture contents (right)

A complete picture portraying the microbial distribution and abundance of stool samples 1, 2, and 3 as well as the 42 pooled microbial culture conditions can be seen in **Figure 8**. Results show a high variability between the microbial profiles of sample 1, 2 and 3 (**Figure 8**). Identical media conditions also appear to significantly differ in their pooled microbial content based on the sample plated. Furthermore, the heat map depicted by **Figure 6** demonstrates that our media conditions can culture most of our target species efficiently. This is important for deciding which media conditions will be chosen for the second phase of the study since plates that are more adept at culturing taxa of interest will be preferred.

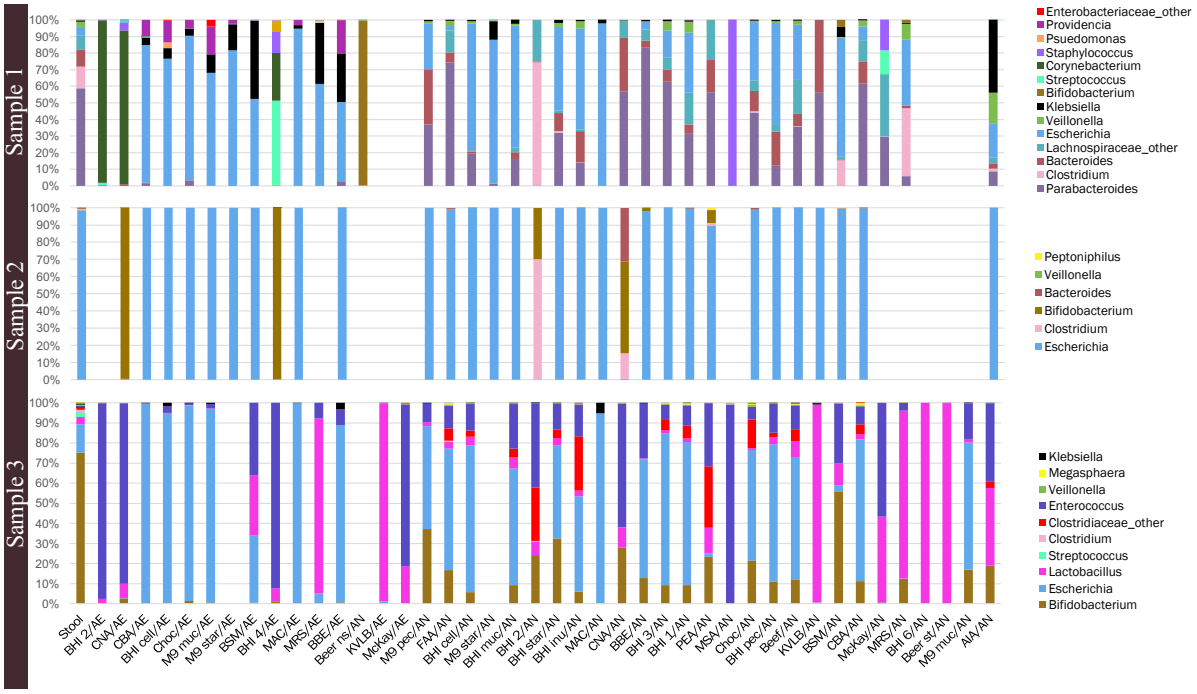


Figure 8: A taxonomic bar chart of microbial distribution and abundance in stool (column 1) and 42 aerobic (AE) and anaerobic (AN) plate pools (column 2 to 43) acquired from the first three pilot samples (sample 1, sample 2, and sample 3).

Figures showing the relative abundance of each individual taxa of interest in all 42 plate pools were constructed to determine which media conditions culture them best. **Figure 9** is an example of the relative abundance of two taxa of interest (bifidobacteria and *Bacteroides*) in all plate pools. BSM, CNA, and Beerens' ns media are best at capturing bifidobacteria anaerobically, while BHI 4 and CNA culture it best aerobically (**Figure 9a**). Moreover, *Bacteroides* is a strict anaerobe and preferably grows on anaerobic CNA and KVLB (**Figure 9b**)

The microbial distribution and abundance information obtained from **Figure 6, 8** and **9** was used to determine a smaller subset of culture conditions that capture most organisms of interest in both microbial culture and stool (**Table 2**).

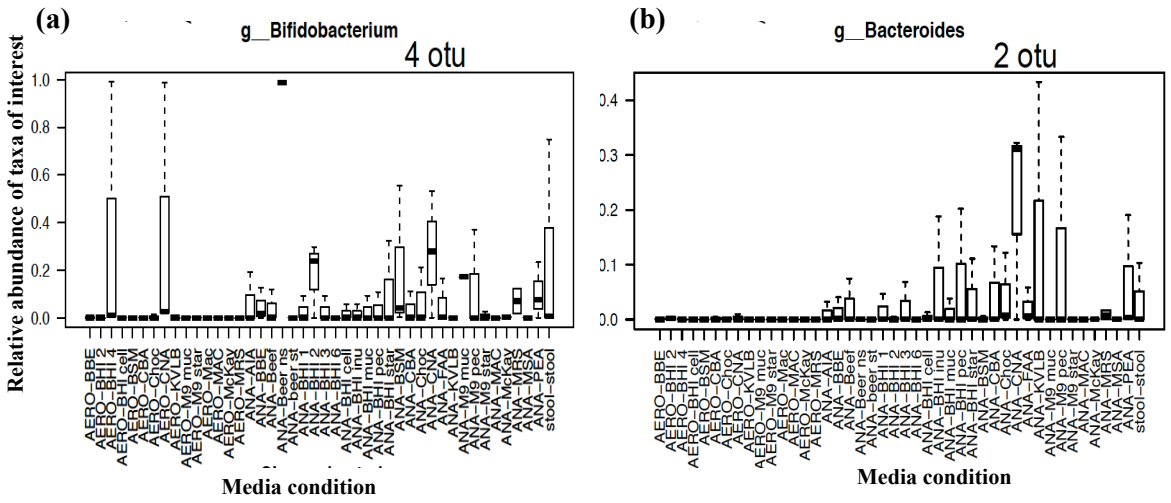


Figure 9: A boxplot showing the mean relative abundance across three pilot sample replicates of (a) bifidobacteria and (b) *Bacteroides* in each of the 42 media conditions and stool. AERO: media condition incubated aerobically; ANA: media condition incubated anaerobically. Figures such as this one were constructed individually for each taxa of interest to determine which media conditions grow them best.

24 plates (7 Aerobic, 17 Anaerobic) were chosen based on their overall coverage of both the OTUs found in stool and the taxa of interest found in both stool and plate pools

3.4 Infant gut community composition before and after solid food introduction based on molecular profiles

Six samples were acquired from a 6-month-old, breastfed, vaginally-born infant, with no history of antibiotic administration, over the course of 25 days, as described in **section 2.2.2** of the methods (**Figure 2**). All stool samples were sent for direct 16S rRNA gene sequencing of the V3 region. Two fresh stool samples (sample 3 and 6) were plated on the 24 media conditions chosen in **section 3.3**, pooled, and sent for sequencing.

Sequence depth of all stool samples and mean sequence depth of fresh samples' plate pools are listed in **Table 6**.

Table 6: Sequence depth of longitudinal stool samples (reads) and mean sequence depth of fresh stool samples' plate pools (reads/sample)

Sample	Day	Sequence depth of stool sample (reads)	Mean sequence depth of plate pools (reads/sample)
1	1	27,486	
2	2	57,274	
3	5	70,004	90,685.5
4	8	68.050	
5	12	64,869	
6	25	24,853	69,653.5

The microbial community in this infant was dominated by species of bifidobacteria before and after solid food introduction (**Figure 10** and **11**). Prior to solid food introduction (day 1), the gut microbial population is predominated by three *Bifidobacterium* phylotypes (OTUs), at a total relative abundance of approximately 85%. Moreover, by day 25 (post-solid food introduction) overall *Bifidobacterium* abundance decreases by approximately 27.5%, however, it remains the most abundant organism in this infant's gut at day 25 (**Figure 10** and **11**).

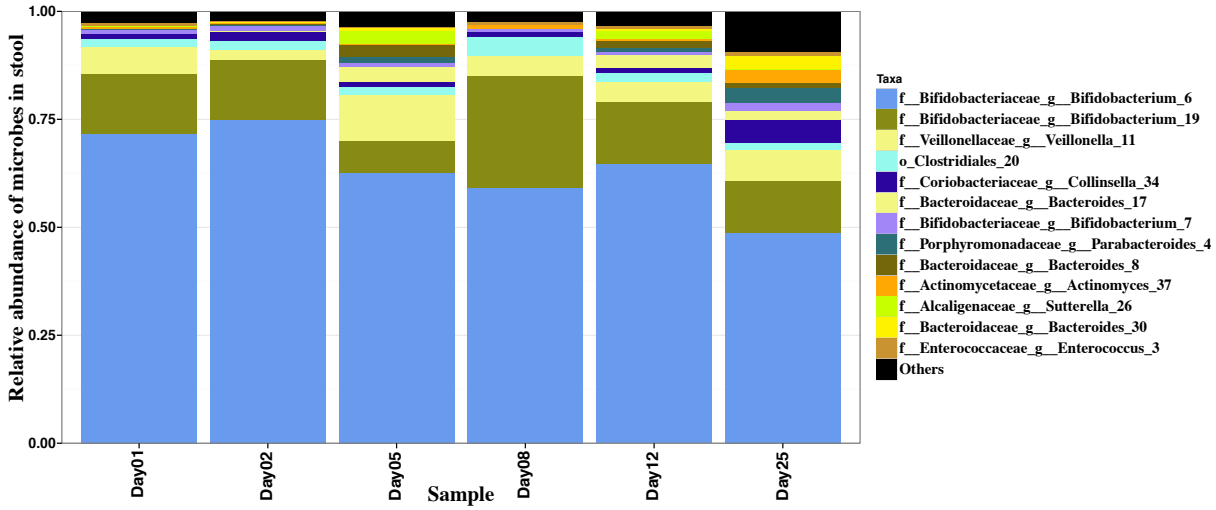


Figure 10: Relative abundance of microbes in stool from day 1 to 25. Food was introduced at day 5.

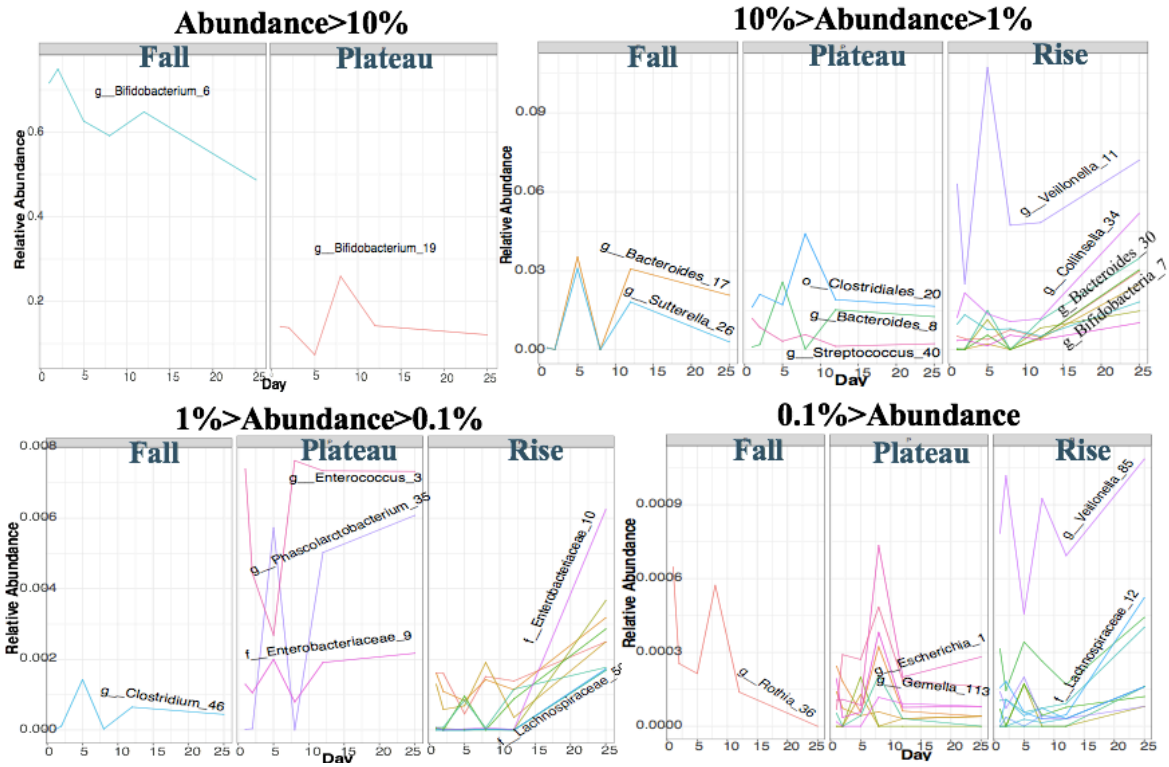


Figure 11: OTUs acquired from longitudinal samples were grouped based on:
1. *Abundance* (higher than 10%, between 10%-1%, between 1%-0.1%, or lower than 0.1%)
2. Whether they displayed an overall rising, plateauing, or falling *pattern* throughout the course of the study.

Following solid food introduction (day 25), there is an overall increase in the microbial diversity present compared to day 1 (before solid food introduction) (**Figure 10**). This is portrayed by an increase in alpha (α) diversity over time (**Figure 12**). Data acquired through 16S rRNA gene sequencing of stool show that an increase in both species richness and community diversity (Species richness and evenness measure) over time contribute to the observable expansion in diversity. As result, new organisms appear to emerge following solid food introduction (e.g. certain OTUs of *Escherichia*, *Clostridium*, *Lachnospiraceae*, *Leclercia*, *Bifidobacterium*, *Bacteroides*, and others). Furthermore, other pre-existing OTUs simply increase in abundance (e.g. *Bacteroides*, *Actinomyces*, *Veillonella*, and *Collinsella*) (**Figure 10** and **11**).

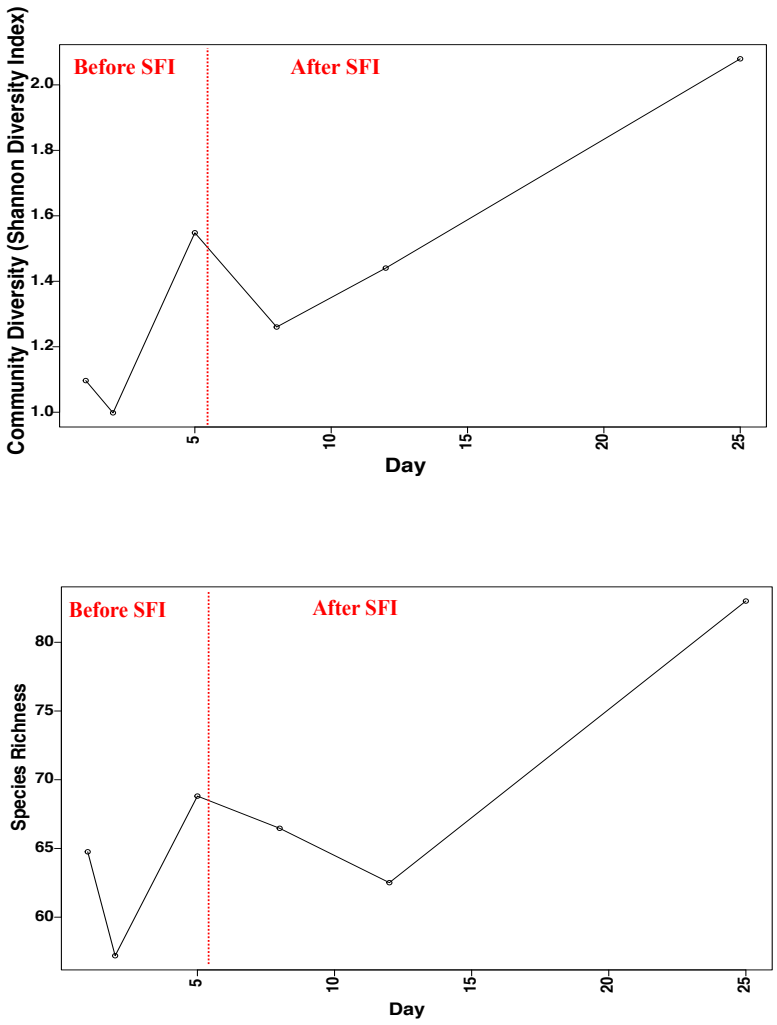


Figure 12: α diversity: Community diversity overtime depicted by the Shannon-Diversity Index (Top) and species richness overtime (Bottom). Red dotted line corresponds to time solid foods were introduced. SFI= Solid food introduction.

β diversity was measured between each of the six longitudinal samples from this infant and the three cross sectional samples used in section 2 for comparison. Inter-individual diversity in the microbiome was much higher than intra-individual diversity. The dissimilarity index between sample 1 (prior to solid food introduction) and sample 6

(following solid food introduction) is the highest compared to the index obtained between sample 1 or 6 and the other stool samples (**Figure 13**).

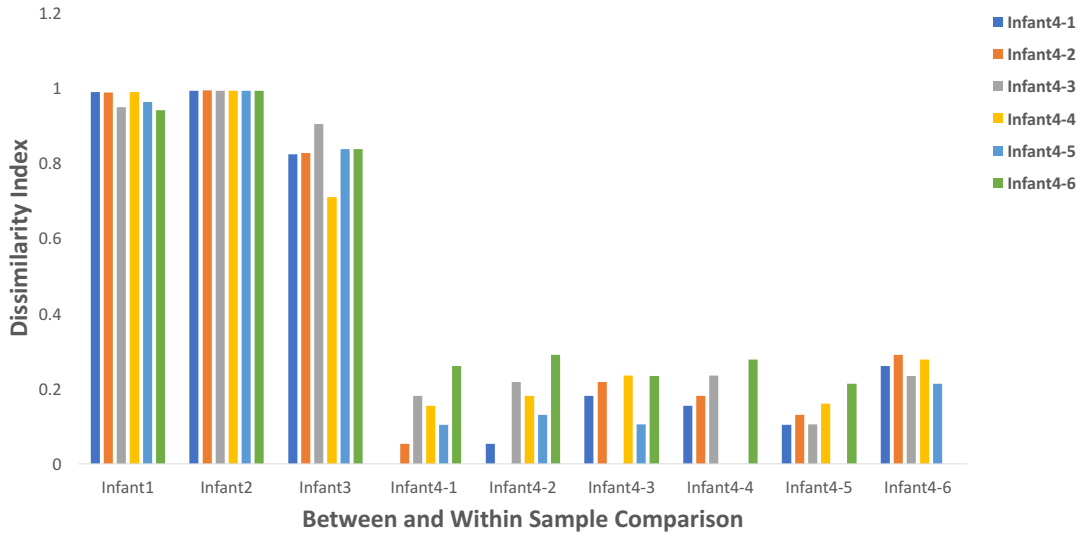


Figure 13: Beta diversity figure comparing inter and intra-individual variability in stool samples. Infant 1, 2, and 3 are cross-sectional stool samples acquired from 3 infants in the pilot study. Infant 4-1 to 4-6 are longitudinal stool samples acquired from one infant in the longitudinal study. The higher the dissimilarity index between samples the more microbial variability is present between them.

Individual OTU patterns appear to differ within the same genera over the course of the study. For example, while the *Bifidobacterium* genus displays a general decreasing trend following solid food introduction, phylotypes 7, 19, and 6 appear to respectively increase, stay the same, and decrease over the course of the study (**Figure 11**). Similarly, the *Bacteroides* genus shows an overall increasing trend, but phylotypes 17, 8, and 30 appear to respectively decrease, stay the same, and increase following the introduction of solid foods (**Figure 11**).

3.5 Microbial culture of the infant gut sample over time

Approximately 2,400 isolates were cultured from the pre-solid food introduction fresh sample, and 829 isolates from the post-solid food introduction fresh sample.

Data in **Figure 14** shows that majority of the taxa we predicted to be abundant were present in our samples. Furthermore, our targeted culture strategy appears to be more sensitive than 16S rRNA gene sequencing of stool alone at capturing them. Only eight taxa (Brevibacterceae, Microbacteriaceae, Paenibacillaceae, Planococcaceae, Sphingomonadaceae, Comamonadaceae, F_Aeromonadaceae, and Deferribacteriaceae) were not captured by targeted culture. Moreover, 27 taxa appear to be missing when using 16S rRNA gene sequencing of stool alone, including the eight taxa missing from culture sequences.

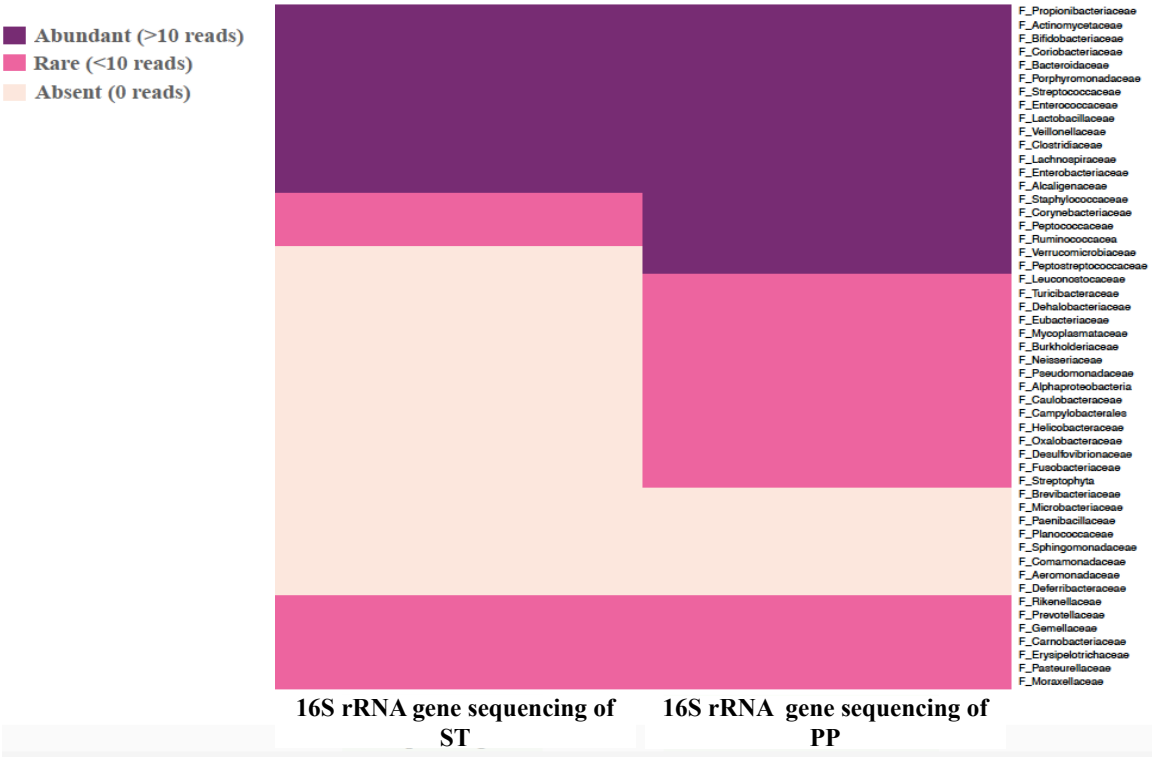


Figure 14: Heatmap comparing the abundance of taxa of interest (vertical axis) found using 16S rRNA gene sequencing data from plate pools of the two fresh longitudinal samples (right) and 16S rRNA gene sequencing of Stool (Left)

3.6 Combining molecular profiles and microbial culture reveals that species evenness increases after solid food introduction

Tracking the emergence and disappearance of OTUs over the course of this longitudinal study can help us understand how the microbial community responds following the transition to solid feeding. Based on the molecular profiles, 33 OTUs were absent then appeared and 58 OTUs were present then disappeared during solid food introduction (**Figures 15** and **16**). DNA isolated from microbial culture plates was profiled in the same way and provides an indication of the microbial groups that were cultured from the fresh sample. When the 16S rRNA gene data is compared between

direct sequencing of DNA from stool samples and DNA obtained from culture plates, we see that approximately 80% of the organisms that appeared to 'emerge' following solid food introduction were, in fact, captured by multiple media conditions from our targeted culture strategy (**Figure 15**). Similarly, approximately 50% of the organisms that appeared to "disappear" following solid food introduction were also captured by multiple culture conditions (**Figure 16**). Although the microbes were detected at extremely low abundances in the plate pools, making them not feasible to isolate, we suspect that their detection indicates that they were already present and viable in our samples at day 5 and 25 but were below the limit of detection of molecular profiling methods. Therefore, α diversity appears to be primarily driven in this case by an increase in species evenness rather than the arrival of new species (increased richness) to the infant gut environment.

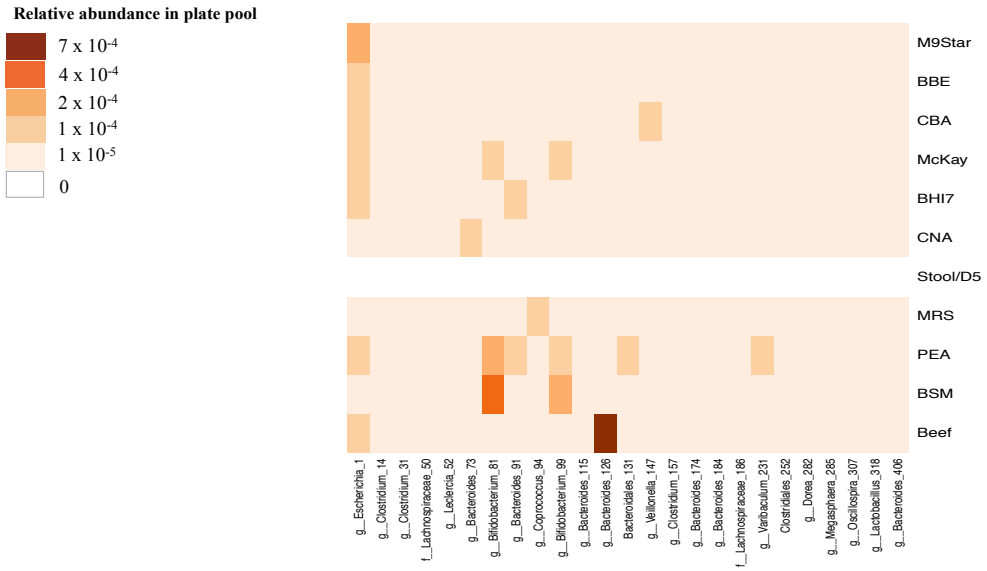


Figure 15: Microbes missing before solid food introduction (Stool/D5) but captured by targeted culture. D5 refers to fresh stool sample acquired on day 5.

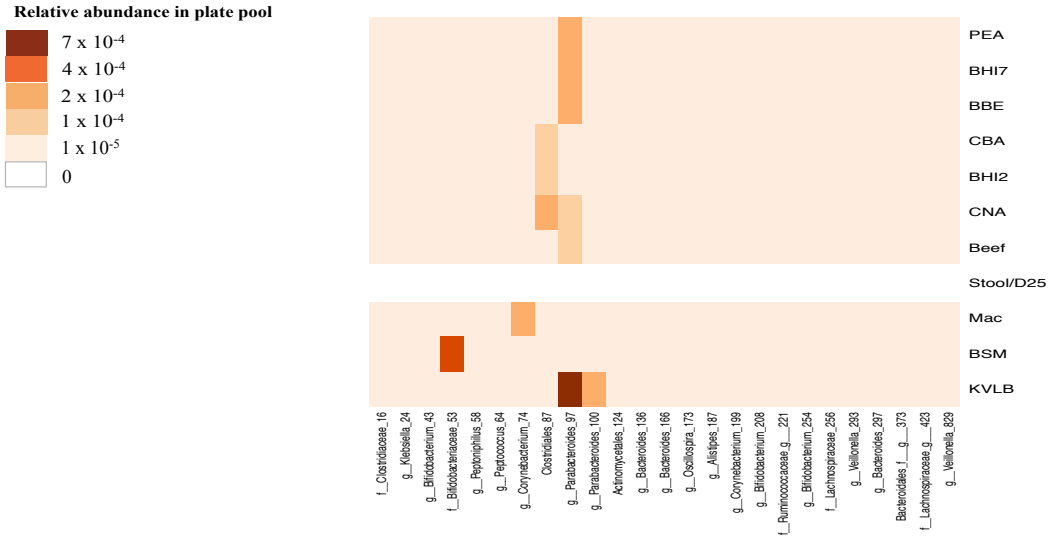
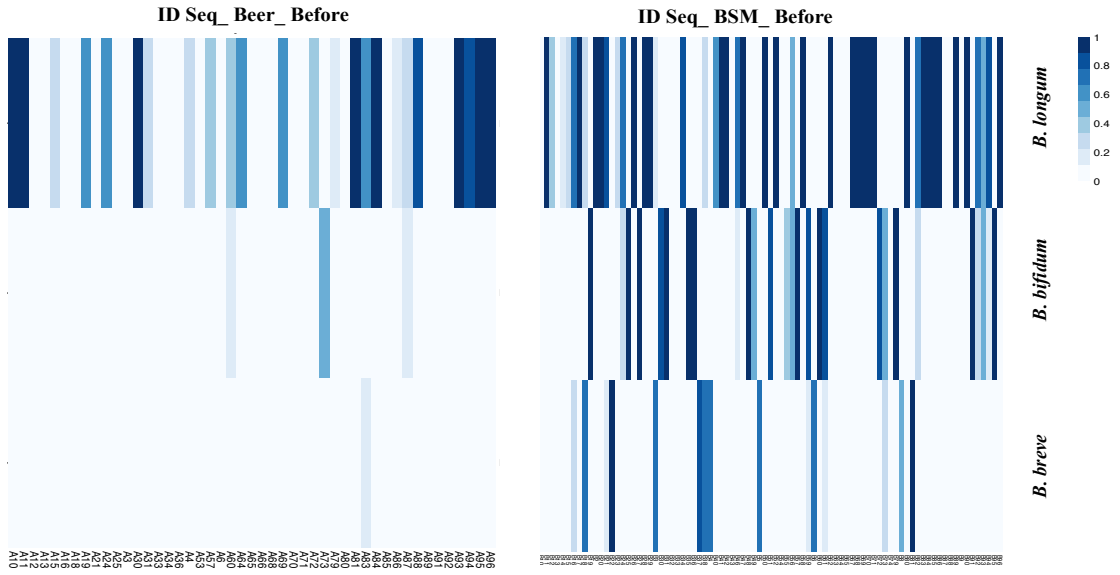


Figure 16: Microbes missing after solid food introduction (Stool/D25) but captured by targeted culture. D25 refers to fresh stool sample acquired on day 25.

3.7 Identity and strain diversity of bacterial isolates

A total of 827 isolates acquired from pre-and post-solid food introduction samples of BSM, BEER, PEA, BHI inu, CNA, CBA, and KVLB were sent for ID sequencing to determine which isolates likely belonged to either of the genera: *Bifidobacterium* and *Bacteroides* (**Figure 17** and **18**).

(a)



(b)

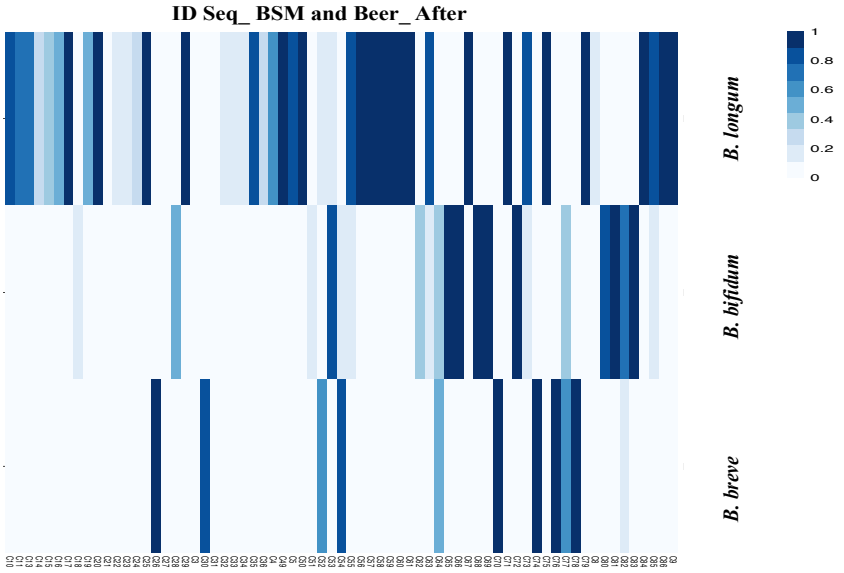
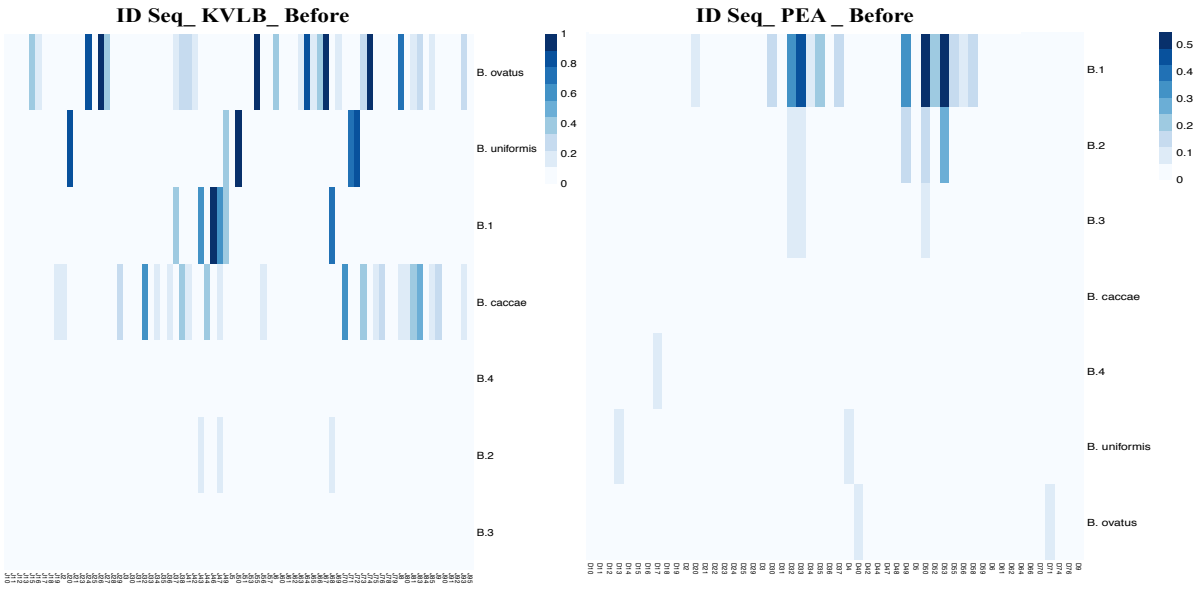
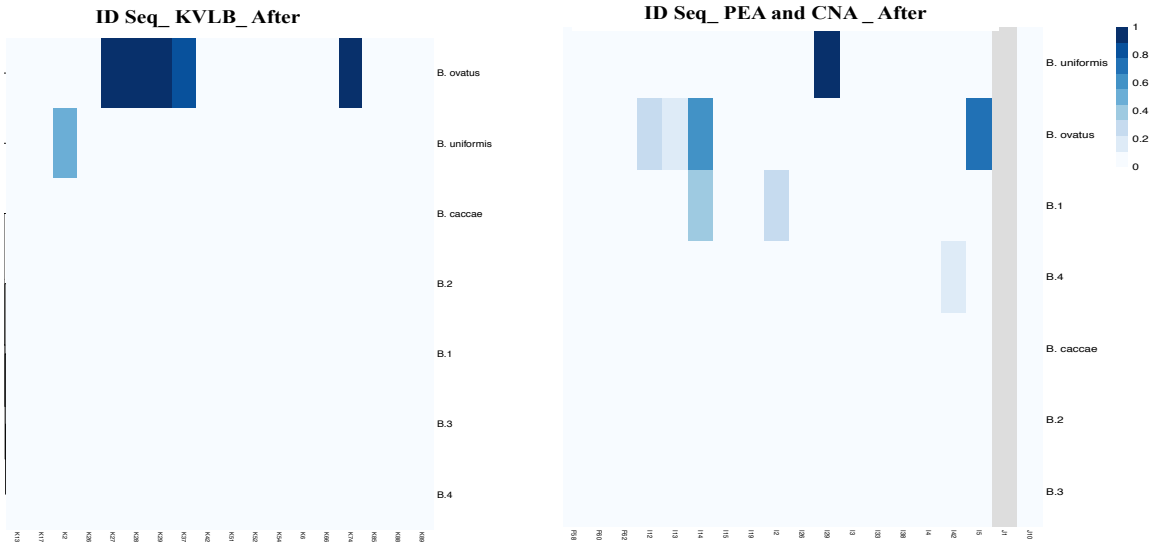


Figure 17: ID sequencing data portraying relative abundance of bifidobacteria acquired from media **(a)** before solid food introduction (BEER_Before, BSM_Before) and **(b)** after solid food introduction (BSM and BEER_After). Each column represents microbial content acquired from one well of a 96-well plate of either BSM or Beerens'. Rows represent species of bifidobacteria (or individual ASVs). Species assignments were estimated based on taxonomic data acquired from SILVAver132 database. Increasing grades of blue represent greater relative abundance, ranging from 0- 100%.

(a)



(b)



<u>Figure 18</u>: ID sequencing data portraying relative abundance of *Bacteroides* acquired from media **(a)** before solid food introduction (PEA_Before, KVLB_Before) and **(b)** after solid food introduction (PEA and CNA_After, KVLB_After). Each column represents microbial content acquired from one well of a 96-well plate of either KVLB, CAN, or PEA. Rows represent species of *Bacteroides* (or individual ASVs). Species were

estimated based on taxonomic data acquired from SILVAver132 database. Increasing grades of blue represent greater relative abundance, ranging from 0- 100%.

Based on ID sequencing results (ranging from 75% to 100% purity), 96 isolates suspected to be species of *Bifidobacterium* (BSM and BEER; **Figure 17**) and 24 isolates suspected to be species of *Bacteroides* (PEA, CNA and KVLB; **Figure 18**) were chosen to undergo RAPD using the AP21 primer. AP21 was chosen based on the scoring criteria in **section 2.5.1** of materials and methods. AP17, AP21, and AP22 received total scores of 7, 12, and 7, respectively (**Figure 19**). Given that a higher total score suggests a better discriminative ability (e.g. bright, more banding), AP21 was determined to be the primer of choice.

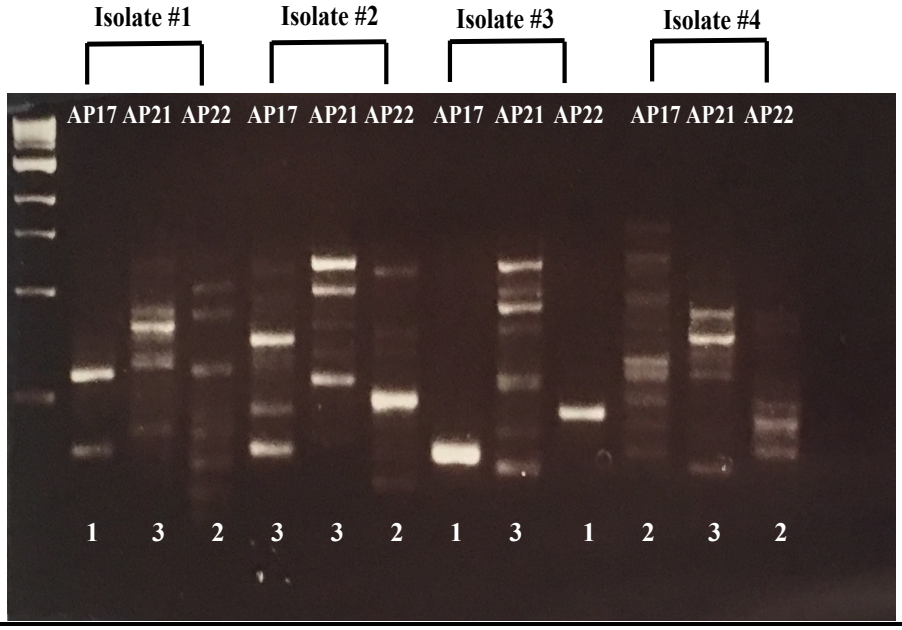
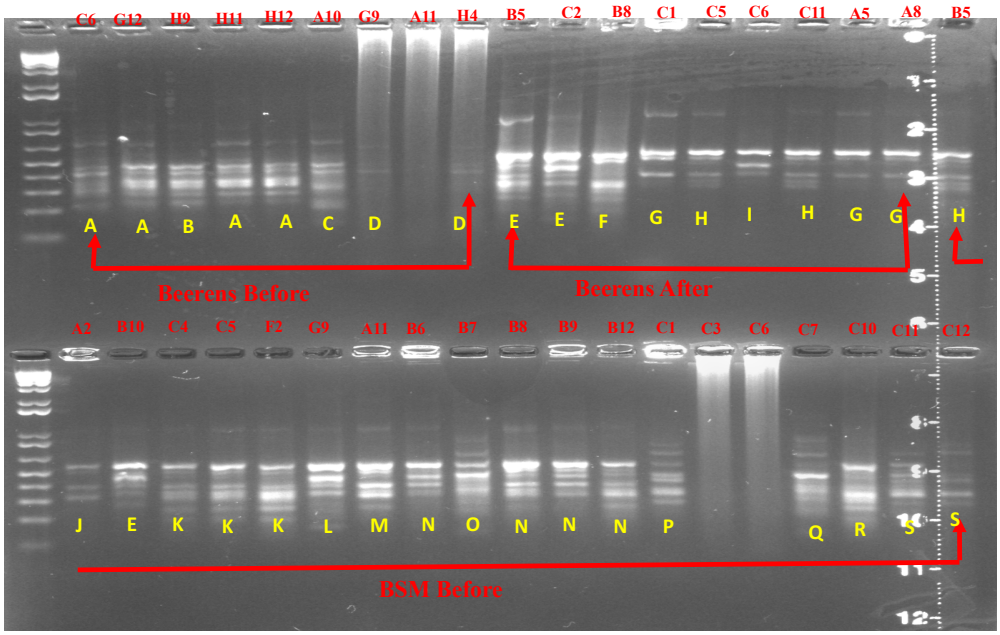


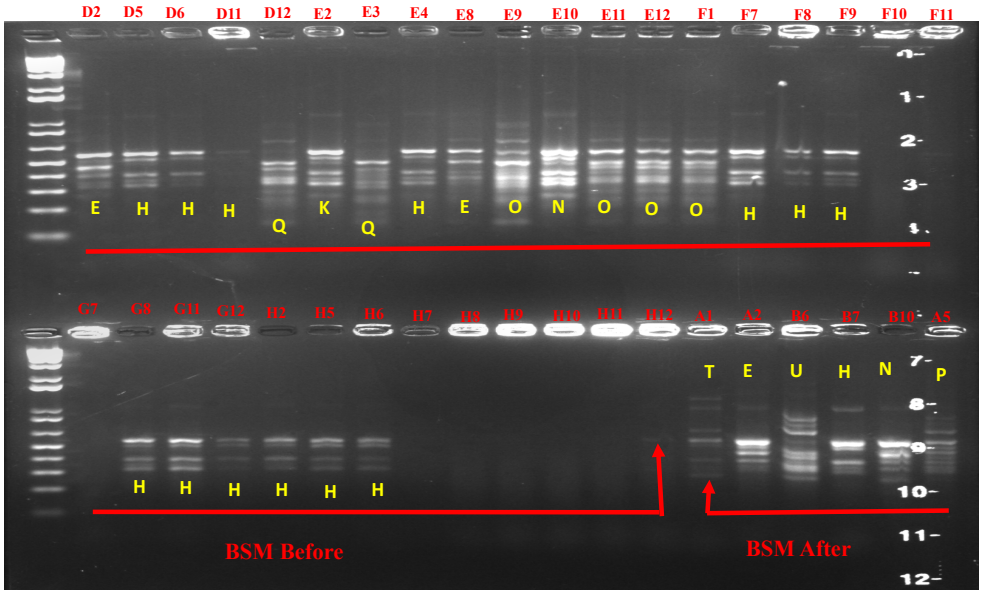
Figure 19: RAPD primer optimization using primers AP17, AP21, and AP22 performed on 4 isolates suspected to be from the bifidobacteria genus

Isolates were grouped based on their banding patterns (Figure **20** and **21**) as per **section 2.5.3**. Isolates with similar banding patterns were designated with the same letter (shown in yellow in **Figure 20** and **21**) and one isolate was chosen from each category and plate. RAPD performed on a total of 120 isolates of *Bifidobacterium* and *Bacteroides* (from pre-and post-solid introduction samples) shows 28 different banding patterns across all isolates (24 distinct patterns in bifidobacteria isolates, and 4 in *Bacteroides*).

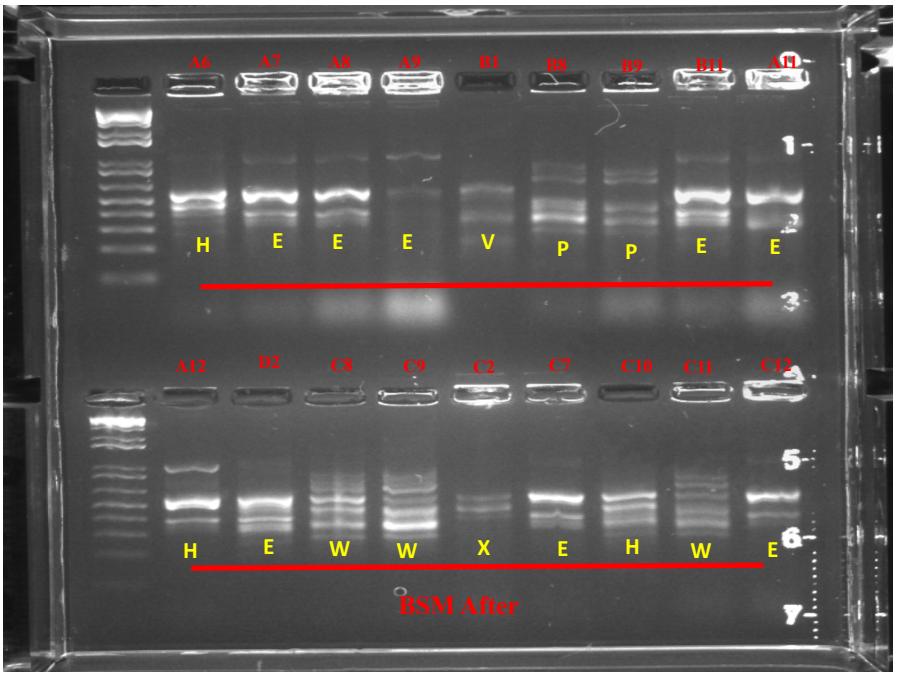
(a)



(b)


D2 D5 D6 D11 D12 E2 E3 E4 E8 E9 E10 E11 E12 F1 F7 F8 F9 F10 F11
E H H H Q K Q H E O N O O O H H H
G7 G8 G11 G12 H2 H5 H6 H7 H8 H9 H10 H11 H12 A1 A2 B6 B7 B10 A5
T E U H N P
H H H H H H
BSM Before
BSM After

(c)


A6
A7
A8
A9
B1
B8
B9
B11
A4
H
E
E
E
V
P
P
E
E
A12
D2
C8
C9
C2
C7
C10
C11
C12
H
E
W
W
X
E
H
W
E
BSM After
1
2
3
5
6
7

(d)

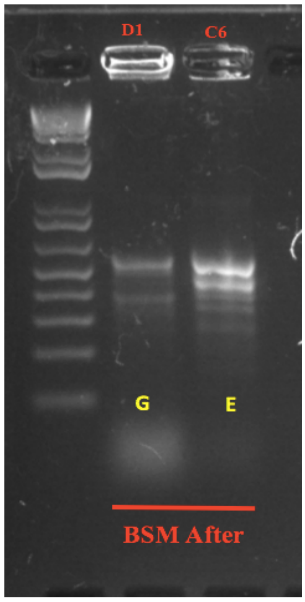
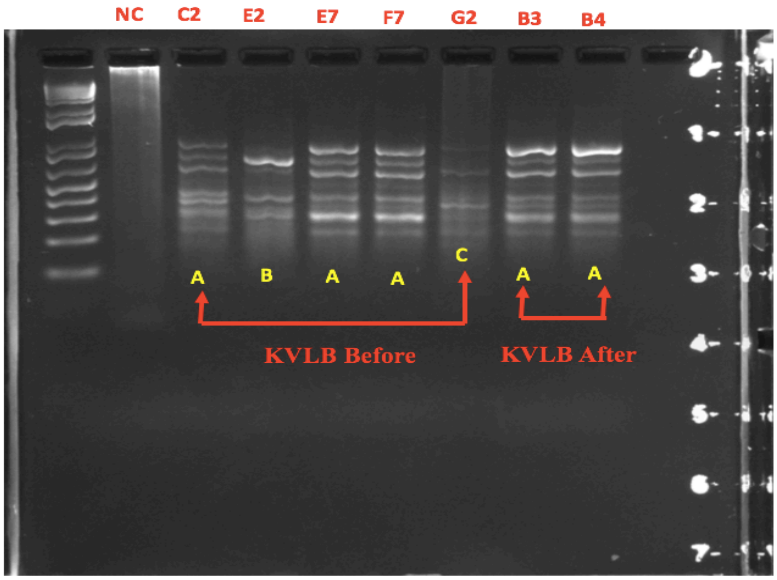


Figure 20: RAPD PCR preformed on 96 bifidobacteria isolates (as identified by ID sequencing) in (a), (b), (c), and (d). Red letters and numbers denote the location of each well the isolates were acquired from. Arrows indicate the media type the isolate was picked from and timing (before/after solid food introduction). Yellow letters are used to distinguish between different patterns (e.g. similar patterns are assigned the same letter).

(a)



(b)

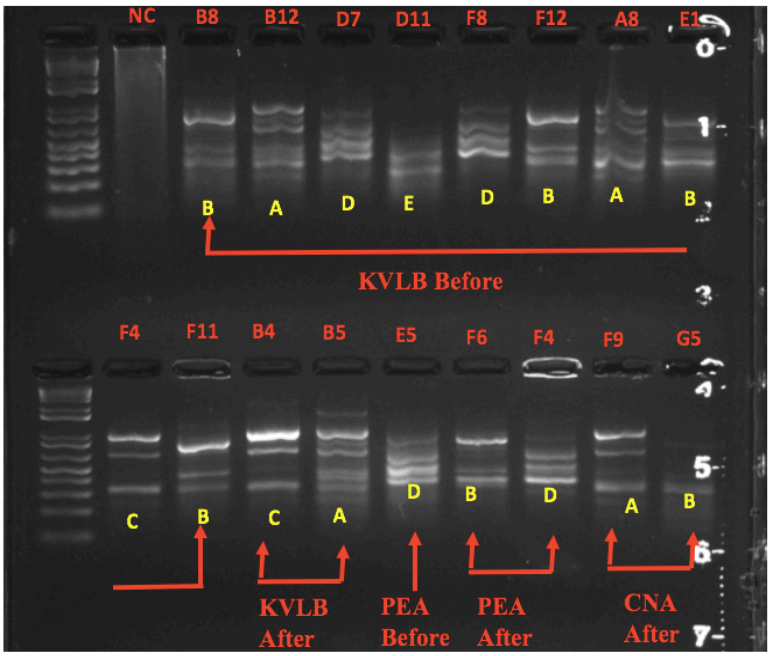


Figure 21: RAPD PCR preformed on 24 *Bacteroides* isolates (identified by ID sequencing) in (a), and (b). Red letters and numbers denote the location of each well the isolates were acquired from. Arrows indicate the media type the isolate was picked from and timing (before/after solid food introduction). Yellow letters are used to distinguish between different patterns (e.g. similar patterns are assigned the same letter).

Species level assignments based on full length 16S rRNA gene sequencing are presented in **Figure 22**. Of the 86 representative isolates chosen for full length 16S rRNA gene analysis, 7 isolates were not sequenced due to a low DNA yield following PCR clean-up, and 18 were unidentified following sequencing. The remainder are presented in **Figure 22** and include *Lactobacillus rhamnosus*, *Enterococcus faecalis*, *Lactobacillus casei*, *Clostridium CYP6*, *Clostridium tertium*, *Paeniclostridium sordelli*, *Bifidobacterium longum, Bacteroides ovatus*, *Bacteroides uniformis*, *Bacteroides vulgatus,*

Paeniclostridium sordelli, *Bifidobacterium breve*, *Clostridium tertium*, *Bradyrhizobium denitrificans* and *Clostridium sordelli*.

Those profiles generated were originally from isolates that were determined to be bifidobacteria and *Bacteroides* by ID sequencing. However, over the course of the study some isolates may have become contaminated, or may have not been pure and contained traces of other organisms that grow better on BHI medium.

Species assignments indicate that the RAPD banding pattern A and C belong to *Bacteroides ovatus*, pattern B to *B.uniformus*, and D to *B.vulgatus*. This indicates that diversity in RAPD patterns is mainly acquired due to species diversity rather than strain diversity. Moreover, *B.ovatus* appears to contain two strains with distinct banding patterns.

3.8 Fibre breakdown by infant gut bacteria

Based on functional assay data performed on isolates from **section 3.7**, *E.faecalis* isolates show 4 distinct patterns of substrate utilization indicating there might be 4 different strains present. On the other hand, almost all *B. breve* and *Paenclostridium sordelli* isolates show the same substrate utilization patterns, indicating there is one strain present, or that different strains within the same species may utilize substrates similarly (**Figure 22**).

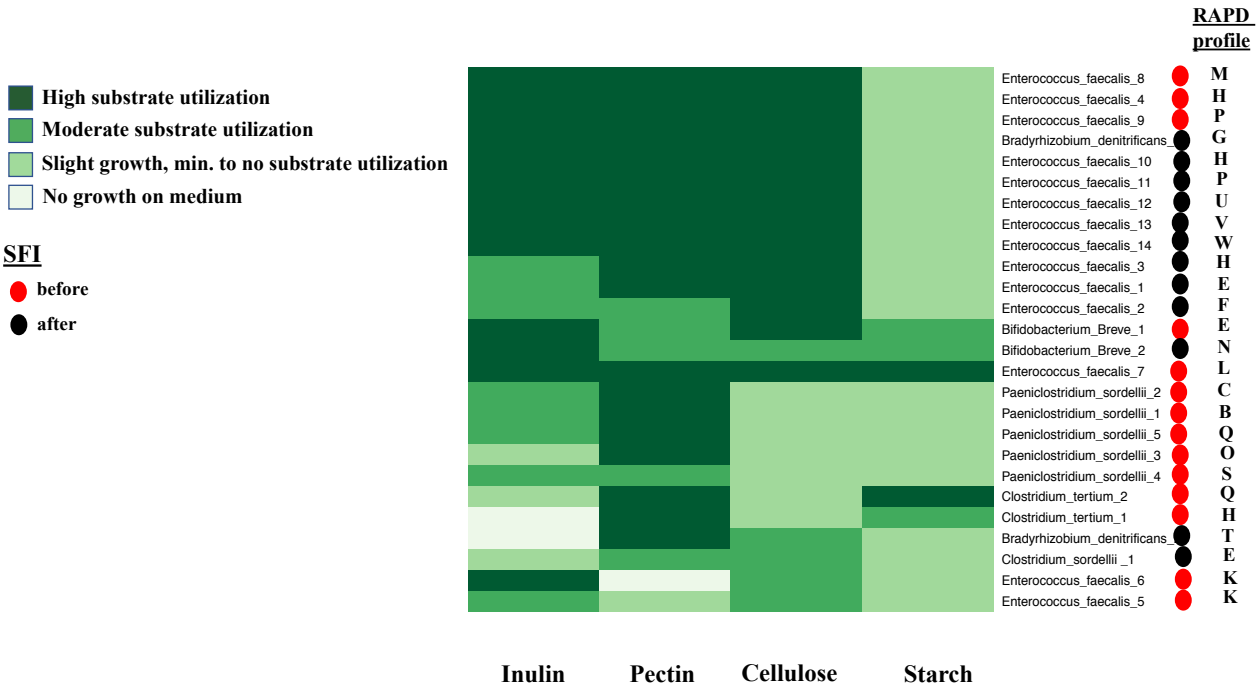


Figure 22: Functional substrate utilization profiles of 26 isolates acquired from the Beerens' and BSM plates and identified by sequencing 700 bp of the 16S rRNA gene. A letter corresponding to a specific RAPD profile is noted next to each isolate. Circles indicate whether isolates were acquired before (red) or after (black) solid food introduction (SFI).

3.9 Strain dynamics over the introduction of solid food

Paenclostridium sordelli appears on both Beerens' and BSM media only prior to solid food introduction, and all isolates from both plates, aside from 4, display similar functional substrate utilization patterns despite their variable RAPD profiles (**Figure 22**). Isolates 8, and 9 of *E. faecalis* appear on BSM medium prior to solid food introduction and have the same substrate utilization profile as isolates 11, 12, 13, 14 that appear on the same medium following solid food introduction, but different RAPD patterns. Alternatively, isolate 4, acquired before solid food introduction, shows both and identical

RAPD and substrate utilization profile as isolate 10, isolated after solid food introduction. This may indicate that the same strain potentially persisted following solid food introduction and retained the same metabolic profile. Alternatively, isolates 5 and 6 from BSM medium before solid food introduction have a similar substrate utilization profile and an identical RAPD pattern, but differ from the remaining isolates. Isolate 7, also acquired from BSM medium before solid food introduction has its own unique profile and differs from all other isolates. Furthermore, isolates 1,2, and 3 of *E. faecalis* acquired from Beerens' medium following solid food introduction share a unique pattern distinguishable from all other *E. faecalis* isolates as well, indicating that a strain's metabolic profile could have potentially shifted following solid food introduction. *B. breve* acquired from BSM medium before solid food introduction displays a similar substrate degradation pattern but a different RAPD profile from the one isolated from the same media type following solid food introduction. Additionally, many isolates acquired before solid food introduction tend to cluster together in terms of substrate utilization. The same can be noted regarding many isolates acquired after solid food introduction (**Figure 22**).

4.0 Discussion

4.1 A targeted bacterial culture strategy to capture bacterial isolates from the infant gut

Most microbiome studies focus on using culture independent molecular methods such as 16S rRNA gene sequencing (Turnbaugh *et al.,* 2007; Turroni *et al.*, 2015, Bergstrom *et al.,* 2014; Koenig *et al.,* 2011, David *et al.*, 2013). Such methods provide a good overview of the gut microbiome but often do not capture low abundance organisms due to insufficient sequence depth. Additionally, such methods do not provide any strain-level resolution or address functional changes over the course of a disturbance. It was previously believed that many organisms in the gut were "unculturable" (Fodor *et al.*, 2012; Stewart, 2012), however, Lau and others have demonstrated that culture-enriched molecular methods capture more diversity from stool than traditional molecular methods (Lau *et al.*, 2016; Browne *et al.*, 2016; Lagier *et al.*, 2012). In this study, we used molecular profiling as a tool to assess the effectiveness of our culture conditions at capturing organisms of interest in the breastfed, vaginally born infant gut over the course of solid food introduction. Molecular profiling of the 16S rRNA gene confirmed that we are, in fact, able to culture the infant gut microbiome. Our data show that our targeted culture strategy, combined with 16S rRNA gene profiling, was more sensitive than direct sequencing of stool alone. We captured 95% of the organisms we targeted, compared with 67% using molecular methods alone. Targeted culturing captured 69% more OTUs than 16S rRNA gene sequencing alone. These results demonstrate that culture adds sensitivity to molecular methods. Using this strategy we cultured 52 out of the 55 taxa of

interest in the infant gut and hence selected a targeted set of media conditions that we can use to culture future infant stool samples. Therefore, culture appears to be a great supplement to our longitudinal study. Using 24 media conditions, we captured a total of 3229 isolates from pre- and post-solid food introduction samples and 52 different bacterial families of interest. The higher sensitivity of culture relative to direct molecular methods allows us to track bacterial isolates overtime, monitor changes in their diversity patterns, and hence determine where the isolates are coming from. This is because culture helps us evaluate whether changes in diversity are attributed to changes in species richness or evenness overtime. The high sensitivity of targeted culturing allows us to detect organisms that fall below the limit of detection of direct molecular methods, and therefore, understand whether those organisms were already present in minute amounts prior to solid food introduction or have recently colonized the infant gut following the disturbance. Previous studies utilizing amplicon-based sequencing and metagenomics primarily attribute increases in microbial diversity, over the first 3 years of life, to an overall change in community diversity (a combined measure of species richness and evenness overtime) (Koenig *et al.*, 2010). Our data (**Figures 15** and **16**) supports this observation but highlights that it is evenness and not species richness that is driving the increase in community diversity.

Using targeted culturing we can isolate microbes of interest *in vitro* and generate a library of bacterial isolates to be studied further. For instance, we can identify each down to the species level and explore strain diversity, we can assess their substrate utilization profiles, measure their substrate usage, and explore strain dynamics following transition

to solid foods. For instance, given more time, we can potentially measure how a group of bacteria, such as bifidobacteria or *Bacteroides*, use the substrates available in the infant gut during the transition period to solid foods.

Despite the many strengths that culture adds to molecular profiling methods, many caveats, such as time and expenses, remain. Culture is both expensive and time-consuming since microbiological media sometimes require costly reagents, immediate processing of samples, and elaborate protocols. Additionally, research utilizing culturing methods often runs into bottlenecks. While conventional molecular profiling involves sequencing stool contents directly, targeted culture requires plating stool on media to obtain colonies, which take time to grow (Furrie, 2006) and often require anaerobic culture conditions. Contamination can occur, necessitating repetition of experiments, and selecting media conditions can drastically impact the results (Furrie, 2006). Additionally, isolating bacterial colonies and streaking them for purity is labor-intensive and subsequently requires identification through molecular techniques (Furrie, 2006), which further increases the cost. Therefore, although culture is a powerful tool to study the gut microbiome when combined with molecular profiling methods, we must be aware of all its challenges.

4.2 The impact of solid food introduction on the infant gut microbiome using molecular and culture methods

Solid food introduction appears to increase diversity of the gut microbiome largely due to an increase in species evenness. Prior to solid food introduction one phylotype of *Bifidobacterium* predominates. However, following solid food introduction

some of the less abundant organisms display an overall upward trend and the dominant *Bifidobacterium* phylotype decreases in abundance. Observable patterns also appear to differ within the same genus, with different phylotypes displaying different patterns. Although the results may appear promising, it is difficult to derive a definitive pattern from the data without applying those strategies to a larger infant sample size. Additionally, inter-individual diversity of the gut microbiome is relatively high, therefore, mechanistic studies are needed if we are to understand how such disturbances affect gut microbial populations during infancy.

4.3 Strain diversity and dynamics in the infant gut during the introduction of solid food

Three species of *Bacteroides* (*B. ovatus*, *B. uniformus*, and *B. vulgatus*), and *Clostridium* (*C. sordelli, C. tertium*, and *C. CYP6*), two species of *Bifidobacterium* (*B. longum* and *B. breve*), *Lactobacillus* (*L. casei* and *L. rhamnosus*), and *Bradyrhizobium (B. denitrificans),* and one species of *Enterococcus* (*E. faecalis*) and *Paeniclostridium* (*P. sordelli*), were isolated throughout the study.

RAPD data shows that the *E. faecalis* isolates had four distinguishable banding patterns and the isolates of *B. ovatus* had two indicating that there might be more than one strain present. Functional screens using inulin, pectin, cellulose, and starch indicated that *B. breve* and *Paeniclostridium sordelli* isolates had the same substrate utilization patterns, indicating that there might be one strain present for each species, or that different strains within the same genera utilize substrates similarly.

Three distinct functional profiles for *E. faecalis* arose prior to solid food introduction (A, B, and C), and two profiles afterward; one completely unique (D) and one shared (A) profile. This indicates that one of the *E. faecalis* strains either switched functionality following solid food introduction to profile D, or a new strain emerged with its own substrate degradation profile. Conversely, *B. breve* acquired from BSM media before solid food introduction displayed the same substrate degradation pattern as the one isolated from the same media type following solid food introduction, which implies that the same strain persisted following solid food introduction.

Functional screens combined with RAPD PCR can help us compare the functionality and banding patterns between isolates of the same species and hence track strains over the course of solid food introduction. However, RAPD profiling is not without its challenges. In this case, we relied on subjective methods to assess and score banding patterns. Although this was done by more than one individual to account for bias, it remains subjective and subject to error and could benefit from a computerized genetic software (e.g. BioNumerics) to analyze banding patterns to increase reliability. Pulse-field gel-electrophoresis could be a complementary method to those used here due to its consistent and reliable results (Parizad *et al.*, 2016) and because the software mentioned above can be used to preform cluster analysis and assess banding patterns to compare isolates (Sharma-Kuinkel *et al.,* 2016).

4.4 Importance and significance

It is important that we understand how environmental factors affect colonization and succession of the gut microbiome during infancy because the outcomes of these processes

have lasting consequences. Solid food introduction is one of the important determinants of gut microbiome health, especially since changes in the infant's diet dictate which communities are established and hence which communities remain into adulthood. Previous infant studies have not completely addressed the effect of solid food introduction on the infant gut microbiome directly, and tended to assess it in the context of many other confounding factors (Koenig *et al.,* 2011; Bergstrom *et al.,* 2014). Factors such as mode of feeding (e.g. bottle-feeding vs. breast feeding) (Coppa *et al.,* 2006; Guaraldi & Salvatori, 2012; Backhed *et al.,* 2015), mode of delivery (vaginal vs. cesarean) (Dominguez-Bello *et al.,* 2010; Penders *et al.,* 2006), and antibiotic administration (Yassour *et al.*, 2016) can alter the microbiome and confound the study of the effect of solid food introduction. The cohort, on which the methods studied here will be used, is specifically composed of breastfed, vaginally-born infants that were sampled intensively for three weeks surrounding the period of solid food introduction. By combining conventional molecular methods with culture strategies developed in this study we will be able to explore functional and strain dynamics of gut bacteria. Furthermore, it allows us to construct a library of isolates that we can study in a plethora of different ways and answer important questions about how they functionally respond to a disturbance such as solid food introduction. Therefore, although my research relies on a sample size of one infant, it is important because it paves the way for a larger study that utilizes this targeted culture strategy and the preliminary analysis methods I used to assess the effect of solid food introduction on a larger sample size of infants. Combining those techniques with metagenomic and metabolomic methods can provide us with a

comprehensive picture, not only about which microbes take part in this transition, but also how they interact as a community with one another and their host, how they alter their phenotypic expression in response to a disturbance, and the metabolites they produce and consume in their environment. Tracking those factors over time can help us understand how gut communities form and hence how this relates to long-term health outcomes.

4.5 Future directions

Isolates were retrieved from frozen stocks of BSM and Beeren's (for bifidobacteria) and PEA, CNA, and KVLB (for *Bacteroides*) and re-stamped on BHI 2 medium during the last aim of the study. This may have favored the growth of organisms other than bifidobacteria and *Bacteroides* that were present in the samples at small amounts (typically <5%). I suggest that isolates should be stamped or streaked on the media they were originally isolated from (e.g. BSM isolates to be streaked on BSM media). This may give the bifidobacteria or *Bacteroides* in the sample an added advantage to grow over contaminants.

5.0 Limitations

Due to the limited timeframe of my research project, I was able to acquire six samples from only one infant in this study. Therefore, despite the promising data acquired from those samples, we should be careful to derive a definitive pattern without applying those strategies to a larger infant sample size. Additionally, samples acquired from this infant were irregular due to the infant's irregular bowl movements. Therefore, there is a large

gap between day 12 and 25 where samples were not provided. This can be limiting since we cannot precisely predict species or OTU dynamics between those days.

Another limitation may have occurred when predicting the functional assay profile for all isolates in the study. Functional assay scoring is still in the works and is largely subjective. Isolates are typically scored for substrate utilization based on the presence of a halo surrounding the colony and its relative size compared to other halos on the same medium. Additionally, pH of metabolites is also deduced by the extent of media discoloration around the colony. Therefore, halo size and extent of discoloration are subjective measures that should potentially be determined by a computerized mechanism to yield more accuracy.

6.0 Conclusion

Most gut microbiome studies have focused on molecular profiles of bacterial communities generated through amplicon based sequencing (16S rRNA gene) or metagenomics. Although providing a good overview of the microbiome, this approach does not address microbial metabolism or interactions. A combined approach involving mechanistic studies of cultured isolates is needed if we are to understand how microbial communities change in response to a disturbance such as solid food introduction.

Our targeted culture strategy was more sensitive than 16S rRNA gene sequencing of stool alone at capturing target microbes. Using this culture strategy, we show that alpha diversity is primarily driven by species evenness rather than richness, as previously shown by culture-independent molecular profiling, following solid food introduction.

Using cultured isolates from the longitudinal study we show that different isolates of the same species utilize substrates differently (e.g. *Enterococcus faecalis*, *Bacteroides ovatus*) and others utilize substrates in the same way (e.g. *Bifidobacterium breve*, *Paeniclostridium sordelli*), indicating that one strain is present for each species, or that different strains within the same species utilize substrates similarly. We also show that isolates of the same species acquired before solid food introduction differ from the isolates acquired after (e.g. *Enterococcus faecalis*), and some remain the same (e.g. *Bifidobacterium breve*)

7.0 References

Agostoni, C., Riva, E., & Giovannini, M. (1995). Dietary fiber in weaning foods of young children. *Pediatrics, 96*, 1002-1005.

Ardissone, A. N., Cruz, D. M., Davis-Richardson, A. G., Rechcigl, K. T., Li, N., Drew, J. C., Murgas-Torrazza, R., Sharma, R., Hudak, M.L., Triplett, E.W., & Neu, J. (2014). Meconium Microbiome Analysis Identifies Bacteria Correlated with Premature Birth. *PLoS ONE, 9*(3).

Arrieta, M., Stiemsma, L. T., Amenyogbe, N., Brown, E. M., & Finlay, B. (2014). The Intestinal Microbiome in Early Life: Health and Disease. *Frontiers in Immunology, 5*.

Arsenic, R., Treue, D., Lehmann, A., Hummel, M., Dietel, M., Denkert, C., & Budczies, J.(2015). Comparison of targeted next-generation sequencing and Sanger sequencing for the detection of PIK3CA mutations in breast cancer. *BMC Clinical Pathology, 15*(1).

Atherly, T., & Ziemer, C. J. (2014). *Bacteroides* isolated from four mammalian hosts lack host-specific 16S rRNA gene phylogeny and carbon and nitrogen utilization patterns. *MicrobiologyOpen, 3*(2), 225-238.

Atlas R. Handbook of microbiological media. Boca Raton: CRC Press; 2004

Avelar, K. E., Pinto, L. J., Antunes, L. C., Lobo, L. A., Bastos, M. C., Domingues, R. M., & Ferreira, M. C. (1999). Production of bacteriocin by *Bacteroides* fragilis and partial characterization. *Letters in Applied Microbiology,29*(4), 264-268.

Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Li, Y., Xia, Y., Xie, Y., Zhong, H., Khan, M.T., Zhang, J., Li, J., Xiao, L., Al-Aama, J., Zhang, D., Lee, Y.S., Kotowska, D., Colding, C., Tremaroli, V., Yin, Y., Bergman, S., Xu, X., Madsen, L., Kristiansen, K., Dahlgren, J., & Wang, J. (2015). Dynamics and

Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host & Microbe,17*(6), 852.

Barrangou, R., & Horvath, P. (2012). CRISPR: New Horizons in Phage Resistance and Strain Identification. *Annual Review of Food Science and Technology, 3*(1), 143-162.

Bartram, A.K., Lynch, M.D., Stearns, J.C., Moreno-Hagelsieb, G., Neufeld, J.D. (2011). Generation of multi-million 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. *Applied and environmental microbiology, 77*(15), 5569-5569.

Baughn, A. D., & Malamy, M. H. (2004). The strict anaerobe *Bacteroides* fragilis grows in and benefits from nanomolar concentrations of oxygen. *Nature, 427*(6973), 441-444.

Beerens, H. (1990). An elective and selective isolation medium for *Bifidobacterium* spp. *Letters in Applied Microbiology, 11*(3), 155-157.

Belenguer, A., Duncan, S.H., Calder, A.G., Holtrop, G., Louis, P., Lobley, G.E., & Flint, H.J. (2006). Two Routes of Metabolic Cross-Feeding between *Bifidobacterium* Adolescentis and Butyrate-Producing Anaerobes from the Human Gut. *Applied and Environmental Microbiology*, *72* (5): 3593–99.

Bergstrom, A., Skov, T. H., Bahl, M. I., Roager, H. M., Christensen, L. B., Ejlerskov, K. T., Molgaard, C., Michaelsen, K.F., & Licht, T. R. (2014). Establishment of Intestinal Microbiota during Early Life: A Longitudinal, Explorative Study of a Large Cohort of Danish Infants. *Applied and Environmental Microbiology,80*(9), 2889-2900.

Bernet, M. F., Brassart, D., Neeser, J. R., & Servin, A. L. (1994). *Lactobacillus acidophilus* LA 1 binds to cultured human intestinal cell lines and inhibits cell attachment and cell invasion by enterovirulent bacteria. *Gut, 35*(4), 483-489.

Bhardwaj, A., Gupta, H., Kapila, S., Kaur, G., Vij, S., & Malik, R.K. (2010). Safety assessment and evaluation of probiotic potential of bacteriocinogenic Enterococcus faecium KH 24 strain under *in vitro* and *in vivo* conditions. *Int J Food Microbiol., 141*, 156–164.

Bourget, N., Simonet, J. M., & Decaris, B. (1993). Analysis of the genome of the five *Bifidobacterium* breve strains: plasmid content, pulsed-field gel electrophoresis genome size estimation and rrn loci number. *FEMS Microbiology Letters, 110*(1), 11-20.

Breitbart, M., Hewson, I., Felts, B., Mahaffy, J.M., Nulton, J., Salamon, P., & Rohwer, F.(2003). Metagenomic analyses of an uncultured viral community from human feces. *J. Bacteriol., 185*, 6220–23.

Briner, A. E., Lugli, G. A., Milani, C., Duranti, S., Turroni, F., Gueimonde, M., Margolles, A., van Sinderen, D., Ventura, M., & Barrangou, R. (2015). Occurrence and Diversity of CRISPR-Cas Systems in the Genus *Bifidobacterium*. *Plos One, 10*(7).

Browne, H.P., Froster, S.C., Anonye, B.O., Kumar, N., Neville, B.A., Stares, M.D., Goulding, D., & Lawley, T.D. (2016). Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature, 533*, 543-546.

Bull, M. J., & Plummer, N. T. (2014). Part 1: The Human Gut Microbiome in Health and Disease. *Integr Med (Encinitas), 13*(6), 17-22.

Callahan, B. J., Mcmurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods, 13*(7), 581-583.

Chakravorty, S., Helb, D., Burday, M., Connell, N., & Alland, D. (2007). A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of Microbiological Methods, 69*(2), 330-339.

Chistoserdova, L. (2013). Is metagenomics resolving identification of functions in microbial communities? *Microbial Biotechnology, 7*(1), 1-4.

Chu, S., Chen, Q., Chen, Y., Bao, Y., Wu, M., & Zhang, J. (2017). Cesarean section without medical indication and risk of childhood asthma, and attenuation by breastfeeding. *Plos One, 12*(9).

Coppa, G., Zampini, L., Galeazzi, T., & Gabrielli, O. (2006). Prebiotics in human milk: A review. *Digestive and Liver Disease, 38*.

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., Ling, A.V., Devlin, A.S., Varma, Y., Fischbach, M.A., Biddinger, S.B., Dutton, R.J., &Turnbaugh, P. J. (2013). Diet rapidly and reproducibly alters the human gut microbiome. *Nature,505*(7484), 559-563.

Costea, P. I., Munch, R., Coelho, L. P., Paoli, L., Sunagawa, S., & Bork, P. (2017). MetaSNV: A tool for metagenomic strain level analysis. *Plos One, 12*(7).

Dominguez-Bello, M. G., Costello, E. K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., & Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences,107*(26), 11971-11975.

Duncan, S.H., Louis, P., & Flint, H.J. (2004). Lactate-Utilizing Bacteria, Isolated from Human Feces, That Produce Butyrate as a Major Fermentation Product. *Applied and Environmental Microbiology, 70* (10), 5810–17.

Eley, A., Greenwood, D., & O'Grady, F. (1984). Comparative Growth of *Bacteroides* Species in Various Anaerobic Culture Media. *Journal of Medical Microbiology, 19*, 195-201.

Fraher, M. H., O'toole, P. W., & Quigley, E. M. (2012). Techniques used to characterize the gut microbiota: a guide for the clinician. *Nature Reviews Gastroenterology & Hepatology, 9*(6), 312-322.

Fodor, A. A., Desantis, T. Z., Wylie, K. M., Badger, J. H., Ye, Y., Hepburn, T., . . . Earl, A. M. (2012). The "Most Wanted" Taxa from the Human Microbiome for Whole Genome Sequencing. PLoS ONE, 7(7)

Frank, J. A., Reich, C. I., Sharma, S., Weisbaum, J. S., Wilson, B. A., & Olsen, G. J. (2008). Critical Evaluation of Two Primers Commonly Used for Amplification of Bacterial 16S rRNA Genes. *Applied and Environmental Microbiology, 74*(8), 2461-2470.

Fuentes-Zaragoza, E., Riquelme-Navarrete, M.J., Sanchez-Zapata, E., & Perez-Alvarez, J.A. (2010). Resistant starch as functional ingredient: A review. *Food Research International, 43*(4), 931-942.

Furrie, E. (2006). A molecular revolution in the study of intestinal microflora. *Gut, 55*(2), 141-143.

Giuffré, A., Forte, E., Arese, M., Davletshin, A., Mastronicola, D., Sarti, P., & Borisov, V. B. (2014). Cytochrome bd oxidase and bacterial tolerance to oxidative and nitrosative stress. *Biochimica Et Biophysica Acta (BBA) - Bioenergetics, 1837*.

Goodman, A. L., Kallstrom, G., Faith, J. J., Reyes, A., Moore, A., Dantas, G., & Gordon, J. I. (2011). Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proceedings of the National Academy of Sciences, 108*(15), 6252-6257.

Gorbach, S.L. (1996). Chapter 95 Microbiology of the Gastrointestinal Tract. In S. Baron (Ed.), Medical Microbiology (4th ed.). Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK7670/

Guaraldi, F., & Salvatori, G. (2012). Effect of Breast and Formula Feeding on Gut Microbiota Shaping in Newborns. *Frontiers in Cellular and Infection Microbiology, 2*.

Hadrys, H., Balick, M., & Schierwater, B. (1992). Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Molecular Ecology, 1*(1), 55-63.

Houghteling, P. D., & Walker, W. A. (2015). Why Is Initial Bacterial Colonization of the Intestine Important to Infants' and Children's Health? *Journal of Pediatric Gastroenterology and Nutrition, 60*(3), 294-307.

Janda, J. M., & Abbott, S. L. (2007). 16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls. *Journal of Clinical Microbiology, 45*(9), 2761-2764.

Jarocki, P., Podleśny, M., Komoń-Janczara, E., Kucharska, J., Glibowska, A., & Targoński, Z. (2016). Comparison of various molecular methods for rapid differentiation of intestinal bifidobacteria at the species, subspecies and strain level. *BMC Microbiology, 16(1)*.

Jovel, J., Patterson, J., Wang, W., Hotte, N., Okeefe, S., Mitchel, T., Perry, T., Kao, D., Mason, A.L., Madsen, K.L., &Wong, G. K. (2016). Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics. *Frontiers in Microbiology, 7*.

Kamada, N., Seo, S., Chen, G. Y., & Núñez, G. (2013). Role of the gut microbiota in immunity and inflammatory disease. *Nature Reviews Immunology, 13*(5), 321-335.

Karlsson, F. H., Ussery, D. W., Nielsen, J., & Nookaew, I. (2011). A Closer Look at *Bacteroides*: Phylogenetic Relationship and Genomic Implications of a Life in the Human Gut. *Microbial Ecology, 61*(3), 473-485.

Kelly, C.J., Zheng, L., Campbell, E.L., Saeedi, B., Scholz, C.C., Bayless, A.J., Wilson, K.E., Glover, L.E., Kominsky, D.J., Magnuson, A., Weir, T.L., Ehrentraut, S.F., Pickel, C., Kuhn, K.A., Lanis, J.M., Nguyen, V., Taylor, C.T., & Colgan, S.P.(2015). Crosstalk between Microbiota-Derived Short-Chain Fatty Acids and Intestinal Epithelial HIF Augments Tissue Barrier Function. *Cell Host & Microbe, 17*(5), 662-671.

Koenig, J. E., Spor, A., Scalfone, N., Fricker, A. D., Stombaugh, J., Knight, R., Angenet, L.T., & Ley, R. E. (2010). Succession of microbial consortia in the developing

infant gut microbiome. *Proceedings of the National Academy of Sciences,108*(Supplement_1), 4578-4585.

Lau, J. T., Whelan, F. J., Herath, I., Lee, C. H., Collins, S. M., Bercik, P., & Surette, M. G. (2016). Capturing the diversity of the human gut microbiota through culture-enriched molecular profiling. *Genome Medicine, 8*(1).

Lagier, J., Armougom, F., Million, M., Hugon, P., Pagnier, I., Robert, C., Bittar, F., Fournous, G., Gimenez, G., Maraninchi, M., Trape, J.F., Koonin, E.V., La Scola, B., & Raoult, D. (2012). Microbial culturomics: Paradigm shift in the human gut microbiome study. *Clinical Microbiology and Infection, 18*(12), 1185-1193.

Lievin, V. (2000). *Bifidobacterium* strains from resident infant human gastrointestinal microflora exert antimicrobial activity. *Gut, 47*(5), 646-652.

Mantis, N. J., Rol, N., & Corthésy, B. (2011). Secretory IgA's complex roles in immunity and mucosal homeostasis in the gut. *Mucosal Immunology, 4*(6), 603-611.

Martens, E. C., Chiang, H. C., & Gordon, J. I. (2008). Mucosal Glycan Foraging Enhances Fitness and Transmission of a Saccharolytic Human Gut Bacterial Symbiont. *Cell Host & Microbe, 4*(5), 447-457.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal,17*(1), 10.

Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G., & Neufeld, J. D. (2012). PANDAseq: Paired-end assembler for illumina sequences. *BMC Bioinformatics, 13*(1), 31.

Mazmanian, S. K., & Kasper, D. L. (2006). The love–hate relationship between bacterial polysaccharides and the host immune system. *Nature Reviews Immunology, 6(*11), 849-858.

Meehan, B. M., Baughn, A. D., Gallegos, R., & Malamy, M. H. (2012). Inactivation of a single gene enables microaerobic growth of the obligate anaerobe *Bacteroides* fragilis. *Proceedings of the National Academy of Sciences, 109*(30), 12153-12158.

Michael G. Surette Laboratory. (n.d.). Retrieved December 11, 2016, from http://www.surettelab.ca/lab/

Microbiology Culture Media and Laboratory Supplies - by Hardy Diagnostics. (n.d.). Retrieved February 12, 2017, from http://hardydiagnostics.com/

Milani, C., Turroni, F., Duranti, S., Lugli, G. A., Mancabelli, L., Ferrario, C., van Sinderen, D., & Ventura, M. (2015). Genomics of the Genus *Bifidobacterium* Reveals Species-Specific Adaptation to the Glycan-Rich Gut Environment. *Applied and Environmental Microbiology, 82*(4), 980-991.

Miteva, V., Selenska-Pobell, S., & Mitev, V. (2001). Random and repetitive primer amplified polymorphic DNA analysis of Bacillus sphaericus. *Journal of Applied Microbiology, 86*(6), 928-936.

Mossie, K. G., Robb, F. T., Jones, D. T., & Woods, D. R. (1981). Inhibition of ribonucleic acid polymerase by a bacteriocin from *Bacteroides* fragilis*. Antimicrobial Agents and Chemotherapy, 20(*4), 437-442.

Mota-Meira, M., Morency, H., Lavoie, M.C. (2005). In vivo activity of mutacin B-Ny266. *J Antimicrob Chemother, 56*, 869–871.

Neu, J., & Rushing, J. (2011). Cesarean Versus Vaginal Delivery: Long-term Infant Outcomes and the Hygiene Hypothesis. *Clinics in Perinatology, 38*(2), 321-331.

Ocallaghan, A., & Sinderen, D. V. (2016). Bifidobacteria and Their Role as Members of the Human Gut Microbiota. *Frontiers in Microbiology*, *7*.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Henry, M., Stevens, H., &Wagner, H. (2015). Vegan: Community Ecology Package. Retrieved October 11, 2018 from http://CRAN.R-project.org/package=vegan.

Parche, S., Beleut, M., Rezzonico, E., Jacobs, D., Arigoni, F., Titgemeyer, F., & Jankovic, I. (2006). Lactose-over-Glucose Preference in *Bifidobacterium* longum NCC2705: glcP, Encoding a Glucose Transporter, Is Subject to Lactose Repression. *Journal of Bacteriology, 188*(4), 1260-1265.

Parizad, E. G., Parizad, E. G., & Valizadeh, A. (2016). The Application of Pulsed Field Gel Electrophoresis in Clinical Studies. *Journal of Clinical And Diagnostic Research, 10*(1), DE01-DE04.

Patel, J. B. (2001). 16S rRNA Gene Sequencing for Bacterial Pathogen Identification in the Clinical Laboratory. *Molecular Diagnosis, 6*(4), 313-321.

Penders, J., Thijs, C., Vink, C., Stelma, F. F., Snijders, B., Kummeling, I., van den Brandt, P.A., & Stobberingh, E. E. (2006). Factors Influencing the Composition of the Intestinal Microbiota in Early Infancy. *Pediatrics,118*(2), 511-521.

Peterson, L. W., & Artis, D. (2014). Intestinal epithelial cells: Regulators of barrier function and immune homeostasis. *Nature Reviews Immunology, 14*(3), 141-153.

Poretsky, R., Rodriguez-R, L. M., Luo, C., Tsementzi, D., & Konstantinidis, K. T. (2014). Strengths and Limitations of 16S rRNA Gene Amplicon Sequencing in Revealing Temporal Microbial Community Dynamics. *PLoS ONE, 9*(4).

Rea, M. C., Clayton, E., O'Connor, P. M., Shanahan, F., Kiely, B., Ross, R. P. & Hill, C. (2007). Antimicrobial activity of lacticin 3147 against clinical *Clostridium difficile* strains. *J Med Microbiol, 56*, 940–946.

Rhee, K., Sethupathi, P., Driks, A., Lanning, D. K., & Knight, K. L. (2004). Role of Commensal Bacteria in Development of Gut-Associated Lymphoid Tissues and Preimmune Antibody Repertoire. *The Journal of Immunology, 172*(2), 1118-1124.

Rivière, A., Gagnon, M., Weckx, S., Roy, D., & Vuyst, L. D. (2015). Mutual Cross-Feeding Interactions between *Bifidobacterium* longum subsp. longum NCC2705 and Eubacterium rectale ATCC 33656 Explain the Bifidogenic and Butyrogenic Effects of Arabinoxylan Oligosaccharides. *Applied and Environmental Microbiology, 81*(22), 7767-7781.

Rodríguez, J. M., Murphy, K., Stanton, C., Ross, R. P., Kober, O. I., Juge, N., Avershina, E., Rudi, K., Narbad, A., Jenmalm, M.C., Marchesi, J.R., & Collado, M. C. (2015). The composition of the gut microbiota throughout life, with an emphasis on early life. *Microbial Ecology in Health & Disease, 26*.

Roessner, U., & Bowne, J. (2009). What is metabolomics all about? *BioTechniques, 46*(5),363-365.

Roy, D., Ward, P., & Champagne, G. (1996). Differentiation of bifidobacteria by use of

pulsed-field gel electrophoresis and polymerase chain reaction. International *Journal of Food Microbiology, 29*(1), 11-29.

Sela, D. A., Chapman, J., Adeuya, A., Kim, J. H., Chen, F., Whitehead, T. R., Lapidus, A., Rokhsar, D.S., Lebrilla, C.B., German, J.B., Price, N.P., Richardson, P.M., & Mills, D. A. (2008). The genome sequence of *Bifidobacterium* longum subsp. infantis reveals adaptations for milk utilization within the infant microbiome. *Proceedings of the National Academy of Sciences, 105(48)*, 18964-18969.

Selenska-Pobell, S., Evguenieva-Hackenberg, E., & Schwickerath, O. (1995). Random and Repetitive Primer Amplified Polymorphic DNA Analysis of Five Soil and Two Clinical Isolates of Rahnella aquatilis. *Systematic and Applied Microbiology*, 18(3), 425-438.

Shaheduzzaman, S. M., Akimoto, S., Kuwahara, T., Kinouchi, T., & Ohnishi, Y. (1997). Genome Analysis of *Bacteroides* by Pulsed-field Gel Electrophoresis: Chromosome Sizes and Restriction Patterns. *DNA Research, 4*(1), 19-25.

Sharma-Kuinkel, B. K., Rude, T. H., & Fowler, V. G. (2016). Pulse Field Gel Electrophoresis. *Methods in Molecular Biology (Clifton, N.J.)*, *1373*, 117–130.

Sharon, I., Morowitz, M. J., Thomas, B. C., Costello, E. K., Relman, D. A., & Banfield, J. F. (2012). Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. *Genome Research, 23*(1), 111-120.

Shen, X. J., Rawls, J. F., Randall, T. A., Burcall, L., Mpande, C., Jenkins, N., Jovov, B., Abdo, Z., Sandler, R.S., & Keku, T. O. (2010). Molecular characterization of mucosal adherent bacteria and associations with colorectal adenomas. *Gut Microbes,1*(3), 138-147.

Shen, C., Lin, M., Lin, H., Lin, C., Fu, L., & Fu, Y. C. (2013). The natural course of eczema from birth to age 7 years and the association with asthma and allergic rhinitis: A population-based birth cohort study. *Allergy and Asthma Proceedings*, *34*(1), 78-83.

Simioni, J., Hutton, E. K., Gunn, E., Holloway, A. C., Stearns, J. C., Mcdonald, H., Mousseau, A., Schertzer, J.D., Ratcliffe, E.M., Thabane, L., Surette, M.G., & Morrison, K. M. (2016). A comparison of intestinal microbiota in a population of low-risk infants exposed and not exposed to intrapartum antibiotics: The Baby & Microbiota of the Intestine cohort study protocol. *BMC Pediatrics, 16*(1), 183.

Sonnenburg, J. L., Angenent, L. T., & Gordon, J. I. (2004). Getting a grip on things: How do communities of bacterial symbionts become established in our intestine? *Nature Immunology, 5*(6), 569-573.

Sonnenburg, E. D., Zheng, H., Joglekar, P., Higginbottom, S. K., Firbank, S. J., Bolam, D. N., & Sonnenburg, J. L. (2010). Specificity of Polysaccharide Use in Intestinal *Bacteroides* Species Determines Diet-Induced Microbiota Alterations. *Cell, 141*(7), 1241-1252.

Sonnenburg, J.L., Xu, J., Leip, D.D., Chen, C.H., Westover, B.P., Weatherford, J., Buhler, J.D., & Gordon, J.I. (2005). Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science, 307* (5717), 1955-9.

Stewart, E. J. (2012). Growing Unculturable Bacteria. Journal of Bacteriology, 194(16), 4151-4160.

Tanaka R., Takayama H., Morotomi M., Kuroshima T., Ueyama S., Matsumoto K., Kuroda, A., & Mutai, M. (1983). Effects of administration of TOS and *Bifidobacterium breve* 4006 on the human fecal flora. *Bifidobacteria and Microflora, 2*(1), 17–24.

Thomson, P., Medina, D. A., & Garrido, D. (2017). Human milk oligosaccharides and infant gut bifidobacteria: Molecular strategies for their utilization. *Food Microbiology, 75*, 37-46.

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C., Knight, R., and Gordon, J.I. (2007). The human microbiome project: exploring the microbial part of ourselves in a changing world. *Nature, 449*(7164).

Turroni, F., Peano, C., Pass, D. A., Foroni, E., Severgnini, M., Claesson, M. J., Kerr, C., Hourihane, J., Murray, D., Fuligni, F., Gueimonde, M., Margolles, A., De Bellis, G., O'Toole, P.W., van Sinderen, D., Marchesi, J.R., &Ventura, M. (2012). Diversity of Bifidobacteria within the Infant Gut Microbiota. *PLoS ONE, 7*(5).

Ulluwishewa, D., Anderson, R. C., Mcnabb, W. C., Moughan, P. J., Wells, J. M., & Roy, N. C. (2011). Regulation of Tight Junction Permeability by Intestinal Bacteria and Dietary Components. *Journal of Nutrition, 141*(5), 769-776.

Weemaes, A., Klasen, I., Go¨ertz, J., Beldhuis-Valkis, M., Olafsson, O., & Haraldsson, A. (2003). T Development of Immunoglobulin A in Infancy and Childhood. *Scandinavian Journal of Immunology , 58*, 642-648.

Wexler, H. M. (2007). *Bacteroides*: the Good, the Bad, and the Nitty-Gritty. *Clinical Microbiology Reviews, 20*(4), 593-621.

Wexler, A. G., & Goodman, A. L. (2017). An insiders perspective: *Bacteroides* as a window into the microbiome. *Nature Microbiology, 2*(5).

Whelan, F.J., Surette, M.G. (2017) A comprehensive evaluation of the sl1p pipeline for 16S rRNA gene sequencing analysis. *Microbiome, 5*(1),100.

Yassour, M., Vatanen, T., Siljander, H., Hämäläinen, A., Härkönen, T., Ryhänen, S. J., Franzosa, E.A., Vlamakis, H., Huttenhower, C., Gevers, D., Lander, E.S., Knip, M., DIABIMMUNE Study Group, & Xavier, R. J. (2016). Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. *Science Translational Medicine, 8*(343).

Ye, Y. (2010). Identification and quantification of abundant species from pyrosequences of 16S rRNA by consensus alignment. *Proceedings (IEEE Int Conf Bioinformatics Biomed)*, 153-157.

Ze, X., Duncan, S.H., Louis, P., & Flint, H.J. (2012). Ruminococcus Bromii Is a Keystone Species for the Degradation of Resistant Starch in the Human Colon. *The ISME Journal*, *6* (8), 1535–43.

Zhang, G., Gao, B., Adeolu, M., Khadka, B., & Gupta, R. S. (2016). Phylogenomic Analyses and Comparative Studies on Genomes of the Bifidobacteriales: Identification of Molecular Signatures Specific for the Order Bifidobacteriales and Its Different Subclades. *Frontiers in Microbiology, 7*.

8.0 Appendix I

8.1 Awards

Medical Sciences Research Day: Graduate Excellence in Research Award
McMaster University, Hamilton, ON, January 2017

Faculty of Health Sciences Graduate Programs Outstanding Achievement Award
McMaster University, Hamilton, ON, May 2018

Faculty of Health Sciences Graduate Programs Excellence Award
McMaster University, Hamilton, ON, May 2018

8.2 Conference Presentations

Presented a poster at the Medical Sciences Research Day held in January 2017 at McMaster University, Hamilton, ON.

Presented a poster at the r 5th annual EMPhasis on Health Symposium held in April 2017 at McMaster University, Hamilton, ON.

Presented a poster at the Institute of Infectious Disease Research Trainee day held in October 2017 at McMaster University, Hamilton, ON.

Presented a poster at the Canadian Society of Microbiologists Conference held in June 2018 at University of Manitoba, Winnipeg, ON.

8.3 R scripts

A1. R scripts used for Aim 1 data analysis and figures

```
## This script contains the OTU table (no proportions)
setwd("~/Desktop/Project/JCSA1to3_picked_otus_abundantotu_gg2011")
initial <- read.csv("SPE_OTU.csv",row.names = 1) ##OTU data upload
initial[1:5, 1:5]
dim(initial)
initial2 <- initial[, 1:(ncol(initial)-1)] # or could use intitial[,-113]
sums <- colSums(initial2) ##column sums
sums <- sort(sums) ##sort by number/sequence
head(sums) ##just to see what my lowest sequence depth is
```

```
meta_table <- read.csv("JCSA_map.csv", row.names = 1, na.strings = "") ## Load the
 meta data sheet
View(meta_table)
factor(meta_table$media)
factor(meta_table$condition)
meta_table <- meta_table[-75,] ##deleting rows (if u want to view the rows use
row.names)
meta_table <- meta_table[-53,]
meta_table <- meta_table[-12,]
meta_table <- meta_table[-27,]


newnames <- row.names(meta_table)
newnames <- gsub(pattern = "pp",replacement = "PP", x = newnames) ## how to replace
 things
row.names(meta_table) <- newnames



initial.new <- t(initial) ##transposing
initial.new <- initial.new [-113,]
newnames2 <- row.names(initial.new)
newnames2 <- gsub(pattern = "st",replacement = "", x = newnames2)
row.names(initial.new) <- newnames2
newnames3 <- row.names(initial.new)
newnames3 <- gsub(pattern = "JSCA",replacement = "JCSA", x = newnames3)
row.names(initial.new) <- newnames3
meta.r <- meta_table[rownames(initial.new),]
final2.r <- data.frame(meta.r, initial.new) ##final merged table

# subset all aerobic
AER <- final2.r[final2.r$condition %in% "AERO",] ##Aerobic table

ANA <- final2.r[final2.r$condition %in% "ANA",] ##Anaerobic table

Stool <- final2.r[final2.r$condition %in% "stool",]


# creating a binary table

m<- matrix(as.numeric(unlist(initial.new)),nrow=nrow(initial.new)) ## convert data
 frame to matrix
m[m<=10] <- 0 #creating binary variables
```

```
m[m>10] <- 1
initial.binary <- as.data.frame(m)

trial <- data.frame(final2.r$condition, initial.binary)
trial2 <- data.frame(final2.r$media, trial)
trial3 <- data.frame(rownames(final2.r), trial2, row.names = 1)

akh <- colnames(trial3)
akh<- gsub(pattern = "final2.r.media",replacement = "Media", x = akh)
colnames(trial3) <- akh
akh2 <- colnames(trial3)
akh2<- gsub(pattern = "final2.r.condition",replacement = "Condition", x = akh2)
colnames(trial3) <- akh2
akh3 <- colnames(trial3)
akh3<- gsub(pattern = "Media",replacement = "media", x = akh)
colnames(trial3) <- akh3


##adding sample codes

meta_table2 <- read.csv("JCSA_map2.csv",row.names = 1)

newnames.r <- row.names(meta_table2)
newnames.r <- gsub(pattern = "pp",replacement = "PP", x = newnames.r) ## how to
 replace things
row.names(meta_table2) <- newnames.r

meta.r2 <- meta_table2[rownames(trial3),]
trial3 <- data.frame(meta.r2, trial3)

akh5 <- colnames(trial3)
akh5<- gsub(pattern = "final2.r.condition",replacement = "condition", x = akh5)
colnames(trial3) <- akh5
akh6 <- colnames(trial3)
akh6<- gsub(pattern = "meta.r2",replacement = "Sample", x = akh5)
colnames(trial3) <- akh6

##seperating samples into 6 tables

St.1 <- trial3[trial3$Sample %in% "stA",] ##stool1
St.2 <- trial3[trial3$Sample %in% "stB",] ##stool2
St.3 <- trial3[trial3$Sample %in% "stC",] ##stool3
```

```
PP.1 <- trial3[trial3$Sample %in% "A",] ##PP for sample 1
PP.2 <- trial3[trial3$Sample %in% "B",] ##PP for sample 2
PP.3 <- trial3[trial3$Sample %in% "C",] ##PP for sample 3

sub1 <- trial3[trial3$condition %in% "AERO",]
sub2 <-  trial3[trial3$condition %in% "ANA",]
sub3 <- rbind(sub1, sub2)

## finding match sums
# make a list of rownames
# startt a for loop

## matching stool.1 to PP.1
st.1.t <- t(St.1[,4:720])
J1 <- t(PP.1[,4:720])
bound <- cbind(st.1.t, J1) ##1
colz <- c(1:38) ## creating a vector
for (i in colz){
  m1 <-ifelse(bound[,1]==0, 0,ifelse(bound[,1]==bound[,i], T, NA))
  sums <- sum(m1, na.rm=TRUE)
  sums2 <- colSums(bound, na.rm= TRUE)
  print("************************")
  print(i)
  print(sums) ##degree of overlap with stool
  print(".....................")
  print(sums2) ##sum of OTUs in each sample
}

## matching PP.1 to PP.1
J1 <- t(PP.1[,4:720])
colz <- c(1:37)
for (i in colz){
  m1 <-ifelse(J1[,1]==0, 0,ifelse(J1[,1]==J1[,i], T, NA)) ##change number after J1
  sums <- sum(m1, na.rm=TRUE)
  sums2 <- colSums(J1, na.rm= TRUE)
  print("************************")
  print(i)
  print(sums)
  print(".....................")
  print(sums2)
}
```

```
#########################################################################
 #########
## matching stool.2 to PP.2
st.2.t <- t(St.2[,4:720])
J2 <- t(PP.2[,4:720])
bound2 <- cbind(st.2.t, J2) ##1
colz <- c(1:32) ## creating a vector
for (i in colz){
  m2 <-ifelse(bound2[,1]==0, 0,ifelse(bound2[,1]==bound2[,i], T, NA))
  sums3 <- sum(m2, na.rm=TRUE)
  sums4 <- colSums(bound2, na.rm= TRUE)
  print("************************")
  print(i)
  print(sums3)
  print(".....................")
  print(sums4)
}

## matching PP.2 to PP.2
J2 <- t(PP.2[,4:720])
colz <- c(1:31)
for (i in colz){
  m2 <-ifelse(J2[,1]==0, 0,ifelse(J2[,1]==J2[,i], T, NA)) ##change number after J1
  sums3 <- sum(m2, na.rm=TRUE)
  sums4 <- colSums(J2, na.rm= TRUE)
  print("************************")
  print(i)
  print(sums3)
  print(".....................")
  print(sums4)
}

#########################################################################
##########
## matching stool.3 to PP.3
st.3.t <- t(St.3[,4:720])
J3 <- t(PP.3[,4:720])
bound3 <- cbind(st.3.t, J3) ##1
colz <- c(1:42) ## creating a vector
for (i in colz){
  m3 <-ifelse(bound3[,1]==0, 0,ifelse(bound3[,1]==bound3[,i], T, NA))
  sums5 <- sum(m3, na.rm=TRUE)
  sums6 <- colSums(bound3, na.rm= TRUE)
  print("************************")
```

```
  print(i)
  print(sums5)
  print(".....")
  print(sums6)
}

## matching PP.3 to PP.3
J3 <- t(PP.3[,4:720])
colz <- c(1:41)
for (i in colz){
  m3 <-ifelse(J3[,1]==0, 0,ifelse(J3[,1]==J3[,i], T, NA)) ##change number after J1
  sums5 <- sum(m3, na.rm=TRUE)
  sums6 <- colSums(J3, na.rm= TRUE)
  print("**********")
  print(i)
  print(sums5)
  print(".....")
  print(sums6)
}



new <- trial3[,colSums(trial3 != 0, na.rm = TRUE) > 0] ## eleminating rows with zeros to
find total OTUs
new2 <- sub3[,colSums(sub3 != 0, na.rm = TRUE) > 0]



##venn diagram
install.packages('VennDiagram')
library(grid)
library(VennDiagram)
draw.single.venn(area = 29, category = "Stool", fill = "light blue", alpha = 0.5)

draw.pairwise.venn(area1 = 58, area2 = 152,  cross.area = 50, category = c("St", "MC"),
 fill = c("red", "lightgrey"))
grid.newpage()
draw.pairwise.venn(area1 = 9, area2 = 47,  cross.area = 8, category = c("St2", "MC2"),
 fill = c("red", "lightgrey"))
grid.newpage()
```

```
draw.pairwise.venn(area1 = 37, area2 = 87,  cross.area = 25, category = c("St3", "MC3"),
 fill = c("red", "lightgrey"))
grid.newpage()
draw.pairwise.venn(area1 = 58, area2 = 187,  cross.area = 52, category = c("Total Stool",
 "Total Plate Pools "),lty = "blank", fill = c("skyblue", "pink1"))



##Viewing what OTUs are presented in each stool sample
St.1[, colSums(St.1 != 0, na.rm = TRUE) > 0] ##OTUs in stool 1
St.2[, colSums(St.2 != 0, na.rm = TRUE) > 0]##OTUs in stool 2
St.3[, colSums(St.3 != 0, na.rm = TRUE) > 0] ##OTUs in stool 3

##Viewing what OTUs are presented in each individual plate pool
PP.1 <- trial3[trial3$Sample %in% "B",]
JJ.1 <- PP.1[PP.1$condition %in% "ANA",]
JAA <- JJ.1[JJ.1$media %in% "CNA",]
JAA[, colSums(JAA != 0, na.rm = TRUE) > 0]
initial[,113]

##finding exactly which OTUs are represented by all PPs
mat <- PP.1[,4:720]
i <- (colSums(mat, na.rm=T) != 0)
matnonzero <- mat[, i] ##drops columns w a colsum of zero



## sample one overlap with top 10 picks
draw.pairwise.venn(area1 = 37, area2 = 66,  cross.area = 25, category = c("St3",
 "Top15"), fill = c("red", "lightgrey"))
grid.newpage()

grepl("Propionibacteriaceae", initial$Consensus.Lineage)
```

A2. R scripts used for Aim 2 and 3 data analysis and figures

```
## installing packages
install.packages("ggplot2")
install.packages("vegan")
install.packages("knitr")
install.packages("reshape2")
install.packages("permute")
```

```
install.packages("lattice")

setwd("~/Google Drive/Stearns_lab_shared/Illumina/Luna/")
library(lattice)
library(permute)
library(vegan)
library(ggplot2)
library(grid)
#library(faraway)
library(reshape2)
#library(lme4)
#library(effects)
options(scipen=999) #get rid of scientific notation


##OTU data
## Sweep instead of / for creating a relative abundance table [post by B.
 Bolker](http://rpubs.com/bbolker/sweep_divide)

meta_table <- read.csv("map.csv", row.names = 1)

OTU <-read.csv("SPE_OTU.csv",row.names=1,check.names=FALSE)

OTU.s <-t(OTU)
meta.OTU.s <- meta_table[row.names(OTU.s),]
merged.s <- data.frame(meta.OTU.s, OTU.s)
Four <- merged.s[merged.s$PID %in% "4",] # table with all kid 4
Pl1 <- Four[Four$Condition %in% "AER",]
Pl2 <- Four[Four$Condition %in% "ANA",]
PlP <- rbind(Pl1, Pl2) # table with kid 4 Plate pools
Four.st <- Four[Four$Condition %in% "ST",] # table with kid 4 stool
Four.O <- Four.st[order(Four.st$Day), c(1:1744)] # order by day
row.names(Four.O) <- Four.O$Day #make rownames same as days

Day <- Four.O[,3]

# Calculate species richness
N <- rowSums(Four.O[,5:ncol(Four.O)])
min(N)

S <- specnumber(Four.O[,5:ncol(Four.O)])

S.rar <-rarefy(Four.O[,5:ncol(Four.O)], min(N))
```

```
species.richness <-rarefy(Four.O[,5:ncol(Four.O)], min(N)) ## use this or S.rar

S.met <- data.frame(Day,species.richness)

newnames <- S.met$Day
newnames <- gsub(pattern = "Day",replacement = "", x = newnames) ## how to replace
 things
S.met$Day <- newnames
row.names(S.met) <- S.met$Day ## those turn factors into numbers so i can line plot
 them


# Calculate community diversity (richness+ evenness)
community.diversity <- diversity(Four.O[,5:ncol(Four.O)])
summary(community.diversity)
D.met <- data.frame(Day,community.diversity)

newnames <- D.met$Day
newnames <- gsub(pattern = "Day",replacement = "", x = newnames) ## how to replace
 things
D.met$Day <- newnames

row.names(D.met) <- D.met$Day

[9:45]
noIPA is the dataframe and in mine the OTU data started in column 24
the otu table you use should has to include singletons etc...
vegan package btw


#This is the pcoa script  [beta diversity measure]
OTU.p <-t(OTU.r) ## relative abundance used here
meta.OTU.p <- meta_table[row.names(OTU.p),]
merged.p <- data.frame(meta.OTU.p, OTU.p)
PC <- merged.p[merged.p$PID %in% "4",] # table with all kid 4
PC.st <- PC[PC$Condition %in% "ST",] # table with kid 4 stool
PC.O <- PC.st[order(PC.st$Day), c(1:1744)] # order by day
row.names(PC.O) <- Four.O$Day

bray3d <- vegdist(PC.O[,5:ncol(PC.O)], method = "bray")
groups3d <- PC.O$Day
mod3d <- betadisper(bray3d, groups3d)
plot(mod3d,main=NULL, ellipse = TRUE, hull = FALSE, label =TRUE)
dev.off()
```

```
## beta diversity between ALL samples (1,2,3,4,5)
OTU.p <-t(OTU.r) ## relative abundance used here
meta.OTU.p <- meta_table[row.names(OTU.p),]
merged.p <- data.frame(meta.OTU.p, OTU.p)
KIDS <- merged.p[merged.p$Condition %in% "ST",]

bray3d <- vegdist(KIDS[,5:ncol(PC.O)], method = "bray")
groups3d <- KIDS$Day
mod3d <- betadisper(bray3d, groups3d)
plot(bray3d,main=NULL, ellipse = TRUE, hull = FALSE, label =TRUE)

#print pdf
pdf("d3_PCoA.pdf", height = 4, width = 4, useDingbats = FALSE)
plot(mod3d,main=NULL, ellipse = TRUE, hull = FALSE, label = TRUE)
dev.off()


## line plot alpha diversity (comm diversity and species richness)
plot(S.met$Day, S.met$species.richness, type = "o", las=2)

plot(D.met$Day, D.met$community.diversity, type = "o", las=2)

boxplot(D.met$community.diversity~D.met$Day, las=2)

boxplot(D.met$community.diversity~D.met$Day,data=D.met, main="Community
 Diversity overtime",
        xlab="Day", ylab="Community Diversity")
## barchart for community diversity and species richness
p<-ggplot(data=D.met, aes(x=Day, y=community.diversity)) +
  geom_bar(stat="identity")
p

m<-ggplot(data=S.met, aes(x=Day, y=species.richness)) +
  geom_bar(stat="identity")
m

## find taxa of interest in data

SS1 <- Four.O[Four.O$Day %in% "Day05",]
SS2 <- Four.O[Four.O$Day %in% "Day25",]
SSS <- rbind(SS1, SS2) ## make table with only stool samples from day 5, 25

grepl("Caulobacteraceae", colnames(PlP)) # gives true or false value
```

```
Bacteria <- SSS[ , grepl( "Alphaproteobacteria" , names( SSS ) ) ] ## extract columns
 with certain bacteria
View(Bacteria)
Bacteria <- PlP[ , grepl( "Alphaproteobacteria" , names( PlP ) ) ]
View(Bacteria)

##heatmap for PP Vs. ST representation of taxa of interest (for longitudinal)
chart <-read.csv("16S vs PP.csv",row.names=1,check.names=FALSE)
chart <- data.matrix(chart)
chart <- chart[complete.cases(chart[,]),]
red <- brewer.pal(7,"RdPu")
pdf("STOOL VS PP",height=10,width=6)
heatmap <- heatmap(chart, Colv=NA, col = red, scale="none")
dev.off()

##heatmap for PP Vs. ST representation of taxa of interest (for ALL 12345)
ALL1 <- merged.s[merged.s$PID %in% "1",] #sample 1
ALL2 <- merged.s[merged.s$PID %in% "2",] #sample 2
ALL3 <- merged.s[merged.s$PID %in% "3",] #sample 3
ALL <- rbind(ALL1, ALL2, ALL3) # combine into 1 table

Pl3 <- ALL[ALL$Condition %in% "AER",]
Pl4 <- ALL[ALL$Condition %in% "ANA",]
PlPS <- rbind(Pl3, Pl4) # table with kid 1,2,3 Plate pools
ALL.st <- ALL[ALL$Condition %in% "ST",] # table with kid 1,2,3 stool

Bacteria <- ALL.st[ , grepl( "Alcaligenaceae" , names( ALL.st ) ) ] ## extract columns
 with certain bacteria
View(Bacteria) ## For stool
Bacteria <- PlPS[ , grepl( "Alcaligenaceae" , names( PlPS ) ) ]
View(Bacteria) ## For PP

grepl("Peptostreptococcaceae", colnames(PlPS)) # gives true or false value

colnames(PlPS) ## list all the column names

chart <-read.csv("16sppall.csv",row.names=1,check.names=FALSE)
chart <- data.matrix(chart)
chart <- chart[complete.cases(chart[,]),]
red <- brewer.pal(7,"RdPu")
pdf("STOOL VS PP ALL SAMPLES",height=10,width=6)
heatmap <- heatmap(chart, Colv=NA, col = red, scale="none")
dev.off()
```

```
##heatmap for PP Vs. ST representation of taxa of interest (for 123 only)
chart <-read.csv("Samples123 16S vs PP.csv",row.names=1,check.names=FALSE)
chart <- data.matrix(chart)
chart <- chart[complete.cases(chart[,]),]
red <- brewer.pal(7,"RdPu")
pdf("STOOL VS PP ALL SAMPLES",height=10,width=6)
heatmap <- heatmap(chart, Colv=NA, col = red, scale="none")
dev.off()

##grouping OTUs based on relative abdunaces
   ## used colnames to find which OTU is associated with which colnumber
AB <- NTT[, c(1, 6, 18)] #0. values
AC <- NTT[, c(1, 3, 5, 7, 8, 11, 16, 19, 22, 23, 24, 27, 30, 33, 35 )] #0.0 values
AD <- NTT[, c(1, 4, 9, 10, 26, 28, 29, 31, 37, 39, 43, 48, 49, 53, 60)] #0.00
AE <- NTT[, c(1, 2, 12, 17, 20, 32, 46, 52, 54, 55, 63, 65, 67, 69, 70, 76, 77, 81, 83, 97,
 130, 169)] #0.000
AF <- NTT[,-c(6, 18, 3, 5, 7, 8, 11, 16, 19, 22, 23, 24, 27, 30, 33, 35, 4, 9, 10, 26, 28, 29,
 31, 37, 39, 43, 48, 49, 53, 60, 2, 12, 17, 20, 32, 46, 52, 54, 55, 63, 65, 67, 69, 70, 76, 77,
 81, 83, 97, 130, 169)] #0.0000
library(ggplot2)
install.packages("tidyverse")
library(tidyverse)

AC_long <- gather(AC, Microbes, Relative_Abundance, -Day) #organize microbes into 1
 column
View(AD_long)
AF_plot <- ggplot(AF_long, aes(x= Day, y= Relative_Abundance, colour= Microbes)) +
 geom_line()+ theme_bw()+
  geom_point(size=0)+ theme(legend.text=element_text(size=8))+ guides(colour =
 guide_legend(override.aes = list(size=3,linetype=0)))  #plot line
AF_plot ## when you save make sure to adjust the height/width

pdf("AD_plot.pdf", height = 4, width = 4, useDingbats = FALSE)

ggplot(AF_long, aes(x = Day, y = Relative_Abundance, fill = Microbes)) +
  geom_bar(stat = "identity") # plot stacked bar chart

## plotting which are rising/falling/plateau
Pat <- read.csv("pattern.csv") # assign pattern- R:rise, F:fall, P:plateau
ACpat <- cbind(AC_long, Pat) #bind with previous table
ACpat_plot <- ggplot(ACpat, aes(x= Day, y= Relative_Abundance, colour= Microbes)) +
 geom_line()+ theme_bw()+ facet_wrap("Pattern")+ geom_point(size=0)+
```

```
 theme(legend.text=element_text(size=8))+ guides(colour = guide_legend(override.aes =
 list(size=3,linetype=0)))  #plot line
ACpat_plot ## when you save make sure to adjust the height/width

Pat2 <- read.csv("pattern2.csv")
ADpat <- cbind(AD_long, Pat2) #bind with previous table
ADpat_plot <- ggplot(ADpat, aes(x= Day, y= Relative_Abundance, colour= Microbes)) +
 geom_line()+ theme_bw()+ facet_wrap("Pattern")+ geom_point(size=0)+
 theme(legend.text=element_text(size=8))+ guides(colour = guide_legend(override.aes =
 list(size=3,linetype=0)))  #plot line
ADpat_plot ## when you save make sure to adjust the height/width

Pat3 <- read.csv("pattern3.csv")
AEpat <- cbind(AE_long, Pat3) #bind with previous table
AEpat_plot <- ggplot(AEpat, aes(x= Day, y= Relative_Abundance, colour= Microbes)) +
 geom_line()+ theme_bw()+ facet_wrap("Pattern")+ geom_point(size=0)+
 theme(legend.text=element_text(size=8))+ guides(colour = guide_legend(override.aes =
 list(size=3,linetype=0)))  #plot line
AEpat_plot ## when you save make sure to adjust the height/width

Pat4 <- read.csv("patterns4.csv")
ABpat <- cbind(AB_long, Pat4) #bind with previous table
ABpat_plot <- ggplot(ABpat, aes(x= Day, y= Relative_Abundance, colour= Microbes)) +
 geom_line()+ theme_bw()+ facet_wrap("Patterns")+ geom_point(size=0)+
 theme(legend.text=element_text(size=8))+ guides(colour = guide_legend(override.aes =
 list(size=3,linetype=0)))  #plot line
ABpat_plot ## when you save make sure to adjust the height/width

## build me a table where all columns with zero value for day5 are extracted
NTS <- NTT[NTT$Day %in% "5",] ## data for day 5 only
NTS <- NTS[,2:ncol(NTS)]
NTS2 <- NTS[,-(which(colSums(NTS) == 0))] #all things in st at day 5
NTM <- NTT[NTT$Day %in% "25",] ## data for day 25 only
NTM <- NTM[,2:ncol(NTM)]
NTM2 <- NTM[,-(which(colSums(NTM) == 0))] #all things in st at day 25

UM <- rbind(NTS, NTM) ## table with both day 5 and 25
UMM <- UM[,-(which(colSums(UM) == 0))] ## drop columns with 0 value for both day
 5 and 25


UM5 <- UMM[,-(which(colSums(UMM) > UMM[2,]))] #keep values where day5= 0 but
 25 isn't
   ## 33 things
```

```
UM25 <- UMM[,-(which(colSums(UMM) > UMM[1,]))] #keep values where day25= 0
 but 5 isn't
    ## 58 things
  # now find the plate pools and whether they have those things that were zeros in UM5
 and 25
Plts <- kidf[kidf$Condition %in% "ANA",]
Pltt <- kidf[kidf$Condition %in% "AER",]
Plt <- rbind(Plts, Pltt)
plt <- Plt[,4:ncol(Plt)]
Dayy <-
 c(5,5,5,5,5,5,5,5,5,5,5,5,25,25,25,25,25,25,25,25,25,25,25,25,25,5,25,25,25,25,25,25,25,
 25,25,25,25)
pltt <- data.frame(Dayy, plt)
D5 <- pltt[pltt$Dayy %in% "5",] # organisms at day 5
D5f <- D5[,-(which(colSums(D5) == 0))]
  ##25 of the things in previous list captured. 8 not
D25 <- pltt[pltt$Dayy %in% "25",] # organisms at day 25
D25f <- D25[,-(which(colSums(D25) == 0))]
  ##27 of the things in previous list captured, rest are not

##for plate pools
matchingNames <- names(D5f)[names(D5f) %in% names(D25f)] ## OTU overlap
 between day 5 and 25
D5f[, matchingNames]
D25f[, matchingNames]
matchingNames

## for 16S stool alone

match <- names(NTS2)[names(NTS2) %in% names(NTM2)] ## OTU overlap between
 day 5 and 25
match


## poster pic [for both day 5 and 25]

D5ff <- D5f[,2:ncol(D5f)] ## Day 5
D5ff <- D5ff[, c(1, 14, 28, 42, 43, 59, 62, 70, 73, 77, 87, 92, 95, 104, 110, 120, 122, 123,
 148, 161, 175, 176, 187, 194, 218)]
Mic <- read.csv("poster.csv")
Mic2 <- data.frame(Mic, D5ff)

m <- colnames(Mic2)
y <- gsub(".*o__","",m)
```

```
colnames(Mic2) <- y

Mic_long <- gather(Mic2, Microbes, Relative_Abundance, -Media_Condition)
Mic_plot <- ggplot(Mic_long, aes(x = Media_Condition, y = Relative_Abundance, fill =
 Microbes)) +
  geom_bar(stat = "identity") + theme(legend.text=element_text(size=7))+ guides(colour
 = guide_legend(override.aes = list(size=3,linetype=0)))
 +theme(axis.text=element_text(size=12),
        axis.title=element_text(size=14,face="bold"))
Mic_plot

# enter heatmap here for day 5

write.csv(Mic2, "Day5.csv")

library(RColorBrewer)
chart <-read.csv("Day5.csv", row.names = 1, check.names=FALSE)
chart <- data.matrix(chart)
chart <- chart[complete.cases(chart[,]),]
red <- brewer.pal(7,"Oranges")
pdf("day5.heatmap",height=10,width=8)
heatmap <- heatmap(chart, Colv=NA, col = red, scale="none")
dev.off()


D25ff <- D25f[, c(16, 23, 41, 48, 52, 57, 65, 75, 82, 85, 98, 108, 127, 131, 137, 144, 150,
 155, 176, 178, 198, 200, 235, 258, 374)] ## Day 25
Micc <- read.csv("poster2.csv")
Micc2 <- data.frame(Micc, D25ff)

m <- colnames(Micc2)
y <- gsub(".*o__","",m)
colnames(Micc2) <- y

Micc_long <- gather(Micc2, Microbes, Relative_Abundance, -Media_Condition)
Micc_plot <- ggplot(Micc_long, aes(x = Media_Condition, y = Relative_Abundance, fill
 = Microbes)) +
  geom_bar(stat = "identity") + theme(legend.text=element_text(size=8))+ guides(colour
 = guide_legend(override.aes = list(size=3,linetype=0)))
 +theme(axis.text=element_text(size=12),

      axis.title=element_text(size=14,face="bold"))
Micc_plot
```

```
# enter heatmap here for day 25

write.csv(Micc2, "Day25.csv")

library(RColorBrewer)
chart <-read.csv("Day25.csv", row.names = 1, check.names=FALSE)
chart <- data.matrix(chart)
chart <- chart[complete.cases(chart[,]),]
red <- brewer.pal(7,"Oranges")
pdf("day250.heatmap",height=10,width=8)
heatmap <- heatmap(chart, Colv=NA, col = red, scale="none")
legend("topright",legend=c("Basal","LumA","LumB","Her2","Claudin","Normal","","Po
 sitive","Negative","NA","","Targeted","Chemo","","Approved","Experimental"),
      fill=c("red","blue","cyan","pink","yellow","green","white","black","white","grey","w
 hite","darkorchid","darkred","white","green","darkgreen"), border=FALSE, bty="n",
 y.intersp = 0.7, cex=0.7)
dev.off()


## Comparison between OTUs found in ST and in PP
NTS <- NTT[NTT$Day %in% "5",] ## data for day 5 only
NTS <- NTS[,2:ncol(NTS)]
NTM <- NTT[NTT$Day %in% "25",] ## data for day 25 only
NTM <- NTM[,2:ncol(NTM)]

UFF <- NTS[,-(which(colSums(NTS) == 0))] #exclude columns with zero for day 5 ST
UFM <- NTM[,-(which(colSums(NTM) == 0))] #exclude columns with zero for day 25
 ST

DA5 <- D5f[,2:ncol(D5f)] ## this chunk is just dealing with PP day 5 and 25
m <- colnames(DA5)
y <- gsub(".*o__","",m)
colnames(DA5) <- y
DA25 <- D25f[,2:ncol(D25f)]
m <- colnames(DA25)
y <- gsub(".*o__","",m)
colnames(DA25) <- y
      ## number of OTUs in ST day 5= 108; in PP day5= 432 (intersect= 79)
      ## number of OTUs in ST day 25= 83; inn PP day25= 533 (intersect= 75)

D <- colnames(UFF)
P <- colnames(DA5)
x <- intersect(D, P) ##intersect OTUs from stool day 5 with PP5
```

```
D1 <- colnames(UFM)
P1 <- colnames(DA25)
x1 <- intersect(D1, P1) ##intersect OTUs from stool day 25 with PP25

draw.pairwise.venn(area1 = 108, area2 = 432,  cross.area = 79, category = c("St5",
 "PP5"), fill = c("red", "lightgrey"))
## ^ overlap between stool and PP on day 5
grid.newpage() ## new page
draw.pairwise.venn(area1 = 83, area2 = 533,  cross.area = 75, category = c("St25",
 "PP25"), fill = c("red", "lightgrey"))
```