THE INTRODUCTION OF SOLID FOODS AND THE INFANT GUT MICROBIOME

MCMASTER UNIVERSITY

Department of Pediatrics

EVALUATING THE RELATIONSHIP BETWEEN DIETARY INTAKE AT THE TIME IMMEDIATELY BEFORE AND AFTER THE INTRODUCTION OF SOLID FOODS AND THE GUT MICROBIOME IN FULL-TERM INFANTS: A LONGITUDINAL STUDY

By CHIARA-MARIA HOMANN, BSc

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree of Master of Science

McMaster University **MASTER OF SCIENCE (2020)**, Hamilton, Ontario (Medical Sciences)

Title: Evaluating the Relationship Between Dietary Intake at the Time Immediately Before and After the Introduction of Solid Foods and the Gut Microbiome in Full-Term Infants: A Longitudinal Study

Author: Chiara-Maria Homann, BSc (Technical University of Munich)

Student number: 400203930

Supervisor: Dr. Katherine Morrison

Committee Members: Dr. Jennifer Stearns, Dr. Russell de Souza, Dr. Eileen Hutton

Number of Pages: xxiii, 213

Lay Abstract

The introduction of solid foods is an important life-event during infancy. This is also when the gut microbiome is developing to its mature state. Since nutrition is an important factor influencing the microbiome, investigating the dietary choices at the introduction to solid foods is the aim of the following study. Here, daily stool samples and food diary entries were collected for 15 healthy, breast-fed infants. It is important to measure the diversity of the bacteria in the gut of an individual (alpha) and between people (beta), as well as bacteria present. Carbohydrates drive the change in alpha diversity, especially fiber. Feeding infants a diet with many different foods shows increased alpha diversity and change in the microbiome immediately after introduction. Interestingly, the infant gut microbiome reacts to fiber in a manner comparable to the adult gut microbiome, i.e. increased bacterial diversity, which is associated with better health outcomes in adults.

Abstract

Background: The introduction of solid foods is an important dietary event during infancy and is associated with a time of dramatic shifts in gut microbial composition. The influence of solid food introduction on gut bacterial dynamics remains understudied.

Methods: 15 healthy, full-term, vaginally born, and breast-fed infants of the *Baby, Food and Mi* sub-study of the *Baby & Mi Study* were investigated. Caregivers were asked to collect daily stool samples and food diaries for 17 days, commencing three days prior to the introduction of solids. Additional stool samples were available up to one year as part of the *Baby and Mi* study. The exposure of interest, nutritional patterns, was analyzed using food composition output from ESHA's Food Processor. The number of food items and food groups introduced were used to calculate dietary diversity scores. The outcome of interest, gut bacterial dynamics, was analyzed using RStudio.

Results: The mean (SD) age at the introduction of solid foods is 5.5 (0.66) months (n = 15). Over the study period, the proportion of estimated energy intake from solid foods was low (7.5%; SD 6.74%) (n = 14[1]). Alpha diversity increased over time and was highest at 1 year. The gut microbial community influenced by dominant bacterial taxa changed with increasing age. With introduction of solids, individual community composition changed, though to a varying extent. Shannon

[1] Nutritional data is only available for 14 infants.

alpha diversity was directly associated with calories from carbohydrates, particularly daily fiber intake. The infant's dietary diversity score was directly associated with alpha diversity and was also positively associated with the degree of change occurring in this time period.

Conclusion: Fiber intake and the dietary diversity scores had the closest relationships to the gut microbiome's alpha diversity and community structure in infants at the time of solid food introduction.

Acknowledgements

I would first like to thank my supervisor Dr. Katherine Morrison (McMaster University, Department of Pediatrics) for the opportunity to complete my Master of Science in her research lab, and for all the guidance I received over the past two years. She consistently allowed this paper to be my own work but steered me in the right the direction whenever she thought I needed it.

I would also like to thank Dr. Jennifer Stearns of my committee (McMaster University, Department of Medicine) for guiding and advising me through the process of getting to know the microbiome world, and for supporting the analyses in this field. Without their valuable input, this thesis would not have been possible.

I would also like to acknowledge Dr. Russell de Souza (McMaster University, Department of Health Research Methods, Evidence, and Impact) for his invaluable knowledge about human nutrition and statistical methods for analysis of the dietary data.

Additionally, I would like to thank Dr. Eileen Hutton (McMaster University, Department Obstetrics & Gynecology, McMaster University) for providing holistic support to the project.

Lastly, I would like to thank Jenifer Li, of the *Baby & Mi* study, as well as all my fellow lab mentors for the support, especially Efrah Yousuf, for answering all my R-related questions, despite being very busy herself with her PhD.

Contents

List of Figures

List of Tables

Abbreviation Index

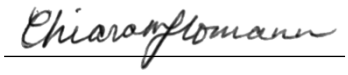
Abbreviation	Meaning
AIC	Akaike Information Criterion
ASV	Amplicon sequence variants
BLW	Baby-Led Weaning
BMI	Body mass index
CRF	Case report form
C-section	Caesarean section
d	Day
DADA2	Divisive Amplicon Denoising Algorithm 2
DNA	Deoxyribonucleic acid
FAO	Food and agriculture organization of the United Nations
FC	Calories from fat and carbohydrates
FOS	Fructooligosaccharides
g	Gram
GBS	Group B streptococcus
HEI	Healthy eating index
HFD	High-fat diet
HMF	Human MicroFlora
HMOs	Human milk oligosaccharides
IAP	Intrapartum antibiotic prophylaxis
IGF-1	Insulin-like growth factor 1
IQR	Interquartile range
IV	Intravenous
kcal	Calories
kg	Kilogram
NICU	Neonatal intensive care unit
OGCT	Oral glucose challenge test
OGTT	Oral glucose tolerance test
OTU	Operational taxonomic unit
PC	Calories from protein and carbohydrates
PCoA	Principal coordinate analysis
PCR	Polymerase chain reaction
PD	Phylogenetic diversity
PERMANOVA	Permutational analysis of variance

PID	Participant ID
PF	Calories from protein and fat
rRNA	Ribosomal ribonucleic acid
RT-PCR	Reverse transcription polymerase chain reaction
SCFA	Short-chain fatty acids
SD	Standard deviation
SDS	Standard deviation score
SE	Standard error
USDA	United States department of agriculture
WHO	World Health Organization

Declaration of Authorship

I, Chiara-Maria Homann, hereby declare and confirm that the master thesis "Evaluating the Relationship Between Dietary Intake at the Time Immediately Before and After the Introduction of Solid Foods and the Gut Microbiome in Full-Term Infants: A Longitudinal Study" has been my independent work, without help from others, and without using anything other than the named sources and aids. This work has not been submitted for any other degree.

Hamilton, 18th of June 2020



Chiara-Maria Homann

1. Introduction

The gut microbiome refers to the bacterial ecosystem in the human gastrointestinal tract that consists of trillions of microbes (Cresci and Bawden, 2015), which has a symbiotic relationship with the human host via metabolic, immunological and nutritional functions (Jandhyala *et al.*, 2015). Disturbances in the gut microbiome have been associated with numerous pathological states, such as obesity and atopy (Turnbaugh and Gordon, 2009), which underlines the importance of a healthy gut microbiome.

Despite the growing body of evidence suggesting the fundamental importance of the gut microbiome and human health, the development of the infant gut microbiome remains understudied. One of the most important events during infancy is the introduction of solid foods (Fallani *et al.*, 2011; Zimmer *et al.*, 2012). The introduction of solid foods initiates the shift towards the more adult-like composition expected at three years of age, as this period may cause dramatic shifts in the composition of the gut microbiota, due to changing ratios of fat, protein, carbohydrate and fiber content (Johnson and Versalovic, 2012). Few studies have investigated early life nutritional choices, apart from breastfeeding versus formula feeding, to date. Therefore, the focus of this thesis is to evaluate the relationship between nutritional choices at the time of introduction to solid foods and gut bacterial dynamics in a cohort of full-term, vaginally born, and healthy infants.

1.1. Nutrition

1.1.1. Introduction of Solid Foods

Infant nutrition is of critical importance due to its impact on growth and development (Organization, 2018) and the influence of early eating patterns on later eating habits and food preferences (Betoko *et al.*, 2013). The World Health Organization (WHO) recommends that breastfeeding is a key component of infant nutrition and encourages continuation of breastfeeding until the age of two years (Organization, 2018). Despite this recommendation, only 34.8% of infants worldwide are exclusively breastfed up to 6 months of age (Organization, 2018), and 65% of all infants worldwide continue to breastfeed until 2 years of age (Unicef, 2019).

One of the most important events during infancy is the introduction of solid foods (Fallani *et al.*, 2011; Zimmer *et al.*, 2012), defined as the time where at least one feeding of breast milk or formula is intentionally replaced by solid foods. Currently, the WHO recommends introducing solid foods at six months of age, as this is when the infant's nutritional and energy requirements begin to exceed what is provided by the mother's breast milk (Organization, 2018). The most important nutrient that drives the need for introduction of solid foods is iron. At six months of age, the iron stores from intrauterine life are depleted and breast milk does not provide sufficient amounts of iron, with 0.03 mg/ 100 g (USDA, 2018). Additionally, at six months of age the gastrointestinal tract has matured to facilitate the digestion

of solid foods (Tandoi *et al.*, 2017). Another consideration for the timing of the introduction of solid foods is the developmental readiness of the infant. Signals of readiness include adequate head and neck control, the ability to sit upright when supported, being interested in food eaten by others, wanting to put things in their mouth and appearing to have an increased appetite (Marks, 2015). Introducing solid foods later than six months could have detrimental effects on the linear growth of the infants, and could lead to nutrient deficiencies (White *et al.*, 2017). Globally, 30% of children are introduced to solid foods before six months of age, while 42% of children are introduced to solid foods later than the recommended time period (White *et al.*, 2017).

1.1.2. Recommendations for the Introduction of Solid Foods

The WHO recommends that 200 kcal/d should come from solid foods from 6 – 8 months of age (Organization, 2002). First foods should be iron-rich since the increased need for iron drives the necessity of introducing solid foods. This includes iron-fortified cereals, meat, fish, chicken, and legumes (Marks, 2015). The foods introduced should have a variety of flavors and textures (Fewtrell *et al.*, 2017), however, these new flavors should be introduced gradually (Tandoi *et al.*, 2017). Nutritional programming from the intrauterine environment (i.e. amniotic fluid), as well as breast milk might facilitate the introduction of family foods (Tandoi *et al.*, 2017), as certain flavors may already be known to the infant. Foods that are

high in fat, salt, and sugar, or low in nutritional value should be avoided when introducing solid foods (Wang *et al.*, 2019).

These guidelines are in contrast to those from the European Society for Pediatric Gastroenterology, Hepatology, and Nutrition. They recommend exclusive breastfeeding up to 4 months of age and predominant breastfeeding up to 6 months, with the introduction of complementary foods between 4 and 6 months. Other considerations related to solid food introduction relate to the timing of introduction of potentially allergenic foods (Fewtrell *et al.*, 2017) and vegan diets are discouraged unless it is medically necessary, and this diet should be overseen by a medical professional. Generally, there are a variety of guidelines to follow when introducing solid foods, but the consensus is that breastfeeding remains important even as solid foods are introduced, and that solid foods should be introduced around 6 months of age. In addition to timing of solid food introduction, the specific nutrients that a child requires for normal growth and development at different ages are an important consideration.

1.1.3. Protein Requirements

Protein requirements are determined based on the understanding of the minimum intake of high-quality protein needed to maintain appropriate body composition for infants of specific ages and to support growth at normal rates. Protein is required for infant growth, which is fastest in the first months of life. As

a result, protein requirements/kg body weight are 75% higher in the first six months of life than they are for adults. Growth slows down rapidly from 6 – 24 months, and the protein requirements also decrease (Garlick, 2006). Interestingly, nutritional guidelines for infants and young children vary by country (Koletzko and Hermoso, 2008). Breast milk intake is utilized as the model of prime protein uptake in the calculation of safe levels and average requirements of protein for different age groups (Garlick, 2006). In Germany, Switzerland and Austria, 2.0 - 2.22 g/kg/d of protein are recommended from birth to under four months of age, 1.2 to 1.6 g/kg/d from four to under one year of age and 1.2 g/kg/d from one to four years of age. The trend that protein requirements decrease per kg/d with increasing age is also seen in the guidelines from other countries (Koletzko and Hermoso, 2008). Guidelines from the World Health Organization are summarized in *Table 1*.

Table 1: Nutritional guidelines from the WHO for energy and protein intake (Koletzko and Hermoso, 2008).

Age	**Energy (kcal/d)**	**Protein (g/d)**	**% Energy**
3-6 months	700	13	7.6
6-9 months	810	14	7.1
9-12 months	950	14	6.0
1-2 years	1150	13.5	4.8
2-3 years	1350	15.5	4.7

Recently, the practice of utilizing protein as the key variable in defining infant nutrition has been called into question due to the possibility that high protein intake

can lead to infants with overweight or obesity. On average, breast milk has 0.9 – 1.2 g/dL of protein (Ballard and Morrow, 2013) and formula can have up to two to three times that amount of protein (Haschke *et al.*, 2016). Infants consuming higher protein formula have significantly higher BMIs and higher body weight, up to six years later (Koletzko *et al.*, 2016; Öhlund *et al.*, 2010; Weber *et al.*, 2014). Furthermore, higher protein intake is also associated with higher fat mass in addition to higher BMI (Voortman *et al.*, 2016). Interestingly, animal protein and not vegetable protein have been linked to higher BMIs and weight at six years in one study (Voortman *et al.*, 2016). It is proposed that eating more protein leads to higher levels of circulating branched-chain amino acids in the bloodstream which in turn induces higher insulin secretion and IGF-1 secretion, which affects the beta-oxidation of fatty acids in the mitochondria and ultimately results in higher levels of fat deposition (Hellmuth *et al.*, 2016; Voortman *et al.*, 2016). Overall, protein is an important nutrient for infant growth, however, introducing too much protein may have detrimental effects on infant body composition and body composition later in life.

1.1.4. Fat Requirements

In contrast to adults, infants need to obtain a higher proportion of their caloric intake from fat. From 0 – 6 months, 40 - 60% of the infant's caloric intake should come from dietary fats and from six months to three years of age, this

declines to 30 - 40% of energy intake (Canada, 2006; Uauy and Dangour, 2009). The high demand for energy from fat is fulfilled by breast milk and most formulas, as half of their energy content is from fat. Lipids must be ingested because endogenous lipid synthesis is quite low in infants. Apart from use as an energy source, dietary fatty acids are needed for the development of the central nervous system, membranes, and the retina (Uauy and Dangour, 2009). Average amounts of fat in human breast milk are 3.2 – 3.6 g/dL (Ballard and Morrow, 2013). Based on the recommendations for fat intake during infancy, a high proportion of fat is expected and healthy in the infant diet.

1.1.5. Carbohydrate Requirements

Digestible dietary carbohydrates are also a major energy source in the early years of life. Lactose, present in breast milk (Stephen *et al.*, 2012), is particularly important during infancy as a primary carbohydrate source. Average contents of lactose in breast milk are 6.7 – 7.8 g/dL (Ballard and Morrow, 2013). In contrast, a recent survey from the UK in infants aged 6 – 12 months acquired most of their carbohydrates from commercial infant foods, cereal products, and milk (Stephen *et al.*, 2012), demonstrating a decreased intake of breast milk and, therefore, lactose. *Table 2* shows the required amounts of carbohydrates grouped according to age.

Table 2: Requirements of carbohydrate intake for the first three years of life (Canada, 2006)

Age	g/d	% energy/d
0 – 6 months	60	35
7 – 12 months	95	44
1 – 3 years	130	43

Aside from lactose, human breast milk also contains human milk oligosaccharides (HMOs). HMOs are oligosaccharides specific to and very prevalent in human breast milk and they range from 3 to 32 saccharides in length (Ballard and Morrow, 2013). Quantities are similar to that of protein, ranging between 5 and 23 g/L (Zivkovic *et al.*, 2011). Breast milk from different mothers varies in the structure of HMOs; this is genetically predetermined and gives rise to 200 different structures (Ballard and Morrow, 2013; Zivkovic *et al.*, 2011), which have different effects on the infant gut, in terms of bacteria present and gut physiology. HMOs seem to be critical for the healthy growth of infants (Zivkovic *et al.*, 2011) and influence the characteristics of the gut microbiome. Thus, all macronutrients play an important role supporting infant growth and development.

1.1.6. Observed Patterns for the Introduction of Solid Foods

Most studies that have looked at population patterns for the introduction of solid foods have taken place in developing countries, while very few studies investigated

these patterns in developed countries. Since this thesis has a population of infants from a developed country, these results are summarized in more detail below. The National Survey of Early Childhood Health, conducted in the United States of America in 2002, showed that solid foods were most commonly introduced from 4 to 6 months (62%), while 19% of respondents introduced complementary foods before 4 months and another 19% at 7 months or later (Kuo *et al.*, 2011). In a Dutch study (BeeBOFT), solid foods were introduced before four months by 21% of the study population, after five months by 38% and between four and five months by 41% (Wang *et al.*, 2019). A study performed in France, the Epifane study, with 3368 mother-infant dyads compared the actual introduction of solid foods with the French guidelines for solid food introduction. In this population, the introduction of solid foods occurred before 4 months for 13% of the study population and before 6 months for 67% of the population. One study, performed in the developing country Brazil, showed closer adherence to the WHO recommendations. Here, 50% of the participants were introduced to solid foods at 6 months of age. These studies demonstrate that the introduction of solids foods occurs most commonly between 4 and 6 months of age in the countries studied, earlier than the WHO guidelines recommend.

The first foods introduced in the Epifane study were cereals, fruit, and vegetables; fish and meat were introduced after 6 months for 90% of the participants. Cow's milk was only ingested by 26% of the participants during the first year and added fats and eggs were introduced at 12 months for 53% and 23%

of the participants, respectively (Boudet-Berquier *et al.*, 2017). Food products most commonly introduced to Dutch infants prior to four months of age were porridge, fruit, vegetables, and sweet beverages[2]. A study performed in Canada addressed the concern that complementary foods before 6 months may displace breast milk and its nutrients. Their results show that the introduction of solid foods, on average, does not fulfill the recommendations regarding introduction of complementary foods (Friel *et al.*, 2010), in terms of age and caloric intake at the time of introduction.

The most common foods introduced in Brazil were cereals, vegetables, beans, and meat, which contains foods with a higher iron content; fruit was generally introduced earlier at around 5 months of age (Lopes *et al.*, 2018). Generally, foods introduced first are cereals, fruits, and vegetables although this varies somewhat by country.

1.1.7. Strategies for the Introduction of Solid Foods

There are two strategies for the introduction of solid foods, namely the more common traditional method, and the newer baby-led approach, which has become more common over the last 10 to 15 years. In the traditional method, infants are spoon-fed with infant foods, with gradual changes in texture. Finger foods are introduced after 8 months in this approach. Baby-led weaning can begin when the

[2] Sweet beverages in this study: fruit juice, fruit juice concentrate, soft drinks, fruit cordial and sweetened dairy beverages.

infant can self-feed and the infant chooses the type and amount of food. Thus, baby-led weaning generally occurs later, and babies fed this way are more likely to receive whole and family foods and less rice cereal (Brown *et al.*, 2017). These infants usually eat more meat, cow's milk, other dairy products, and sweets, while traditionally fed infants eat more commercial products. In the study by Brown et al., there were no significant differences in macronutrient intakes, with 10% of intake coming from protein, 45% from fat and 45% from carbohydrates. The baby-led approach is associated with higher maternal education level, lower maternal anxiety, older maternal age and the baby's temperament (Brown *et al.*, 2017). In contrast, in a study of 202 infants in New Zealand, baby-led feeding was associated with higher intakes of protein, fat and saturated fat, and similar intakes of carbohydrates, sugar, fiber and energy (Erickson, 2015). The differences between these approaches may explain different proportions of the macronutrients ingested, as BLW infants tend to eat more protein and fats and may be introduced to solid foods later than the traditionally fed infants.

1.2. The Microbiome

1.2.1. The Gut Microbiome

The gut microbiome refers to the bacterial ecosystem of the human gastrointestinal tract that consists of trillions of microbes (Cresci and Bawden, 2015), including bacteria, archaea, fungi, and viruses (Jandhyala *et al.*, 2015). The gut microbiome influences host metabolism in multiple ways including nutrient metabolism, especially of carbohydrates (see **1.2.3. The Gut Microbiome and Nutrition**), xenobiotic and drug metabolism, bile acid metabolism, antimicrobial protection via competition and production of bacteriocins, modulation of the immune system and the permeability of the gut barrier (Jandhyala *et al.*, 2015). Recognition of the multiple functions of the gut microbiome, has led to research to understand the factors influencing this ecosystem Potential influencing factors identified to date include age, geographical location, antibiotic use, diet, stress and genetic predisposition (Cresci and Bawden, 2015; Jandhyala *et al.*, 2015). Gut bacterial characteristics often examined in studies are alpha diversity and beta diversity. Alpha diversity describes intra-individual variability, i.e. how many bacterial taxa are present in the infant gut and how evenly these are distributed. There are several alpha diversity metrics including the Shannon index, bacterial richness, the Chao1 index and the Simpson index. Beta diversity shows the differences in gut microbiome composition between individuals (inter-individual variability) and therefore represents a similarity score between populations

allowing for comparisons, as well as a measure of community structure (Morgan and Huttenhower, 2012) (see section 3: **Methodology** for more details).

1.2.2. The Infant Gut Microbiome

In comparison to the adult gut microbiome, the infant gut microbiome is simpler and less diverse (in terms of alpha diversity), but more unstable, going through many changes over time (Blanton *et al.*, 2016; Turroni *et al.*, 2012). The most prevalent bacterial phyla in infants are Firmicutes and Actinobacteria, followed by Proteobacteria and Bacteroidetes. However, there is high inter-individual variability as evidenced by the observation that the phylum Bacteroidetes is not represented in all infants (Azad *et al.*, 2013). Colonization of the infant's gut occurs by facultative anaerobes early on from the bacterial phyla Firmicutes and Actinobacteria, including *Enterococcus, Escherichia/Shigella, Streptococcus* and *Rothia*. After the oxygen concentration decreases and an anaerobic environment develops, aerotolerant and strictly anaerobic bacteria colonize the gut, such as *Lactobacillus, Bifidobacterium, Collinsella* and *Veillonella* from the phyla Actinobacteria and Firmicutes. As the infants grow older, alpha diversity within infants increases and beta diversity between infants decreases (Backhed *et al.*, 2015) and inter-individual variation is much higher in infants than adults (Palmer *et al.*, 2007). The composition of the microbiome reaches an adult-like profile by the age of three years (Yatsunenko *et al.*, 2012), with a transition away from a

Bifidobacterium-dominated microbiota (Cresci and Bawden, 2015). This change has been linked to the time of weaning from breast milk (Backhed *et al.*, 2015).

Multiple factors are associated with the composition of the infant microbiome. Prenatal impacts are still controversial and mostly unknown; it has been suggested that the intrauterine environment may not be sterile and could influence the infant microbiome prior to birth (Perez-Munoz *et al.*, 2017). Proposed mechanisms for the access of microbes to the intrauterine environment include transport of bacteria from the mother's gut via the bloodstream to the placenta (Matamoros *et al.*, 2013).

Apart from possible prenatal influences, there are three key factors known to influence the infant gut microbiome: delivery mode, early infant feeding, and early exposure to antibiotics. Infants born vaginally have a bacterial composition resembling the mother's vaginal bacterial communities (*Prevotella, Sneathia, Lactobacillus, Escherichia/Shigella, Bacteroides, Bifidobacterium)*, while infants born via cesarean section have a gut microbiome resembling the bacterial communities of the mother's skin (*Enterobacter hormaechei, H. influenza, Staphylococcus, Propionibacterium*) (Backhed *et al.*, 2015; Matamoros *et al.*, 2013). Infants born via cesarean section generally have lower colonization rates, and higher variation than those born vaginally (Penders *et al.*, 2006).

Another important factor influencing the microbiome is type of feeding after birth. Breastfeeding contributes to higher counts of *Lactobacillus* and bifidobacteria. This is thought to occur as these bacteria, as well as some

Bacteroides species, can use human milk oligosaccharides as an energy substrate. Additionally, human breast milk contains live bacteria that could also contribute to colonization (Backhed *et al.*, 2015; Zivkovic *et al.*, 2011). These include species of *Staphylococcus, Streptococcus, Serratia, Pseudomonas, Corynebacterium, Ralstonia,* and *Propionibacterium*, as well as *Bifidobacterium* and *Lactobacillus,* which are members of the bacterial phyla Firmicutes, Actinobacteria and Proteobacteria (Mueller *et al.*, 2015). However, another study showed that dominant bacteria in the milk microbiome were of the phylum Proteobacteria (Pannaraj *et al.*, 2017), describing how the milk microbiome varies by individual and by methodological differences between studies. In a study looking at the vertical transfer of bacteria from breast milk to the infant, the gut microbiome of the infant more closely resembled their own mother's milk microbiome than a random one, and changes in daily milk intake were associated with changes in gut microbial composition (Pannaraj *et al.*, 2017). However, the method of transmission from maternal milk to the infant remains unclear (Mueller *et al.*, 2015). Formula-fed infants have higher gut microbial diversity and richness than breastfed infants (Azad *et al.*, 2013); however, a higher abundance of potentially pathogenic bacteria are also found in the microbiome of these infants, especially *C. difficile* and *E. coli* (Penders *et al.*, 2006).

The third most important factor known to influence the infant gut microbiome is exposure to antibiotics (Mueller *et al.*, 2015) – either intrapartum or after birth. Infants exposed to intrapartum antibiotics as prophylaxis for group B

Streptococcus (IAP; GBS) had different community structure compared to unexposed infants up to 12 weeks of age and the colonization by Firmicutes was delayed. Additionally, alpha diversity metrics were significantly lower in those exposed to IAP (Stearns *et al.*, 2017). In a Finnish study, marked differences in the gut microbiomes of those exposed to IAP and those unexposed were seen up to six months of age, which is longer than what was observed in the study by Stearns and colleagues. Bacteroidetes were more abundant in non-exposed infants, while Firmicutes were less abundant. Postnatal antibiotic use was associated with higher relative abundance of species of *Clostridium* (Tapiainen *et al.*, 2019). Overall, antibiotics have been shown to have a great impact on the still-developing infant gut microbiome, although these early reports suggest the influence may be transient.

Other factors influencing the infant gut microbiome include hospitalization after birth, characteristics of the home environment such as older siblings or pets (Penders *et al.*, 2006), gestational age at birth (Arboleya *et al.*, 2012), prebiotic use (Turroni *et al.*, 2012), geographical location (Yatsunenko *et al.*, 2012), ethnicity (Yatsunenko *et al.*, 2012), and diet at introduction of solid foods (**see 1.2.2.1. Introduction of solid foods and the infant gut microbiome**) (Matamoros *et al.*, 2013).

1.2.2.1. Introduction of Solid Foods and the Infant Gut Microbiome

Although major shifts in the microbial composition have been described in infants when weaning to solid food occurs (Johnson and Versalovic, 2012), little is known of the influence of the dietary composition of the solid foods. In a study by Koenig, et al. a change in the abundance of some bacteria was noted after solid food introduction in one infant. In particular, an increase of Bacteroidetes, bacteria that can degrade plant polysaccharides, was noted. The ingestion of breast milk, and especially HMOs in that breast milk, prior to the introduction of solid foods upregulates genes responsible for the degradation of plant-based polysaccharides even before the introduction of solid foods. It has been suggested that plant-based solid foods should, therefore, be introduced first as the infant may already harbor bacteria able to metabolize these materials. The study by Koenig and colleagues also suggests that the pre-existing genes for plant-based polysaccharide digestion may be the reason why some studies did not show a change in microbiota composition after the introduction of rice cereal (Koenig *et al.*, 2011), as the infant gut is already prepared to digest these plant-based substrates.

Another study also found that the introduction of solid foods causes a change in microbiome composition, however, this study did not find significant changes in phylogenetic diversity with these changes in microbiome composition. An earlier introduction of solid foods was associated with a faster maturation of the gut microbiota, underlining that introducing solids initiates the shift toward the adult-like microbiota. Additionally, the consumption of breast milk was shown to

have strong effects on the gut microbiome even after the introduction of solid foods (Pannaraj *et al.*, 2017). In a Spanish study, 13 infants were followed through the first year of life and the introduction of solid foods resulted in a decrease in bacterial richness, caused by a loss of rare taxa. However, the community structure of the gut microbiota became more complex, with an increase of *Bacteroides* and *Ruminococcus*, and a decrease of *Escherichia*. This study describes the introduction of solid foods as a disturbance to the gut microbiome that leads to permanent alterations (Valles *et al.*, 2014). Thus, the introduction of solid foods is consistently associated with alteration in the gut microbiome, but the specific changes are inconsistent highlighting the need for further research in this area.

1.2.3. The Gut Microbiome and Nutrition

Dietary choices are known to impact alpha diversity, beta diversity and taxonomic distribution of the gut microbiota in adults and rodents. This section briefly summarizes what is known of the relation of nutrient intake to the gut microbiome in the adult gut as this may illustrate or be similar to what happens in the infant gut.

Approximately 12 g/d of undigested proteins can reach the colon, where they are metabolized by the gut microbiome (Beaumont *et al.*, 2017). Bacterial groups correlated with amino acid degradation are of the Firmicutes phylum, especially Clostridiales, as well as the Bacteroidetes phylum (*Odoribacter*) (Beaumont *et al.*,

2017). Primary fermenters of amino acids are *Fusobacterium, Propionibacterium, Streptococcus* and *Bacteroides spp.* this means that these are more active and have potential for growth with increased protein intake of the host (Davila *et al.*, 2013).

The impact of the “Western” diet, with high levels of dietary fat and animal protein, on the microbiome has been investigated recently (Yang *et al.*, 2017). In mice, a high-fat diet (HFD) results in reduced bacterial richness (Candido *et al.*, 2017). As more bile acids are secreted with a diet higher in fat content, bile-resistant bacterial taxa increase in the gut including *Alistipes, Bilophila* and *Bacteroides*, as shown in a human study. Conversely, members of *Firmicutes* that metabolize plant polysaccharides, including *Roseburia, Eubacterium rectale* and *Ruminoccus bromii* are reduced (David *et al.*, 2014).

Much of the research on the influence of carbohydrate intake on the microbiome has focused on dietary fiber. Prebiotics or “microbiota accessible carbohydrates” are the main energy source for bacteria in the gut (Davani-Davari *et al.*, 2019), and selectively support the growth of specific bacteria, although responses by bacteria vary by the individual host and type of fiber. The consensus in many studies is that the ingestion of prebiotics increases the abundance of bifidobacteria (Holscher, 2017). Specific to the infant gut, HMOs, which are prebiotics in human breast milk, have the ability to select for beneficial bacteria, namely *Bifidobacterium* species (Zivkovic *et al.*, 2011). The species most selected for are *Bifidobacterium infantis* and *Bifidobacterium bifidum*; persistence of

bifidobacteria in the gut are supported with the ingestion of HMOs. Studies investigating the effect of prebiotics on the gut microbiome demonstrate that these have beneficial effects on the human host. Since the bacteria primarily fermenting the prebiotics are lactic acid bacteria, a reduction in the gut pH is caused, which can also lead to a change in microbial composition. Abundances of acid sensitive *Bacteroides* species decrease (Davani-Davari *et al.*, 2019; Sonnenburg *et al.*, 2016), however, this has not been observed in all studies (De Filippo *et al.*, 2010). A study performed in mice showed that a low fiber diet led to a decrease in alpha diversity (Sonnenburg *et al.*, 2016). Generally, a diet high in fruit, vegetables and fiber has been linked to increased bacterial richness and diversity (Jandhyala *et al.*, 2015). This is also supported by a study performed in pregnant adult women, which used the healthy eating index (HEI) as a measure for diet quality. High diet quality was positively associated with increased Shannon diversity, especially with increased intake of whole grains and vegetables, which are high in fiber (Laitinen and Mokkala, 2019). The consensus of the literature, therefore, is that a healthy diet, high in fiber increases bacterial richness and diversity.

While studies have focused on the influence of specific macronutrients, few studies have considered the overall diversity of the diet and its impact on the gut microbiome. In the few studies available, a diverse diet is positively associated with gut microbiota diversity measures and greater resilience to environmental changes has also been noted (Claesson *et al.*, 2012; Heiman and Greenway, 2016). Another study did not show an increased diversity of the microbiome with

increased dietary diversity, however, this study demonstrated that the stability of the microbiome was correlated with dietary diversity (Johnson *et al.*, 2019), which is also an indication of a healthy gut microbiome due to a lower susceptibility to perturbances. Overall, this data suggests that a healthy diet with a high amount of fruits and vegetables, as well as high dietary diversity is beneficial for the gut microbiome.

1.2.4. Probiotics and the Gut Microbiome

Probiotics are defined by the FAO and WHO as “live microorganisms which, when administered in adequate amounts, confer a health benefit on the host”. With this definition, probiotics are considered food ingredients, as they are safe to use for healthy populations, probiotics used in therapeutic circumstances do not fall under this definition (Morelli and Capurso, 2012). Common probiotic strains used belong to either *Bifidobacterium* or *Lactobacillus* genera (de Vrese and Schrezenmeir, 2008). Certain foods are considered to have probiotic effects, as they have recognized probiotic bacterial strains. Fermented foods, such as long-ripened cheeses, kefir, salami, sauerkraut, sourdough, etc. (Marco *et al.*, 2017) are also considered to have probiotic effects, as they have recognized probiotic bacterial strains. Probiotic usage influences the gut microbiome composition temporarily. In a study looking at infants fed formula supplemented with Bifidobacteria in comparison with those who did not, supplementation was

associated with a reduction of *Bacteroidaceae*. However, any differences in microbiome composition were no longer visible at 2 years of age (Bazanella *et al.*, 2017) and the influences are subtle (Quin *et al.*, 2018). The effect of probiotics on the gut microbiome, therefore, remains unclear; any effects are only temporary.

1.3. The Importance of the Gut Microbiome in Human Health

1.3.1. Microbiome, Diet and Obesity

The worldwide prevalence of childhood obesity has increased from 32 to 41 million infants and young children from 1990 to 2016 ((ECHO), 2017). The origins of childhood obesity are highly complex and include numerous factors that interact

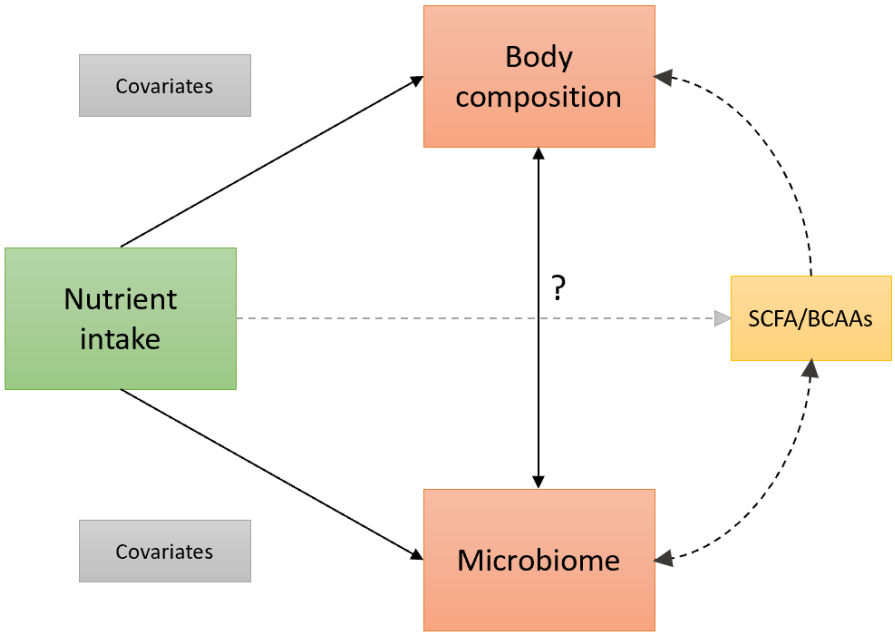


Figure 1: An overview of the relationship between diet, the microbiome and body composition

with each other. Recently, another factor for childhood obesity has been introduced—the **gut microbiome** (Chang and Neu, 2015; Kumar and Kelly, 2017).

Links between the microbiome and obesity have been identified. In *Figure 1* the relationship between diet, the gut microbiome and body composition is summarized.

Gnotobiotic mice with human microbiota transplantation have been used as proof of principle for the interactions between host and microbiome (Williams, 2014). When these mice were fed a Western diet, the microbial community shifted visibly within one day, with increased levels of Firmicutes; this underlines that the microbiome can change rapidly in response to diet (Turnbaugh *et al.*, 2010), as mentioned in **1.2.3. The Gut Microbiome and Nutrition**. When gut microbiota from conventionally raised mice was transplanted into germ-free mice, body fat increased even with lower caloric intake (Backhed *et al.*, 2004), emphasizing the impact of the gut microbiome on body composition.

After there was evidence for a connection between the gut microbiome and obesity in mice, human studies were conducted. In humans, in a study of 138 vaginally born full-term infants, a lower ratio of *S. aureus* to Bacteroidetes was associated with higher body size (BMI SDS, defined as body mass index standard deviation score) in the first three years of life (Vael *et al.*, 2011). Lower levels of *S. aureus* in normal weight children compared to children with overweight / obesity was also noted in another study, and these authors suggested that *S. aureus*

triggers low-grade inflammation, which is a characteristic of obesity (Kalliomaeki *et al.*, 2008). The ratio of Firmicutes to Bacteroidetes is described as a possible biomarker of an obesity-related gut microbiome in multiple studies, and this ratio can be influenced by diet (De Filippo *et al.*, 2010; Ley *et al.*, 2006; Schwiertz *et al.*, 2010; Turnbaugh and Gordon, 2009). Generally, lower diversity can be seen in obese participants (Turnbaugh *et al.*, 2009). All of these studies indicate that there is a relationship between the gut microbiome and obesity. As this relationship is already evident in young subjects, the assembly of a healthy gut microbiome early in life may defend against obesity development.

1.4. The Purpose of this Work

The introduction of solid foods is an important life event in the development of a child and occurs at a time when the microbiome is being established. This is thought to occur as new substrates for the bacteria reach the colon (Fallani *et al.*, 2011), beginning the microbial community changes that ultimately lead to a more adult-like composition of the microbiota (Koenig *et al.*, 2011). Evaluating the relationship of dietary intake patterns of macronutrients (fat, protein, and carbohydrates) and the gut microbiome at a critical window in a child's development could further our understanding of this early development of the gut microbiome. Very few studies have focused on the influence of infant nutrition, beyond breastfeeding, on the infant gut microbiome.

Therefore, the aim of this thesis is to investigate dietary intake during the period of the introduction of solid foods and determine if the pattern of solid food introduction is related to gut bacterial dynamics in healthy, breast-fed infants. This should further our understanding of a factor that impacts the gut microbiome in early life, before the adult-like microbiota is defined at around three years of age. This is especially important, as the adult gut microbiome is linked to health and to the development of common diseases such as obesity.

2. Study Objectives

2.1. Research Questions and Project Objectives

Research Question 1: How do characteristics of the gut microbiome (alpha diversity, beta diversity, taxonomic abundances) change over the 17-day period around the introduction of solid foods?

Objective 1: Identify characteristics of the infant gut microbiome over the sub-study period, in terms of alpha diversity, beta diversity and taxonomic distribution.

Hypothesis (H1): It is hypothesized that alpha diversity will increase, and beta diversity will decrease over the introduction of solid foods and out to one year. The bacterial community of the infant gut microbiome will be dominated by bifidobacteria before the introduction of solids, and as the infants age, the abundance of bifidobacteria will decrease.

Research Question 2*:* Are these changes in the characteristics of the gut microbiome linked to nutrient exposures?

Objective 2: Identify patterns in the introduction of solid foods, in terms of macronutrient intake, food category intake and dietary diversity. Evaluate the relationship between dietary characteristics and the changes occurring in the gut microbiome.

Hypothesis (H1): Differences in nutritional intake patterns will influence the change in characteristics of the gut microbiome in the first two weeks of solid food introduction.

Figure 2 shows the key exposures that will be evaluated in this project, i.e. the potential impact that nutrition, in terms of macronutrient intake, food group intake and dietary diversity, may have on the gut microbiome independent of infant and maternal antibiotic exposure and age at introduction of solid foods.

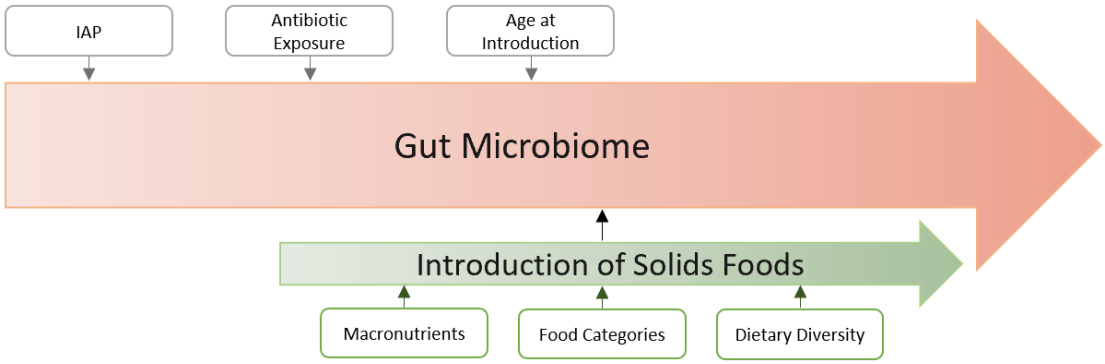


Figure 2: Relationships to be evaluated in this thesis

3. Methodology

This project is embedded in the *Baby, Food & Mi* sub-study of the *Baby & Mi* cohort (see section 3.1 below). In this sub-study, 15 infants from the *Baby & Mi* cohort had intensive data collection over a 17-day period, beginning 3 - 4 days prior to the planned introduction of solid foods. In this 17-day period, caregivers were asked to collect stool samples on a daily basis (2 fresh samples, 15 frozen samples), and to fill out daily food diaries. The theoretical progression through the *Baby, Food & Mi* study is shown in *Figure 3*, where infants remain exclusively breastfed for the first days of the diary and the introduction to solid food occurs on approximately day 4 of the diary.

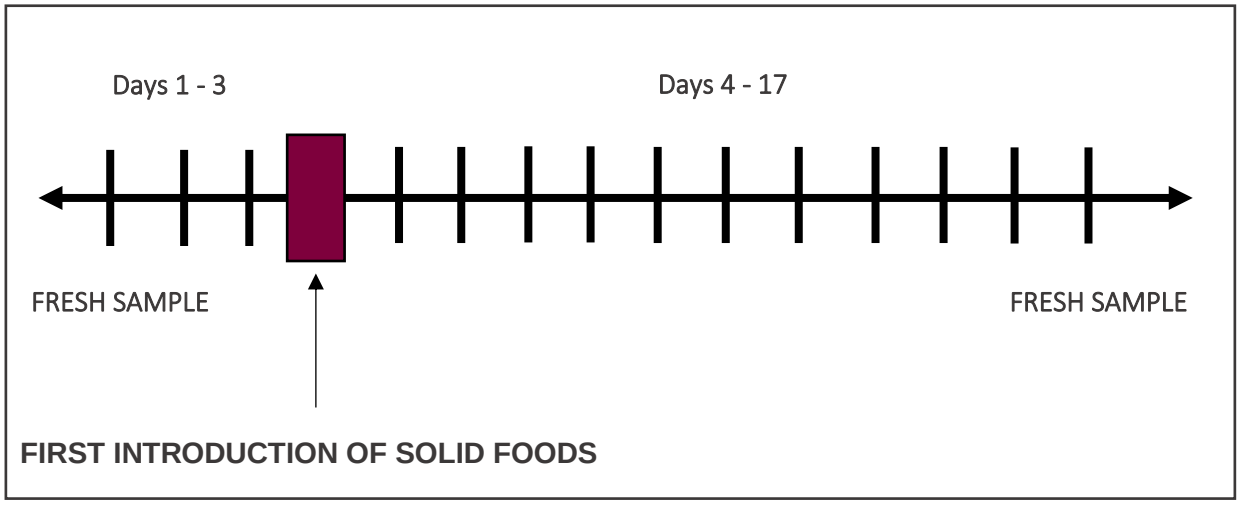


Figure 3: Schematic outline of the study timeline. Daily stool sample collection (max. 17 samples), two fresh samples collected for culturing and metabolomics

3.1. Study Design and Population

Baby, Food & Mi is a longitudinal and observational cohort study, and is a subset of the *Baby & Mi* study. Participants (n = 240) were recruited to the *Baby & Mi* study from midwifery clinics in the Hamilton region (low-risk pregnancies with a planned vaginal birth). They consented to participation during the pregnancy and, at the time of birth, inclusion criteria were re-examined to ensure the infant remained eligible for the study. Inclusion criteria included singleton birth that occurred at full term and the caregiver needed to be able to communicate in English.

Inclusion criteria for the *Baby, Food & Mi* substudy were a willingness to collect samples frequently around the time of introduction of solid foods and the exclusion criteria were: C-section delivery, admission to the neonatal intensive care unit (NICU), weaning of breastmilk before the introduction of solid foods and the use of oral or intravenous (IV) antibiotics within 4 weeks of starting solid foods. These criteria were stringent in an effort to remove some of the factors known to influence the gut microbiome.

In addition to the samples collected throughout the intensively sampled period, samples from the *Baby & Mi* study were collected and used in some of the analyses described below. These samples were collected at day 3, day 10, 6 weeks, 12 weeks, 5 months and 1 year. Questionnaires regarding infant medications, such as antibiotics and probiotics, as well as maternal antibiotic use

were collected at these time points, with additional questionnaires collected at birth, baseline, 4 months and 6 months.

3.2. Exposure

The exposure of interest for this project is the nutrient intake over the period of solid food introduction. This was described as the macronutrient intake (grams of fat, protein, carbohydrate), estimated caloric intake from the macronutrients (calories and per cent of estimated total energy intake), and specific food groups ingested (see section **3.2.4.**). Calories from the macronutrients, as well as the food groups and food items introduced were used to determine food categories (see section **3.2.5.**), as well as dietary diversity scores (see section **3.2.6.**). As this is an exploratory study, a primary exposure was not defined.

3.2.1. Food diaries

To evaluate dietary exposure, participants of the *Baby, Food & Mi* study completed a food diary for 17 consecutive days around the time of solid food introduction. The majority of the days captured in the diary were to occur after the introduction of solid foods. The instructions for filling out the food diary and an example showing how the food diary should be filled out were given at the front of the study diary (see *Fig. 39,* appendix A). Quantity, type of food, the brand name

(if applicable), as well as time of eating were collected. Further, parents were to note if the food was tolerated.

Food diaries are open-ended assessments that record all food and drinks ingested over a predetermined time-period. Food diaries have the advantage of being a prospective means of collecting nutritional intake, and allow for diverse eating patterns, but they can be sensitive to respondent fatigue. Additionally, while food diaries reflect the current diet, this may not be the same as the “usual” dietary pattern. Becoming aware of dietary habits may also lead to the Hawthorne effect, which describes that gaining awareness of certain habits leads to changes in behavior. Further, respondents may under- or over-report their intake, i.e. changing the entries to seem more socially acceptable (Ortega *et al.*, 2015).

In a study looking at the validity of estimated food diaries versus weighed intake in children aged 6 to 24 months it was shown that differences in terms of energy and macronutrients between the two methods were less than 4% and were insignificant. However, both methods over-estimated the energy expenditure based on knowledge of food intake of 7.3% (Lanigan *et al.*, 2001). This study shows that estimating food intake based on household measures can be as accurate as weighed intakes in young children. Based on this evidence and the study design, it was thought that a food diary would best capture the first solid foods given to these infants, though it was understood that these were likely to be estimates only.

3.2.2. Macronutrient Intake

Each item from the food diary was entered into the Food Processor software (© ESHA). Caregivers described intakes as precisely as possible, i.e. either in tablespoons, teaspoons, cups, or milliliters. If this was not the case, quantities were described as slices, halves, thumb-sized, etc., which was then interpreted and standardized between food diaries. The resulting output was exported to Microsoft® Excel for calculations. The Food Processor calculates grams of the macronutrients, as well as total calories ingested. From the mass values of the macronutrients, calories from the macronutrients were calculated using the 9-4-4-7 rule with adjustment for fiber. The 9-4-4-7 rule describes that one gram of fat has 9 kcal, one gram of protein and one gram of carbohydrates have 4 kcal each, one gram of alcohol has 7 kcal and one gram of fiber has 2 kcal. Therefore, total calories for this project were calculated by adding the individual calories of the macronutrients together. Fiber, in the carbohydrate component, has fewer calories, as the non-soluble part is not digestible. Therefore, total calories from carbohydrates was calculated as follows:

$$\left(\left(Carbohydate\ (g) - Fiber\ (g)\right) * 4\frac{kcal}{g}\right) + \left(Fiber\ (g) * 2\frac{kcal}{g}\right)$$

$$= Total\ calories\ from\ carbohydrates\ (kcal)$$

Variables created using the calories from the macronutrients are calories from protein and fat added together (PF), calories from carbohydrates and protein

added together (PC) and calories from fat and carbohydrates added together (FC). These are used in linear models (see section **3.7.2**.)

3.2.3. Proportion of Energy from Solid Foods

The proportion of energy from solid foods at introduction is based on the caloric needs for the infants at this time using FAO (food and agriculture organization) guidelines. Therefore, these numbers are not exact, as the amount of breast milk ingested is unknown and cannot be accurately estimated. The FAO guidelines use age in months and weight in kg for the calculation, using the following factors based on age in months: 99 kcal/kg/d for 4 – 5 months, 96.5 kcal/kg/d for 5 – 6 months and 95 kcal/kg/d for 6 – 7 months (Experts, 1985). The estimated proportions of energy from solid food were calculated as follows:

$$Proportion\ of\ energy\ from\ solid\ foods = \frac{Calories\ from\ solid\ foods\ (kcal)}{Caloric\ needs\ for\ infant\ age\ and\ weight\ (kcal)} * 100\ \%$$

3.2.4. Food group intake

Food groups were manually assigned by looking through the food diaries, according to *Table 13 in appendix A*. The food groups established *a priori* were fruit, vegetables, grains/beans/legumes/nuts, oils, dairy, meat, and confections.

The quantity chosen for food groups is times per day, because serving sizes for infants vary greatly and were unlikely to be accurate.

Depending on the source, the usual food groups described in nutrition recommendations are fruits, vegetables, grains, dairy, protein, confections and fats in the United States (USDA, 1992) or fruits and vegetables, grains and protein in Canada's new food guide (Canadian_Government, 2019). However, since this project differentiates between infants that have been introduced to meat and those that have not, beans, legumes and nuts were added to the grain food group. Additionally, beans, legumes and nuts were introduced more rarely than items belonging to the other food groups.

3.2.5. Grouped Characteristics of Dietary Intake

Infants were categorized into groups based on 4 dietary characteristics or patterns: protein intake, dominant macronutrient, dominant food group, and vegetarian vs omnivorous. Infants were divided into those with high and low protein intake based on being above or below the median intake of this population. The dominant macronutrient in the infant diet, arising from solid foods, was determined according to which macronutrient made up more than 50% of the infant's diet by energy; the categories are carbohydrate based, fat based and 50/50 (percentage of calories from fat and carbohydrates are approximately equal, between 47% and 54%). The dominant food group was determined by the food group that was most

common in the infant's diet (>50% of their diet). The categories are fruit and vegetable based, grain based and mixed. The last category was based on whether the infant ate a vegetarian or an omnivorous diet.

3.2.6. Dietary Diversity Scores

To determine dietary diversity, four scores were created based on the number of food items and food groups consumed and on intake of foods thought to be probiotics or prebiotics. All of the scores were normalized to the number of days of solid food intake by dividing by # of days.

To create the first dietary diversity score, food diaries were inspected and the number of food items, as well as the number of food groups for each infant were documented, as well as the number of days that food was given over the sub-study period. The score was calculated with the following formula:

$$Dietary\ Diversity\ Score\ 1 = \#\ of\ food\ items * \#\ of\ food\ groups / \#\ of\ days.$$

To create the second dietary diversity score, only the number of food groups for each infant, as well as the number of days of food was included. The score was calculated as follows:

$$Dietary\ Diversity\ Score\ 2 = \#\ of\ food\ groups / \#\ of\ days * 10.$$

To create the third dietary diversity score, the number of different food items was included in the calculation, as well as the number of days of food introduction. The score was calculated as follows:

$$Dietary\ Diversity\ Score\ 3 = \#\ of\ food\ items / \#\ of\ days * 10.$$

The last diversity score was calculated considering foods that may have greater effects on the gut microbiome, i.e. prebiotic and probiotic foods. These foods were tallied from the food diaries for each of the infants, and the number of days with foods introduced were considered as well. The formula used looks like this:

$$Dietary\ Diversity\ Score\ 4 = \frac{(2 * \#\ of\ Prebiotic\ foods) + \#\ of\ Probiotic\ foods}{\#\ of\ days} * 10.$$

Prebiotic foods were weighted more heavily than probiotic foods, as they are known to impact the gut microbiome, while effects of probiotic foods are temporary. Foods considered prebiotic included garlic, onions, bananas, oats, apples, flaxseed, wheat bran, whole grain, cruciferous vegetables, legumes, honey, coconut, berries, and corn products (Davani-Davari *et al.,* 2019). Foods considered probiotic include anything fermented, e.g. yoghurt, pickled foods, cheese, tempeh, and sourdough (Marco *et al.,* 2017). Foods were classified as either prebiotic or probiotic retrospectively.

3.3. Outcome

The main outcome of this project is the composition of the gut microbiota of the infants across the time-period where solid foods are first introduced. Stool samples from these time points were collected by the parents and frozen in the provided bags. DNA isolation and sequencing was carried out by the McMaster Metagenomics Facility with the following protocol. DNA was extracted from 0.1 g of stool with mechanical lysis using 2.8 mm ceramic beads and 0.1 mm glass beads for 3 min at 3000 rpm in 800 µl of 200 mM sodium phosphate monobasic (pH 8) and 100 µl guanidinium thiocyanate EDTA N-lauroylsarcosine buffer (50.8 mM guanidine thiocyanate, 100 mM ethylenediaminetetraacetic acid and 34 mM N-lauroylsarcosine) as previously described (Stearns *et al.,* 2015; Stearns *et al.,* 2017). This extract was then purified with the MagMAX-96 DNA Multi-Sample Kit (Life Technologies, Carlsbad, CA) on the MagMAX Express-96 Deep Well Magnetic Particle Processor (Applied Biosystems, Foster City, CA). The DNA was quantified using a Nanodrop 2000c Spectrophotometer (Thermo Scientific, Mississauga, ON Canada). Amplification of the bacterial 16S rRNA gene V3 region (150 bp) tags was performed as previously described (Bartram *et al.,* 2011) with the following changes: 5 pmol of primer, 200 µM of each dNTP, 1.5 mM MgCl2, 2 µl of 10 mg/ml bovine serum albumin, and 1.25 U Taq polymerase (Life Technologies, Carlsbad, CA, USA) were used in a 50 µl reaction volume. The PCR program used was as follows: 94 °C for 2 min followed by 30 cycles of 94 °C for 30 s, 50 °C for 30 s, and 72 °C for 30 s, then a final extension step at 72 °C for 10

min. Illumina libraries were sequenced in the McMaster Genomics Facility with 250 bp sequencing in the forward and reverse directions on the Illumina MiSeq instrument. The completed run was de-multiplexed with Illumina's Casava software. Adapter, primer, and barcode sequences were trimmed from sequencing reads with cutadapt (Martin, 2011) then amplicon sequence variants (ASV) were inferred from the sequenced data using the DADA2 pipeline (Callahan *et al.,* 2016). 16S rRNA sequencing involves PCR amplification of the 16S rRNA gene, followed by high-throughput sequencing of these genes (Jo *et al.*, 2016). Out of the nine hypervariable regions of 16S rRNA, the Surette lab uses the V3 hypervariable region which looks at nucleotides 433-497 (from *E. coli* nomenclature). The V3 region can distinguish bacterial genus- or species-levels well and is able to differentiate closely related Enterobacteriaceae. The sequenced data was run through the Divisive Amplicon Denoising Algorithm 2 (DADA2) pipeline in R to produce ASVs (amplicon sequence variants), as opposed to the previously commonly used Operational Taxonomic Unit (OTU) classification. This means that the sequences are not grouped based on their similarity, but the exact sequences are reflected by the output of the pipeline where differences of as little as one nucleotide can be distinguished. This leads to a more visible fine-scale variation that could arise due to ecological niches, temporal dynamics, and population structure (Callahan *et al.*, 2016). The sensitivity and specificity of ASVs is just as good or even superior to OTUs, because grouping does not occur and comparisons can be made between different studies (Callahan *et al.*, 2017).

Taxonomy was assigned using the SILVA rRNA database (release 132), the highest taxonomic rank assigned here is the genus level. The Greengenes database was used to distinguish species level taxonomy (13_8 release) (DeSantis *et al.,* 2006; McDonald *et al.,* 2012). SILVA and Greengenes taxonomic assignment matched for the organisms, unless otherwise indicated.

3.4. Microbiome analysis

The key metrics of the gut microbiome that were investigated are alpha diversity, beta diversity and taxonomic abundances. Gut microbiome data was analyzed with the *phyloseq* (1.28.0), *microbiome* (1.6.0) and *vegan* (2.5-6) packages in R (R- 3.6.0).

Figure 4 gives an overview of the samples used for each analysis. The analyses are explained below.

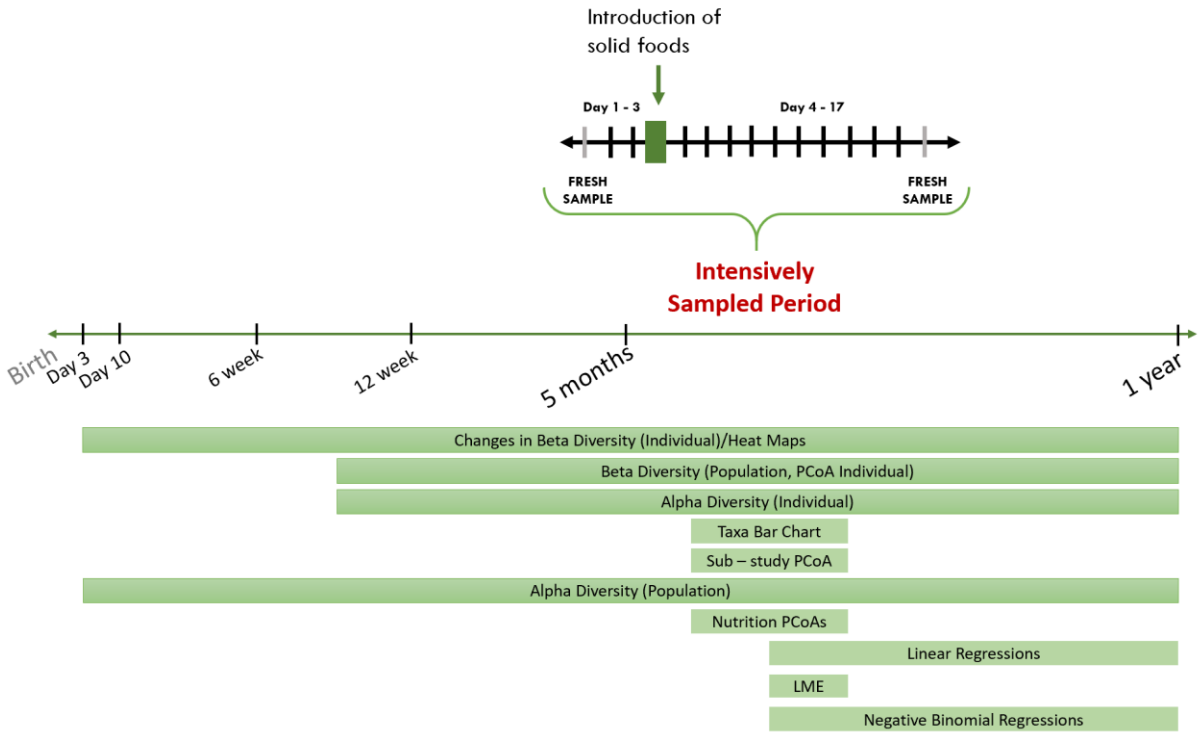


Figure 4: Samples available for analysis, the most important samples are shown in a larger font, i.e. 5 month and one year, as well as the intensively sampled period.

3.4.1. Alpha Diversity Calculations/Plots

Alpha diversity describes the diversity of bacterial taxa in a single sample. This was calculated with the *vegan* as implemented with the *phyloseq* package (McMurdie and Holmes, 2013) and the *plot_richness* function and plotted for each individual infant over time, including samples collected after 3 months of age. The introduction of solid foods usually occurred at around 5 months of age. Two

measures of alpha diversity were calculated, namely observed species richness and the Shannon index. Observed species richness is the most basic alpha diversity measure; this shows the number of ASVs observed in each sample, as observed diversity increases, the number of ASVs increases.

As the Shannon index increases, the higher the alpha diversity of the individual (Morgan and Huttenhower, 2012). The Shannon index weights richness and evenness evenly, in comparison to some other measures of alpha diversity (Wagner *et al.*, 2018). The formula for the Shannon Index is, where p is a frequency measure of each individual taxon, i:

$$Shannon\ Index: H' = -\Sigma_{i=1}^{S}(p_i(\ln(p_i))$$

For alpha diversity plots describing the full *Baby, Food & Mi* cohort, boxplots for each time point, including early visits, were constructed using *phyloseq* and *ggplot2* (Wickham, 2016), and the sub-study samples were plotted individually in a separate graph, as to not create averages for this period leading to a simplification of the data.

To analyze the impact of nutritional choices on alpha diversity for the food categories (*a priori* groupings), the last sample of the sub-study available, i.e. the sample taken 10 – 20 days after the introduction of solid foods, of the intensively sampled sub-study was used to calculate and plot alpha diversity. The first sample obtained prior to the introduction of solid foods was used for analysis to determine whether the alpha diversity measures between groups were already significantly

different before the introduction of solid foods, indicating that the introduction of solid foods may not have had an effect. This was done if significant differences were seen between groups in the tests comparing means for the last sample. Outliers, defined as being smaller than the first quartile subtracted by 1.5 * the interquartile range (IQR) or larger than 1.5 * the IQR added to the third quartile were removed from the dataset for multiple comparison analysis. These statistical tests included tests of variance homogeneity, distribution normality of the residuals and appropriate tests for the comparison of means, either a t-test, Wilcoxon test, Kruskal-Wallis test, or a Tukey-Kramer test (see **3.7.1**.)

3.4.2. Beta Diversity Calculations/Plots

Beta diversity is quantified using the Bray-Curtis dissimilarity measure, which has the following equation, where S_i and S_j are either the number of species or the relative abundance of each species[3] in populations *i* and *j* respectively and C_{ij} is the total number of species of the population with the lowest amount of species:

$$Bray - Curtis\ dissimilarity: BC_{ij} = \frac{S_i + S_j - 2C_{ij}}{S_i + S_j}$$

[3] Both methods are used in the literature.

As the Bray-Curtis dissimilarity index increases, diversity between the samples increases. The values range from 0 to 1, if the value is 1 then no species are shared between individuals (Morgan and Huttenhower, 2012).

Beta diversity was calculated using the *phyloseq* package and the *ordinate* and *plot_ordination* functions. This created principal coordinates analysis (PCoA) plots for each individual infant. Principal coordinate analysis depicts beta diversity; this is a method that reduces the dimensions of the data set, so that beta diversity relationships can be shown in scatterplots. Principal coordinates, usually the first two, are plotted (Goodrich *et al.*, 2016). Additionally, the differences between samples for each individual infant were plotted to show changes over time. This used the Bray-Curtis dissimilarities calculated using the *vegdist* function of the *vegan* package (Oksanen *et al.*, 2019). In these plots, the value was plotted on the day of the later sample, i.e. if the difference of interest in beta diversity is between the day 3 and the day 10 samples, the calculated dissimilarity is plotted on day 10.

On the population level, i.e. including all samples from all participants of the sub-study, a PCoA plot was created for samples collected later than 90 days, as well as boxplots for the changes from time point to time point, and samples from the sub-study categorized as “intensive before”, “intensive early” and “intensive late”. For these groupings, averages of the Bray-Curtis dissimilarity values were calculated. The before samples, usually 2, were averaged to create the “intensive before” value, 2 – 3 samples after introduction were averaged to create the

"intensive early" variable, and the last 2 samples were averaged to create the "intensive late" variable.

The degree of movement on the PCoA was qualitatively assessed, and the participants were divided into three groups: little change, i.e. samples from one participant are closely clustered and movement along either axis does not exceed 0.1, great change, i.e. movement along either axis of the PCoA plot exceeds 0.4, and moderate change, where movement along the axis lies between 0.1 and 0.4, throughout the study period.

PERMANOVAs (permutational analysis of variance) were conducted, using 9999 permutations, with the vegan package in R on Bray-Curtis dissimilarity matrices calculated from ASV relative abundances. The *a priori* groupings (protein intake, dominant food group, dominant macronutrient, vegetarian/omnivore), calories from macronutrients, grams of fiber and dietary diversity scores, as well as the covariates probiotic use, antibiotic use, GBS prophylaxis and gestational hyperglycemia, were tested for their association with variation in the microbiome composition. This was done for the last sample in the sub-study period, as well as all sub-study samples.

3.4.3. Relative Abundance

Relative abundance was calculated in R by transforming the original counts to proportions of total counts of bacterial classifications. Relative abundance was

used to create heat maps using the *heatmap.2* function from the gplots package (3.0.3) in R. Key bacterial ASVs were determined from the heat maps for the sub-study period only, when relative abundance was greater than 0.35 more than once for a specific ASV in combination with the top 10 ASVs for all samples, which was determined using taxa bar charts.

3.4.4. Prevalence

Prevalence of bacterial ASVs was assessed using the microbiome package (Lahti and Shetty, 2012-2019) and the *core* function in R. The *core* function allows a cut-off point to be set, so that the ASVs that are present in x % of samples are listed. Prevalence was plotted using the *plot_core* function. The cut-off point for these analyses was 50%.

3.5. Covariates

Many factors may impact the gut microbiome at the introduction of solid foods, apart from the type of complementary foods. These factors include delivery mode, feeding (i.e. breast milk vs. formula), home environment (siblings and/or pets), antibiotic usage, probiotic usage, infant sex and age at introduction to solid foods, as well as age (days) of infants at the time of sample collection. Due to the exclusion criteria of the intensively sub-sampled group, delivery mode can be disregarded as a covariate. None of the infants in the intensively sampled sub-

group were receiving formula at the time of introduction of solid foods, and only three of them had received formula in their lifetimes. The other factors are all captured in the case report forms filled out at the study visits (CRFs from day 3 to 6 months) and can be accounted for. Data surrounding the potential covariates can be seen in **Chapter 4: Results.** Data surrounding perinatal factors which may impact the gut microbiome is collected as described above from the birth and baseline CRFs and includes GBS prophylaxis, gestational diabetes, and maternal pre-pregnancy BMI (used to determine incidence of maternal obesity).

3.5.1 Calculations and Classifications

3.5.1.1. Maternal Pre-pregnancy BMI (body mass index)

Maternal pre-pregnancy height and weight were collected on the birth form of the *Baby & Mi* study, this was self-reported either in centimeters and kilograms or in pounds and feet. The measurements were first converted to centimeters and kilograms for the calculation of the BMI. BMI is a convenient measure to categorize subjects into the following groups: underweight, normal weight, overweight and obese. BMI is calculated as follows:

$$BMI = \frac{weight\ (kg)}{height\ (m)^2}$$

BMI cut-off points are the following: below 18.5 kg/m^2 is considered underweight, from 18.5 to 24.9 kg/m^2 is considered normal weight, 25.0 to 29.9 kg/m^2 is

considered overweight and 30.0 kg/m^2 and above is classified as obese (Weir and Jan, 2019).

3.5.1.2. Oral Glucose Challenge Test

Gestational diabetes is captured on the birth form of the *Baby & Mi* study, the results of a 60-minute oral glucose challenge (OGCT) are included. The OGCT is considered a glucose-loading test, because the expecting mothers ingest 50 g of glucose and venous glucose levels are measured at baseline and one hour later. Glucose levels are considered elevated if above 7.8 mmol/L. Generally, the recommendation to assess gestational diabetes is using the oral glucose tolerance test (OGTT) after an abnormal finding was seen in the OGCT, however, this test was not carried out in all of the participants, which is why the OGCT is chosen for these purposes (van Leeuwen *et al.*, 2012).

3.7. Statistical Methods

Table 3: Overview of the statistical methods used, and variables included in the methods with classification into either continuous (C) or categorical (K).

Microbiome Outcome	Calculated Measure	Statistical Analysis	Samples Used
Research Question I: How do characteristics of the gut microbiome (alpha diversity, beta diversity, taxonomic abundances) change over the introduction of solid foods?			
I.i. Alpha Diversity	Shannon diversity index Observed species richness	- simple linear regression with age in days (C) - used estimate_richness for average before and average after samples to compare before and after quantitatively (Welch test)	- Individual infants: all samples beyond three months of age; Cohort: all samples from all infants from day 3 to one year - Linear regression: all samples from day 3 to year one - Welch test: only sub-study samples, separated into before and after the introduction of solid foods
I.ii. Beta Diversity	Bray-Curtis dissimilarity	- PCoA plot for all sub study samples - PCoA for after samples only - PCoA of all samples > 90 days - Average BC dissimilarity – Before, Early, Late (boxplot), and other study visits	- PCoA: all sub-study samples - PCoA: only after samples for the sub-study - PCoA: all samples > 90 days - Boxplot: all samples

Research Question II: Are these changes in the characteristics of the gut microbiome linked to nutrient exposures?			
II.i. Alpha Diversity - **Macronutrients and Fiber** - **Food categories** - **Dietary Diversity scores (DDS)**	Shannon diversity index Observed species richness	-Multiple comparisons for first/last sample of sub-study for the food categories -LME Analysis: random effect is PID, fixed effects vary by model. Fixed effects include: CHO calories (C), Protein calories (C), Fat calories (C), Fiber (g/d)(C), dominant macronutrient (K), dominant food group (K), protein grouping (K), vegetarianism (K), PF calories (C), PC calories (C), CF calories (C), dietary diversity scores (C), age in days (C), age at introduction (C), GBS prophylaxis (K) depending on the model.	-Multiple comparisons: first and last sample of the sub-study -LME Analysis: sub-study samples after the introduction of solid foods
II.ii. Beta Diversity - **Macronutrients and Fiber** - **Food categories** - **DDS**	Bray-Curtis dissimilarity	- PCoA plots colored by calories of macronutrients (C), g/d of fiber (C), food category (K) or dietary diversity score (C) - PERMANOVAs unadjusted for each - run for all “after” samples, and for last sample only	- PCoA: only samples after solid food introduction of the sub-study period -PERMANOVAs: last sample only, and all after samples of the sub-study

II.ii. Beta Diversity - **DDS**	Qualitative assessment of degree of change over sub-study period	- Spearman correlation of degree of change and dietary diversity score - simple linear regression of degree of change and dietary diversity score (C)	- No samples used. Observations from PCoA plot and calculated dietary diversity scores.
II.iii. Taxonomic Distribution	Count Data of ASVs	Negative binomial regressions with offset for total counts of the ASV were carried out. Variables chosen for models seen in *Table 5*.	-All samples after the introduction of solid foods
Covariates			
Alpha Diversity	Shannon diversity index Observed species richness	-Multiple comparisons for last sample of sub-study for the covariates: probiotic usage (K), antibiotic usage (K), GBS prophylaxis (K), maternal obesity (K), hyperglycemia during pregnancy (K)	- Multiple comparisons: last sample of sub-study, first sample of sub-study
Beta Diversity	Bray-Curtis dissimilarity	- PCoA plots colored by the covariates (K) - PERMANOVAs unadjusted for each	- PCoA: all after samples of the sub-study - PERMANOVA: last sample only and all after samples of the sub-study

Created variables are calories from protein and fat added together (PF), calories from carbohydrates and protein added together (PC) and calories from fat and carbohydrates added together (FC). Additionally, for the

covariates, maternal obesity was derived from pre-pregnancy BMI and gestational diabetes/hyperglycemia during pregnancy from the results of the OGCT.

3.7.1. Comparison of Means

Tests to compare means were used throughout this analysis, including the t-test, Welch test, Tukey-Kramer test and Kruskal-Wallis test, depending on the prerequisites for each test, i.e. variance homogeneity, distribution of the outcome and distribution of the residuals (*Fig. 5*). These prerequisites were tested using the Fligner-Killeen test and the Shapiro-Wilk test. The Wilcoxon test and Kruskal-Wallis test were used when the prerequisites were non-parametric; these tests are less likely to show significant results. These tests were used to compare alpha diversity between *a priori* groupings (RQ2), alpha diversity at different time points (RQ1), as well as the change in beta diversity for different time points (RQ1).

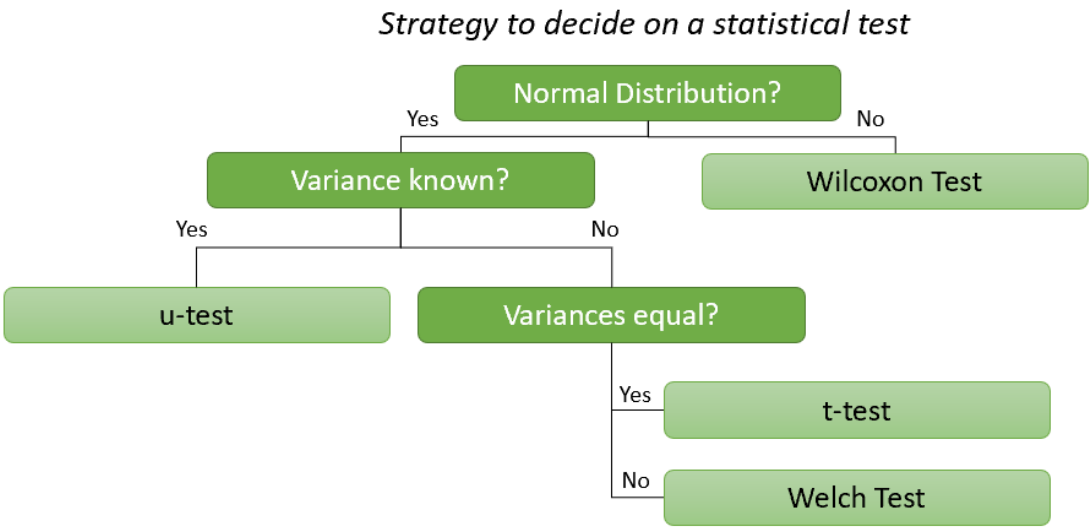


Figure 5: Flowchart to depict which test was chosen for comparison of means (not for multiple comparisons). Multiple comparisons used were either a Tukey test for parametric data or a Kruskal-Wallis for non-parametric data.

3.7.2. Linear Modeling (Linear Regression and Linear Mixed Effects Analysis)

Multiple linear models were performed to investigate the research questions of this study, each analysis corresponds to a microbiome outcome in *Table 3*. To investigate research question one, simple linear regressions were carried out to determine the relationship between infant age in days and measures of alpha diversity (I.i, *Table 3*).

To further explore research question 2: “Are the changes in the gut microbiome linked to nutrient exposures?”, linear mixed effects analyses and multiple comparisons were carried out (II.i, *Table 3*). All of the following linear mixed effects analyses use the intensively sampled samples, as well as any samples following up to one-year, i.e. this is a longitudinal approach for analysis of the gut microbiome. Linear mixed affects analysis was used to determine the relationship between alpha diversity and the calories from the macronutrients. The fixed effects vary by model (illustrated in *Table 4*), while the random effect is the same for all models, namely participant ID. This was accomplished using the *lme4* package (1.1-23) (Bates *et al.*, 2015). The difference between the standard multivariate models and the energy partition models is that the standard multivariate model includes total calories ingested by each infant, while the energy partition model only adds calories of the remaining macronutrients, i.e. if the relationship between x and calories from fat is being investigated, the second

variable added is calories from protein and carbohydrates, instead of total calories. These models are described in more detail elsewhere (Willett *et al.*, 1997).

Table 4: Models for linear mixed effects analyses and negative binomial regressions.

Variables Included					
Macronutrient Calories – Shannon diversity					
Unadjusted	Calories from the macronutrients	-	-	-	-
Standard Multivariate 1	Calories from the macronutrients	Total calories	-	-	-
Standard Multivariate 2	Calories from the macronutrients	Total calories	Age in days	Age at introduction	GBS prophylaxis
Energy partition 1	Calories from the macronutrients	Calories from remaining macronutrients	-	-	-
Energy partition 2	Calories from the macronutrients	Calories from remaining macronutrients	Age in days	Age at introduction	GBS prophylaxis
Macronutrient Calories – observed species richness					
Unadjusted	Calories from the macronutrients	-	-	-	-
Standard Multivariate 1	Calories from the macronutrients	Total calories	-	-	-
Energy partition 1	Calories from the macronutrients	Calories from remaining macronutrients	-	-	-
Fiber (g/d) – Shannon diversity and observed species richness					
Unadjusted	Fiber (g/d)	-	-	-	-

Standard multivariate 1	Fiber (g/d) + Calories from carbohydrates	Total calories	-	-	-
Standard multivariate 2	Fiber (g/d) + Calories from carbohydrates	Total calories	Age in days	Age at introduction	GBS prophylaxis
Food categories – Shannon diversity and observed species richness					
Unadjusted	Food category	-	-	-	-
Standard multivariate 1	Food category	Total calories	-	-	-
Standard multivariate 2	Food category	Total calories	Age in days	Age at introduction	-
Standard multivariate 3	Food category	Total calories	Age in days	Age at introduction	GBS prophylaxis
Dietary Diversity – Shannon diversity and observed species richness					
Unadjusted	Dietary diversity score	-	-	-	-
Standard multivariate	Dietary diversity score	Total calories	Age in days	Age at introduction	GBS prophylaxis

3.7.3. Negative Binomial Regression

For research question two, the relationship between the key bacterial ASVs and the dietary variables were evaluated using negative binomial regressions (II.iii., *Table 3*). This controls for participant ID. In the model an offset for total counts was added to create a better fit of the model to the data. The models used are seen in *Table 5*.

Table 5: Models used for negative binomial regression analyses.

Variables Included			
Macronutrient Calories			
Unadjusted	Calories from the macronutrients	-	-
Standard Multivariate 1	Calories from the macronutrients	Total calories	-
Standard Multivariate 2	Calories from the macronutrients	Total calories	Age in days
Energy partition 1	Calories from the macronutrients	Calories from remaining macronutrients	-
Energy partition 2	Calories from the macronutrients	Calories from remaining macronutrients	Age in days
Fiber (g/d)			
Unadjusted	Fiber (g/d)	-	-
Standard multivariate 1	Fiber (g/d)	Total calories	-
Standard multivariate 2	Fiber (g/d)	-	Age in days
Standard multivariate 3	Fiber (g/d)	Total calories	Age in days
Food categories			
Unadjusted	Food category	-	-
Standard multivariate 1	Food category	Total calories	-
Standard multivariate 2	Food category	-	Age in days
Standard multivariate 3	Food category	Total calories	Age in days
Dietary Diversity			
Unadjusted	Dietary diversity score	-	-

Standard multivariate	Dietary diversity score	Total calories	Age in days
Standard multivariate 1	Dietary diversity score		Age in days
Standard multivariate 2	Dietary diversity score	Total calories	Age in days

This was done using the *glmmTMB* package (1.0.1) and function in R (Brooks *et al.*, 2017). The negative binomial regression model assumes that the dependent variable follows the negative binomial distribution and is usually used to model count variables that may be over-dispersed (Group, 2020).

3.7.4. Spearman Correlation

For research question two, the relationship between the dietary diversity scores and the degree of movement on the population PCoA plot was analyzed using a Spearman correlation. The Spearman correlation was chosen, because the independent variable, degree of movement, is an ordinal variable and differences between groups may not be equivalent, disqualifying a Pearson correlation.

All code for this project can be found in the GitLab repository under the following link: https://gitlab.com/ChiaraHomann/baby-and-mi----introduction-of-solids-and-the-gut-microbiome---ch

4. Results

4.1. Study Population

The *Baby, Food & Mi* sub-study of the *Baby & Mi* cohort has 15 participants, from whom stool samples and food diaries were collected. A number of potential covariates for the analysis of the composition of the gut microbiome can be seen in *Table 6,* along with characteristics of the infants at the time of first introduction of solid foods. Maternal pre-pregnancy BMI was in the normal weight category with a mean (SD) of 24.1 (4.16) kg/m^2, only 5 of the 15 mothers of participants had a BMI considered overweight or obese pre-pregnancy. The average maternal age was 32.4 years (SD = 2.67). In this study population, 4 mothers had gestational diabetes, which is 27% of the cohort. Intrapartum antibiotic prophylaxis against group B streptococcus was administered during labour to 3 of 15 mothers (20%).

Table 6: Description of maternal characteristics, and infant characteristics at the time of introduction to solid foods

	n (%)	**Median**	**Range**	**Mean (SD)**
Pre-pregnancy BMI (kg/m^2)	15 (100)	23.5	17.7, 34.0	24.1 (4.16)
Maternal age (years)	15 (100)	32.4	27.4, 36.9	32.4 (2.67)
Gestational diabetes[1] (y)	4 (27)	-	-	-
Glucose levels after OGCT (mmol/L)	11 (73)	7.1	4.3, 9.0	6.6 (1.7)
GBS prophylaxis (y)	3 (20)	-	-	-
Gestational age at birth (weeks)	15 (100)	40.0	38.3, 41.7	40.0 (0.95)
Age at introduction of solid foods (months)	15 (100)	5.79	3.98, 6.5	5.5 (0.66)
Weight-for-length z-score	15 (100)	-0.52	-2.47, 1.38	-0.20 (1.098)
Female Sex	7 (47)	-	-	-

Infant oral antibiotic use before introduction of solid foods	1 (7)	-	-	-
Maternal oral antibiotic use before introductions of solid foods	1 (7)	-	-	-
Infant probiotic use before introduction of solid foods	5 (33)	-	-	-

Gestational age at birth was 40.0 (0.95) weeks on average (SD), which is full term. The age at introduction of solid foods was 5.5 (0.66) months on average, ranging between approximately 4 and 6 months. Only one infant received oral antibiotics before the introduction of solid foods, at 4 months of age (5.5 weeks prior to the start of the sub-study period), and five infants received probiotics. Probiotic use started at 6 weeks of age for most infants and two infants were still being administered them at 5 months. Probiotics used were either BioGaia, Genestra HMF Baby B or not named by the caregiver. Since infants were still breastfed throughout the introduction period, maternal antibiotic use is also of interest. Only one mother used antibiotics prior to the introduction of solid foods, at the 6-week time point.

4.3. Nutritional Data

4.3.1. Overview of Food Diary Collection

Food diaries are the gold standard for collecting dietary data (**see 3.2.1.**); however, recording complete dietary information for 17 days can be tedious, especially when having to care for an infant simultaneously. The completion rates for the diaries in this study population can be seen in *Fig. 40 of appendix A*; nutritional data is only available for 14 infants. In general, the completion of the food diaries was very thorough—very few days were left blank. Most of the caregivers abided to the proposed timeline for the study, with the introduction of solid foods occuring around day 4 (mean (SD): 3.3 (1.78) days), with the exception of one participant, who introduced solid foods on day 10. The last day of the study period (day 17) is often left blank, because the last stool sample is collected on day 17; depending on when the infant passes stool, the study ends “early”. Occasional days were left blank or did not introduce solid foods; the mean (SD) number of days where this occurs is 0.93 (1.223) days.

4.3.2. Average intake over the first 17 days after solid food introduction

The initial solid food intake was very small, estimated to be 4.7% (IQR: 9.23%) of required energy intake – based on infant age and weight. The proportion of energy from solid foods at the time of introduction to solid foods is shown in *Figure 6*.

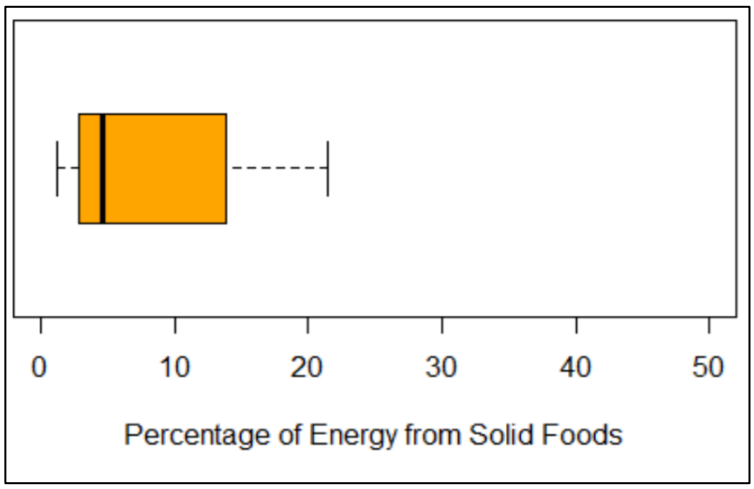


Figure 6: Boxplot of the percentage of energy (on average) from solid foods in the first ~14 days of introduction (8 – 15 days).

From *Figure 6* it is clear that the majority of infants (over 50%) consume less than 5% of their expected energy needs from solid foods over this initial period of solid food intake. Six out of fourteen infants had greater than 5% of their energy from solid foods but the proportion of energy from solid foods ranged from 1.2% to 21.4%, averaged over ca. fourteen days from commencement of solid foods.

Table 7 shows the average number of calories from each macronutrient over the time period where solid foods were introduced, after excluding days without solid food introduction. The macronutrient with the lowest average (SD) overall intake was protein with 6.79 (6.708) kcal, ranging from 0.43 kcal to 16.7 kcal per day, followed by fat with 21.2 (22.23) kcal, ranging from 0.64 kcal to 69.3 kcal per day. Carbohydrates contributed most to caloric intake with an average of 27.5 (27.02) kcal/day, ranging from 3.47 kcal to 86.3 kcal/day. Total average caloric intake was 50.7 (50.70) kcal per day, ranging from 10.9 kcal to 132 kcal.

Table 7: Average daily caloric intakes from the individual macronutrients over the days were solid food was introduced

	Average Protein (kcal/d)	**Average CHO (kcal/d)**	**Average Fat (kcal/d)**	**Average Total (kcal/d)**
01-057	0.43	10.8	0.64	10.9

02-043	2.67	8.67	9.48	20.8
02-045	7.26	39.1	49.9	94.3
02-046	0.50	3.47	4.14	8.11
02-048	3.28	16.0	9.08	28.4
02-049	4.06	40.6	10.1	28.7
03-033	0.62	3.80	3.50	7.92
04-061	8.38	6.19	27.0	41.6
04-066	2.13	15.4	2.94	20.4
04-068	2.13	15.4	2.94	20.4
06-019	2.51	24.7	7.65	34.9
06-020	15.4	26.6	69.3	118
06-021	16.7	86.3	29.2	132
06-023	9.98	20.3	17.7	48.0
Group mean (SD) (n = 14)	**6.79 (6.708)**	**27.5 (27.02)**	**21.2 (22.23)**	**50.7 (50.70)**

The day to day variation in macronutirent intake from solid foods over the 17-day time period is shown in *Figure 7.* As expected, the number of calories ingested from solid foods gradually increases over time. Protein intake seems more stable over time than the other macronutrients. Carbohydrates show a gradual incline in calories and the greatest day-to-day variation was in fat intake.

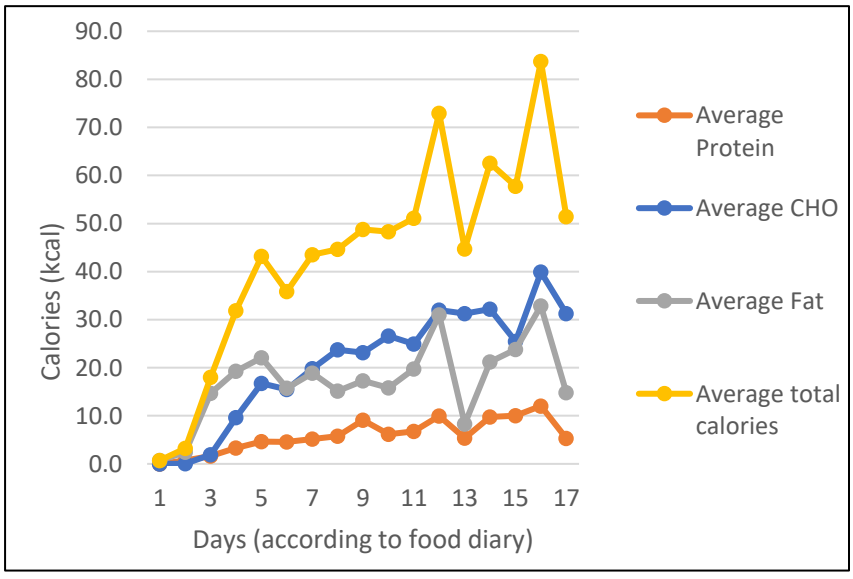


Figure 7: Average daily caloric intake from the individual macronutrients over the first 17 days after introduction of solid food

Looking at the last day alone, mean (SD) of the estimated energy consumed from solid foods is 11.8 % (15.9%), which demonstrates the large variability in the dataset. The minimum amount of energy from solid foods is 1.04% and the maximal amount is 62.2%, which is much higher than the energy intake for most of the participants (see *Figure 41 in appendix A)*.

4.3.3. Nutrient Intakes on a Day-to-Day Basis

After entering the food diaries into the Food Processor (ESHA) and calculating caloric intake using the 9-4-4-7 rule with adjustment for fiber, the output was used to create the graphs seen in *Figure 8* for each individual participant.

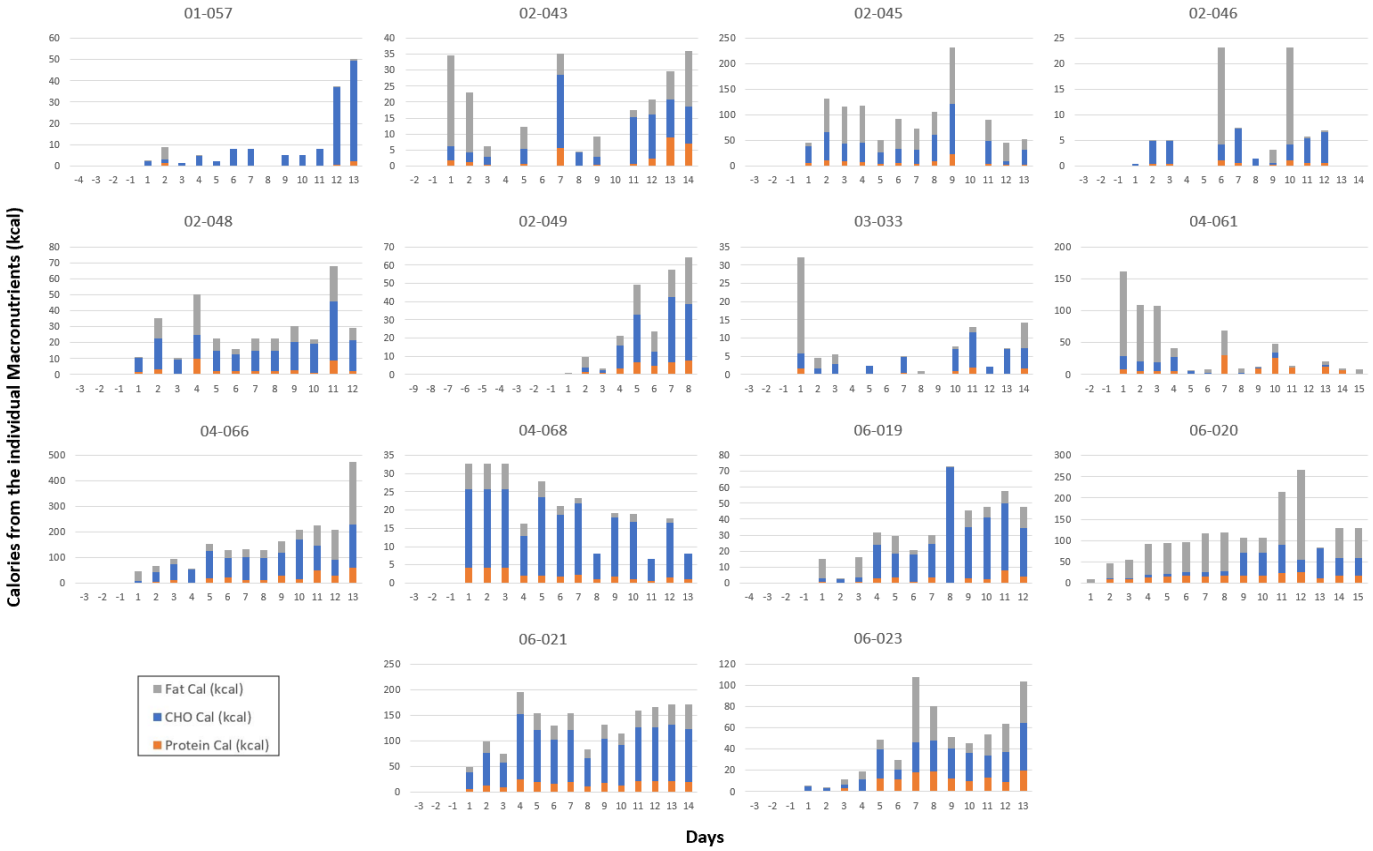


Figure 8: Calories from the individual macronutrients for each participant individually on a day-to-day basis. Blanks indicate either no food was introduced, or the diary was incomplete. Negative days signify days before the first introduction of solid foods; day 1 is the first day where solid foods were introduced.

As expected, for most infants total caloric intake increased over time (n = 7), while other infants showed a decline in caloric intake over time (n = 3). The remaining infants had fluctuating intakes of calories, showing irregular peaks

throughout the introductory period (n = 4). The dominant macronutrient for the infants was determined using *Figure 8*. For most participants, carbohydrates outweigh the contributions made to caloric intake by the other macronutrients, the exception is when fat has a high contribution or when fat and carbohydrates have an equal contribution. Protein intake is relatively low in most participants, contributing to less than 20% of the calories. For the protein intake *a priori* grouping, infants were classified around the group median intake of protein (See *Figure 42* in *appendix A* for a clearer figure). *Figure 9* summarizes the groupings for the participants according to their dominant macronutrient and protein intake.

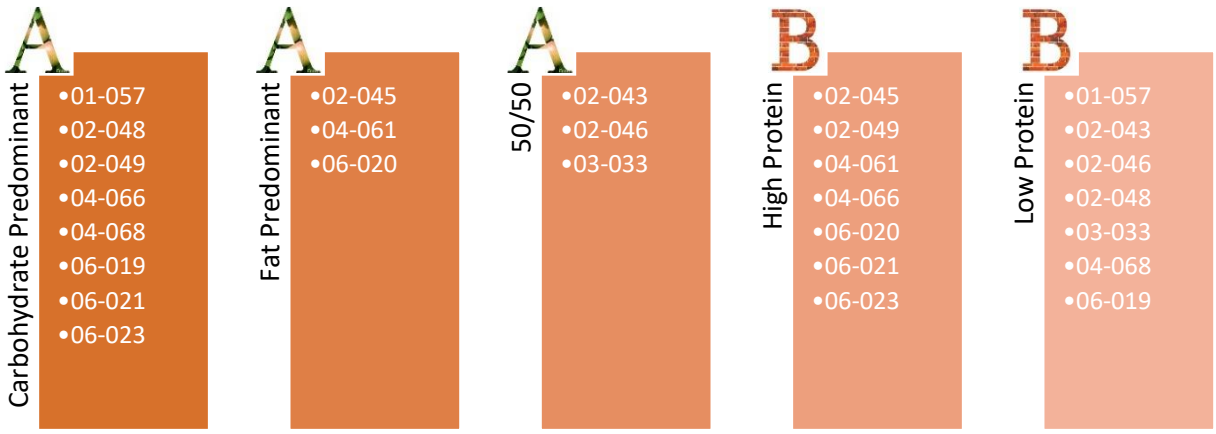


Figure 9: A priori groupings of the infants according to macronutrient intakes. A: groupings for dominant macronutrients, i.e. carbohydrate dominant, fat dominant and carbohydrate/fat mixed diet. B: protein groupings, i.e. high or low protein intake.

4.3.4. Fiber Intake

Fiber intake was low at the introduction of solid foods with a median (IQR) intake of 0.32 (1.015) g/d, and a range of 8.98 g/d. *Figure 10* shows the distribution of daily fiber intake (g/d) over the sub-study period. Most of the infants are ingesting very low amounts of fiber at the introduction of solid foods. Individual intakes of fiber for each day of the sub-study can be seen in *Fig. 43 of appendix A.*

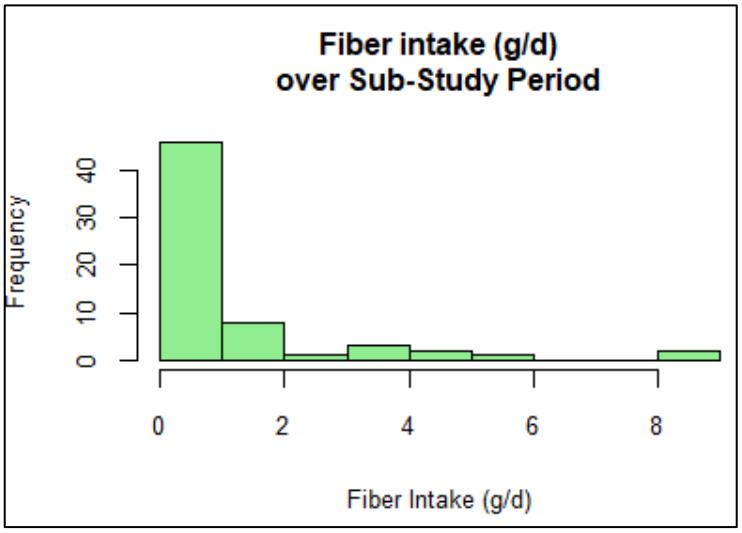


Figure 10: Histogram of the distribution of fiber intake over the sub-study period (g/d).

4.3.5. Food Groups

In addition to macronutrients, the food entries from the diaries were sorted into food groups according to *Table 13* in Appendix A and as described in the methods.

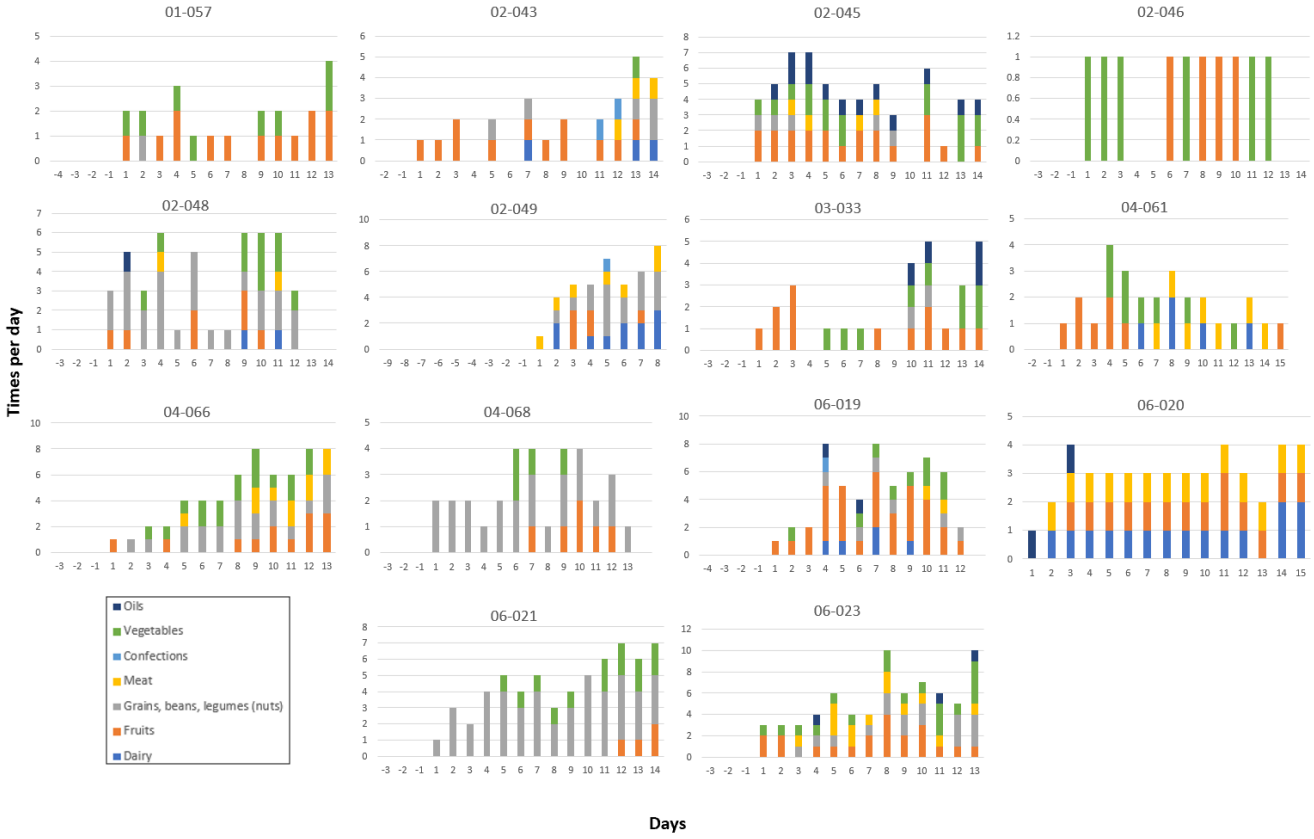


Figure 11: Food group intake for each participant on in individual, day-to-day basis. Blanks indicate either no food was introduced, or the diary was incomplete. Negative days signify days before the first introduction of solid foods; day 1 is the first day where solid foods were introduced.

From these graphs (*Fig. 11)*, three groups of infants were classified on the basis of the food groups that were most common in the infants' diets. These groups are: fruit and vegetable dominant (n = 6), grain dominant (n = 3) and mixed (n =

5). Dominant in this case means that more than 50% of the diet is derived from a specific food groups with 04-061 being the exception, as meat explains the same percentage as fruit and vegetables individually. In both the fruit and vegetable dominant and grain dominant group, there are infants that have received either no animal products, have received only meat, or have received both meat and dairy. The infants in the mixed group have all had meat and/or dairy introduced. *Figure 12* shows which participants are in each category.

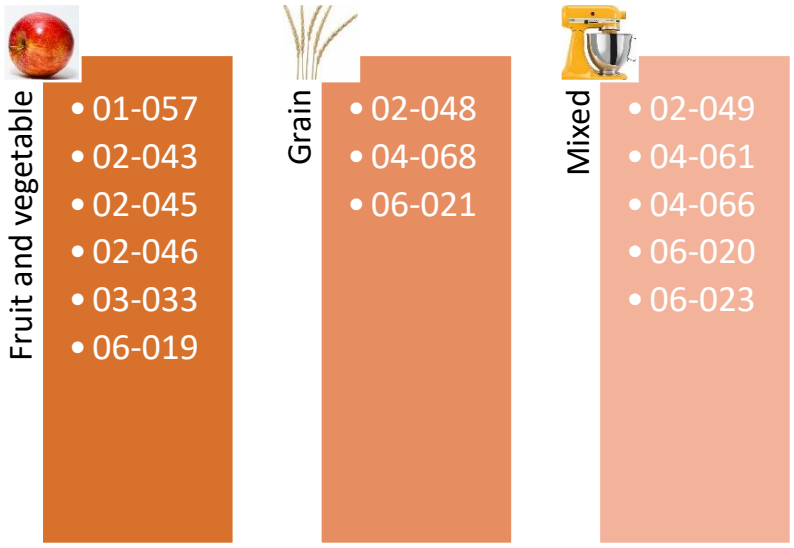


Figure 12: Overview of the participants in each dominant food group a priori *grouping.*

Infants were also categorized according to their introduction to animal products, either vegetarian (n = 5) or omnivorous (n = 9). The vegetarian infants are: 01-057, 02-046, 03-033, 04-068, 06-021, while the omnivorous infants are: 02-043, 02-045, 02-048, 02-049, 04-061, 04-066, 06-019, 06-020, 06-023. Vegetarian in this case means vegan, as dairy products were always introduced in combination with meat, however, meat isn't necessarily introduced in combination with dairy.

4.3.6. Dietary Diversity Scores

Dietary diversity scores were calculated as seen in **3. Methodology (3.2.6.)**. The following values were seen for the infants of the sub-study (*Table 8*), where score one integrates both food groups and items, score two integrates food groups only, score three integrates food items only and score four emphasizes foods known to impact the gut (prebiotics and probiotics).

Table 8: Descriptive statistics for the dietary diversity scores.

Diversity Score	**Mean (SD)**	**Range**	**Min, Max**
1 – Food groups and items	6.34 (5.05)	14.83	0.8, 15.63
2 – Food groups only	3.6 (1.45)	4.3	2, 6.3
3 – Food items only	12.8 (8.06)	27.3	4, 31.3
4 – Pre-/probiotic foods	7.8 (5.54)	16.7	0, 16.7

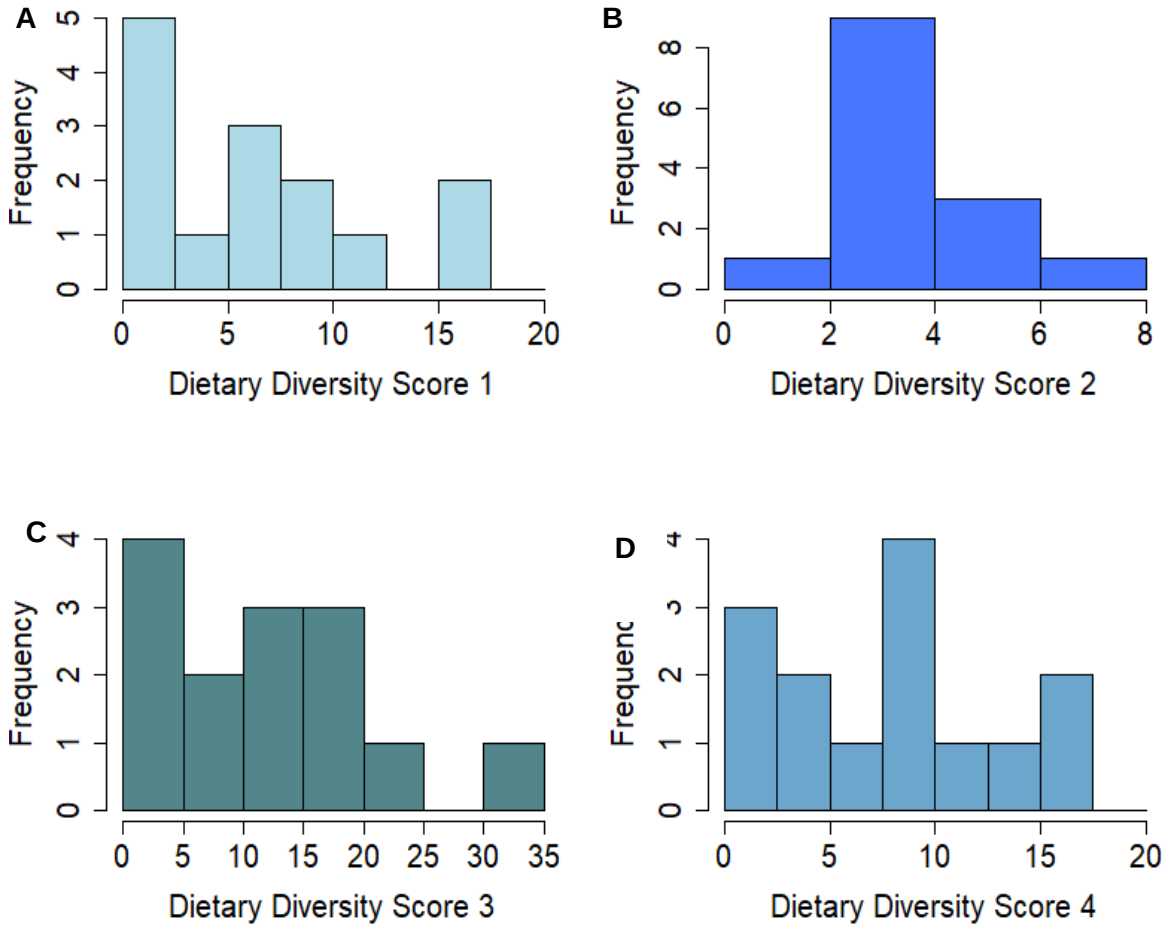


Figure 13: Histograms showing the distribution of dietary diversity scores. A: Dietary diversity score 1, B: Dietary diversity score 2, C: Dietary diversity score 3, D: Dietary diversity score 4.

The first dietary diversity score in this sample has a mean (SD) of 6.34 (5.05) with a range of 14.83 food groups * food items per day. The minimum dietary diversity score is 0.8 and the maximum score is 15.63 food groups * food items per day. The histogram (*Fig. 13*, A) for the first dietary diversity score shows that a high proportion of the infants have low diversity in their diet during the

introductory period and two infants have a much higher score compared to the others. Generally, the histogram is multimodal with a skew to the right.

The second dietary diversity score based on food groups in this sample has a mean (SD) of 3.6 (1.45), with a range of 4.3. The minimum value is 2, while the maximum is 6.3. This histogram (*Fig.13*, B) shows that a diversity score between 2 and 4 is most common. It is also a right-skewed graph.

The third dietary diversity score, based on the number of food items shows a mean (SD) score of 12.8 (8.06), and a range of 27.3. The minimum score is 4.0 and the maximum score is 31.3. The histogram (*Fig. 13,* C) is normally distributed with a right skew, with the majority of infants having a score between 0 and 15.

Dietary diversity score number four was based on prebiotic and probiotic foods. The mean (SD) score here was 7.8 (5.54), with a range of 16.7. The minimum value is 0 and the maximum value is 16.7. The histogram (*Fig. 13*, D) is symmetrical, but not normally distributed. The mean is very similar to the median, and the largest number of infants had a score between 7.5 and 10.

4.4. Covariates

Other factors that could impact the gut microbiome at the introduction of solid foods are the use of oral antibiotics before introduction of solid foods (n = 1), the use of probiotics before the introduction of solid foods (n = 5), GBS prophylaxis (n = 3), maternal obesity and hyperglycemia during pregnancy (n = 4). Probiotics used varied between infants and included Genestra HMF Baby B probiotics (*Lactobacillus salivaris, Lactobacillus paracasei, Bifidobacterium bifidum* and *Bifidobacterium animals ssp. Lactis*), BioGaia (*Lactobacillus reuteri* DSM 17938), or non-specified probiotics. Alpha diversity did not differ when groups were compared for exposure to antibiotics, probiotics, intrapartum antibiotics, maternal obesity, or elevated glucose levels.

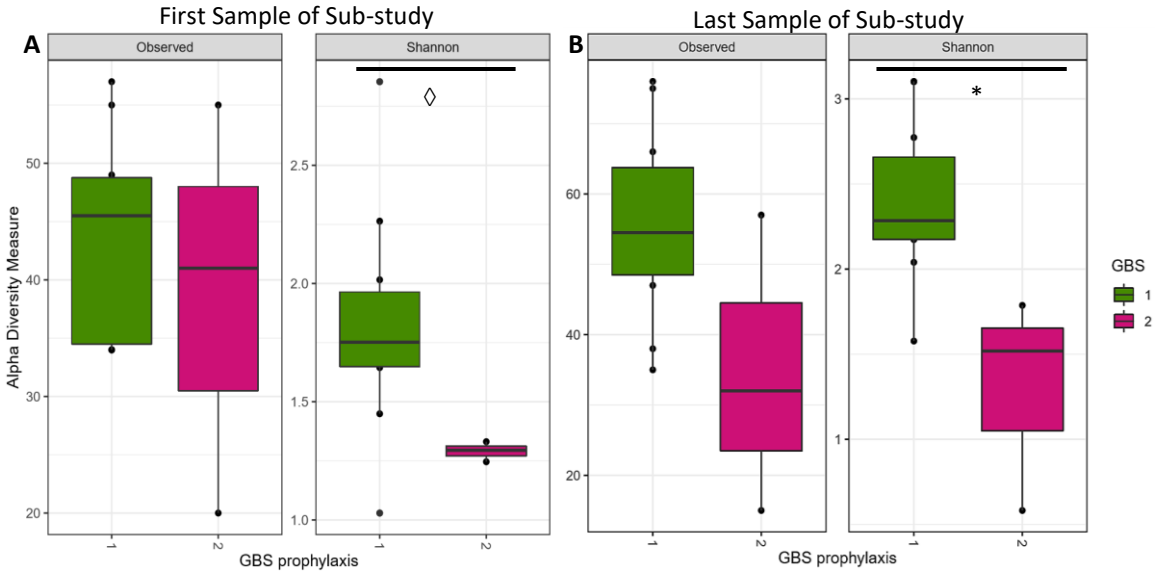


Figure 14: A: Shannon and observed alpha diversity for the covariate "GBS prophylaxis" using the first sample of the intensively sampled sub-study only. B: Shannon and observed alpha diversity for the covariate "GBS prophylaxis" using the last sample of the intensively sampled sub-study. Green (1) = not exposed to GBS prophylaxis, pink (2) = infants exposed to GBS prophylaxis.

Only GBS prophylaxis showed a trend towards significance for the Shannon index for the first sample of the sub-study period ($p < 0.1$), which became significant after the sub-study period ($p < 0.05$), where the infants exposed to intrapartum antibiotics had lower alpha diversity than those who were not exposed (*Fig. 14*). Age in days of the infant trended towards a significant association with Shannon alpha diversity ($p = 0.08$, linear mixed effects analysis). When adding both age in days and the age at introduction, the linear mixed effects model is significant ($p = 0.0012$), although age at introduction alone was not significant ($p = 0.94$). Both variables are highly correlated ($rho = 0.94$). Thus, GBS prophylaxis, age in days and age at introduction to solid foods are used as variables in the linear models for alpha diversity.

For beta diversity, PCoA plots were constructed for all covariates mentioned above and PERMANOVAs were performed for the end of the intensively sampled period (i.e. last available sample). There was a trend towards significance ($p = 0.066$) for antibiotic use, however only one infant received antibiotics, therefore this may not be helpful. Infant 02-046 received antibiotics at 4 months of age and was introduced to solid foods at 5.4 months. The other potential covariates show no clear differentiation between exposed and non-exposed infants. Most interesting is the PCoA plot for probiotic use (*Fig. 15*), because it seems as though

more infants with probiotics are on the right side of the PCoA (PERMANOVA was not significant based on the last sample, however (p = 0.55)).

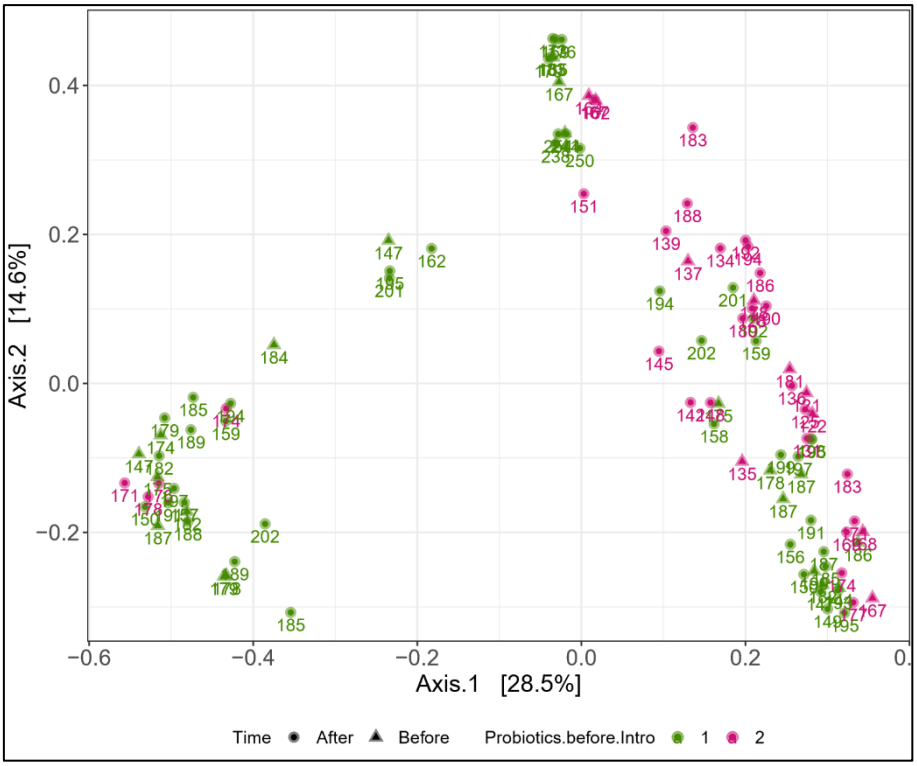


Figure 15: PCoA plot colored according to probiotic use before introduction. Green (1) = not exposed, pink (2) = exposed. Triangular symbols indicate samples collected after introduction; circular symbols indicate samples collected before introduction.

An interesting observation is that the two infants that received BioGaia probiotics showed the most movement in the population based PCoA (*Fig. 15*), and both were dominated by *Enterobacteriaceae_Escherichia/Shigella_1* before the introduction of solid foods.

4.5. Microbiome Results: Overview of Stool Sample Collection

Figure 16 shows the stool samples collected related to the time of solid food introduction, highlighted boxes were sent for processing, and green boxes represent fresh samples, which processed. An average (SD) of 7 (0.73) samples were sent for processing and were used for the analyses below.

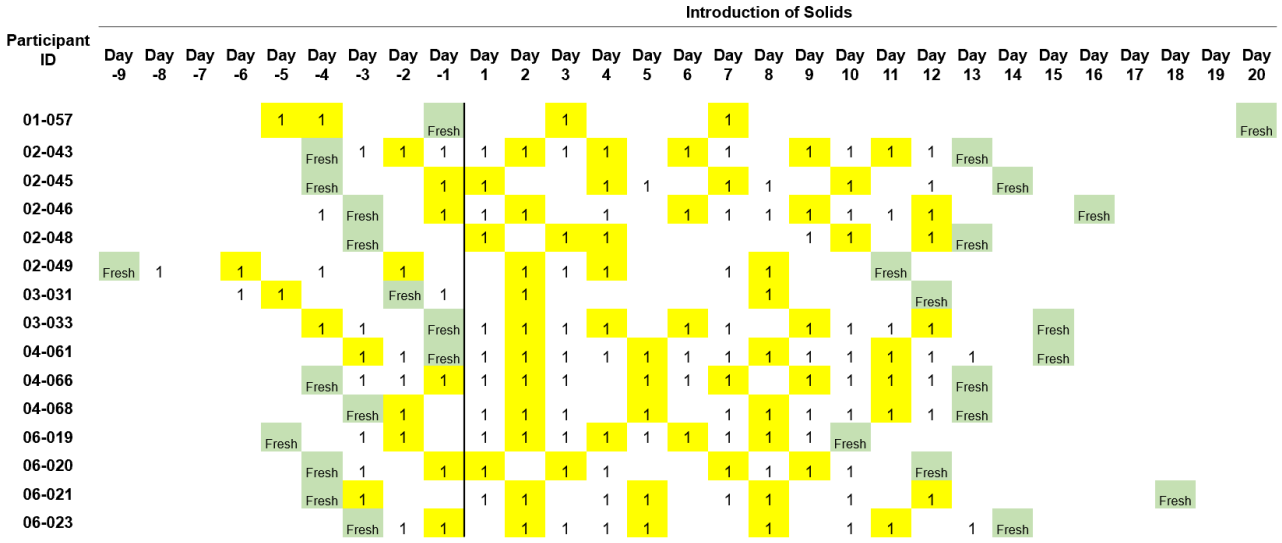


Figure 16: Overview of all stool samples collected and sent for processing in the Baby, Food and Mi study.

4.6. Microbiome Results: Before solid food introduction

4.6.1. Alpha Diversity

Before the introduction of solid foods, alpha diversity (intra-individual variability) is significantly lower than after the introduction of solid foods. The mean Shannon index for the “before” samples of the sub-study period is 1.68, while the mean Shannon index for the “after” samples of the sub-study period is 1.90 ($p = 0.047$, Welch test). Mean observed species richness for the “before” samples was

45.3 and 52.7 for the “after” samples (p = 0.0085, Welch test). Decreases in the Shannon index were observed in the days prior to solid food introduction for 8/15 of the infants of the sub-study, while four infants show increases and two remain relatively stable. On the other hand, observed species richness often shows increasing values before the introduction of solid foods (9/15). Few show decreasing observed species richness before introduction (4/15), and few remain stable (2/15). Trends in alpha diversity identified after the introduction of solid foods can be seen in **4.7.1. Alpha Diversity**. Individual graphs showing these trends can be seen in appendix B (*Figures 45 – 59)*.

4.6.2. Beta Diversity

These 15 study participants had a number of fecal samples collected prior to the introduction of solid foods (on average 3 samples, though one participant had only a single sample).

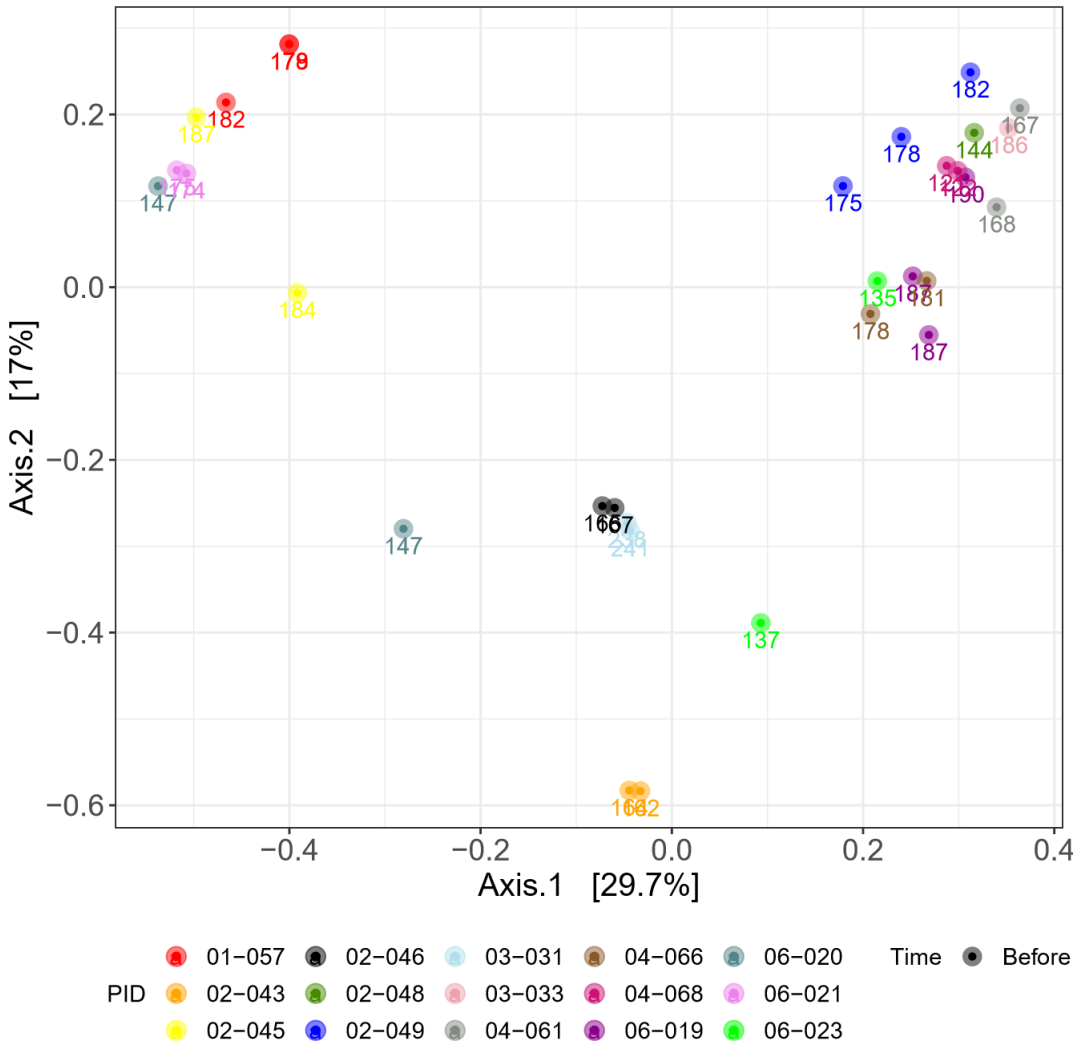


Figure 17: PCoA plot showing only the samples before the introduction of solid foods for the sub-study period only. Colors represent different individuals. Age in days of the sample can be seen underneath the plotted point.

These are shown individually in *Figure 17*. One will note that the points of the same color, representing the multiple samples for each individual before solid food introduction remain relatively close together. There are, however, a few exceptions, including 02-045, 06-020, 06-023, 02-049 and 06-019.

4.6.3. Taxonomic Distribution

As expected for a population that was fully breastfed before the introduction of solid foods, most of the infants are dominated by *Bifidobacteria*. The composition of the bifidobacterial ASVs in the samples varies by the individual. The bifidobacterial ASVs that are most present are ASV 3, ASV 2, ASV 5, ASV 8, ASV 15 and ASV 10. If *Bifidobacteriaceae_Bifidobacterium_3* is most abundant, which is suspected to be *Bifidobacterium longum*, this ASV makes up a great proportion of the ASVs present, while dominance by other bifidobacterial ASVs shows more variation in the other ASVs present in the infant gut, especially in terms of the bifidobacterial ASVs. Some infants (n = 3) do not show dominance of bifidobacterial ASVs; these are dominated by *Enterobacteriaceae_Escherichia/Shigella_1* (suspected *Escherichia coli*), *_Lachnospiraceae_18* (suspected *Ruminococcus gnavus*) or *Bacteroidaceae_Bacteroides_22* (no Greengenes match found). Although all participants were breastfed, there is high inter-individual variability in the composition of the gut microbiome prior to the introduction of solid foods.

4.7. Microbiome Results: Changes after the introduction of solid foods

4.7.1. Alpha Diversity

4.7.1.1. Individual Level

On average, Shannon alpha diversity increases over time although some infants show either decreases in Shannon alpha diversity (n = 2) or remain relatively stable (n = 2). Observed species richness shows greater variation as five infants have increasing alpha diversity, five remain stable, one decreases over time and four show a combination of increasing and decreasing alpha diversity values. Graphs showing these trends on an individual basis can be seen in the appendix (*Figures 45 – 59)*.

4.7.1.2. Population Level

Alpha diversity trends for all study visits and all individual participants can be seen in *Fig. 18*. Population-wise, alpha diversity increases over time as expected. 1-year samples have a much higher value of alpha diversity than the other time points, especially when looking at observed species richness. The Shannon index shows greater variation in alpha diversity for the different study

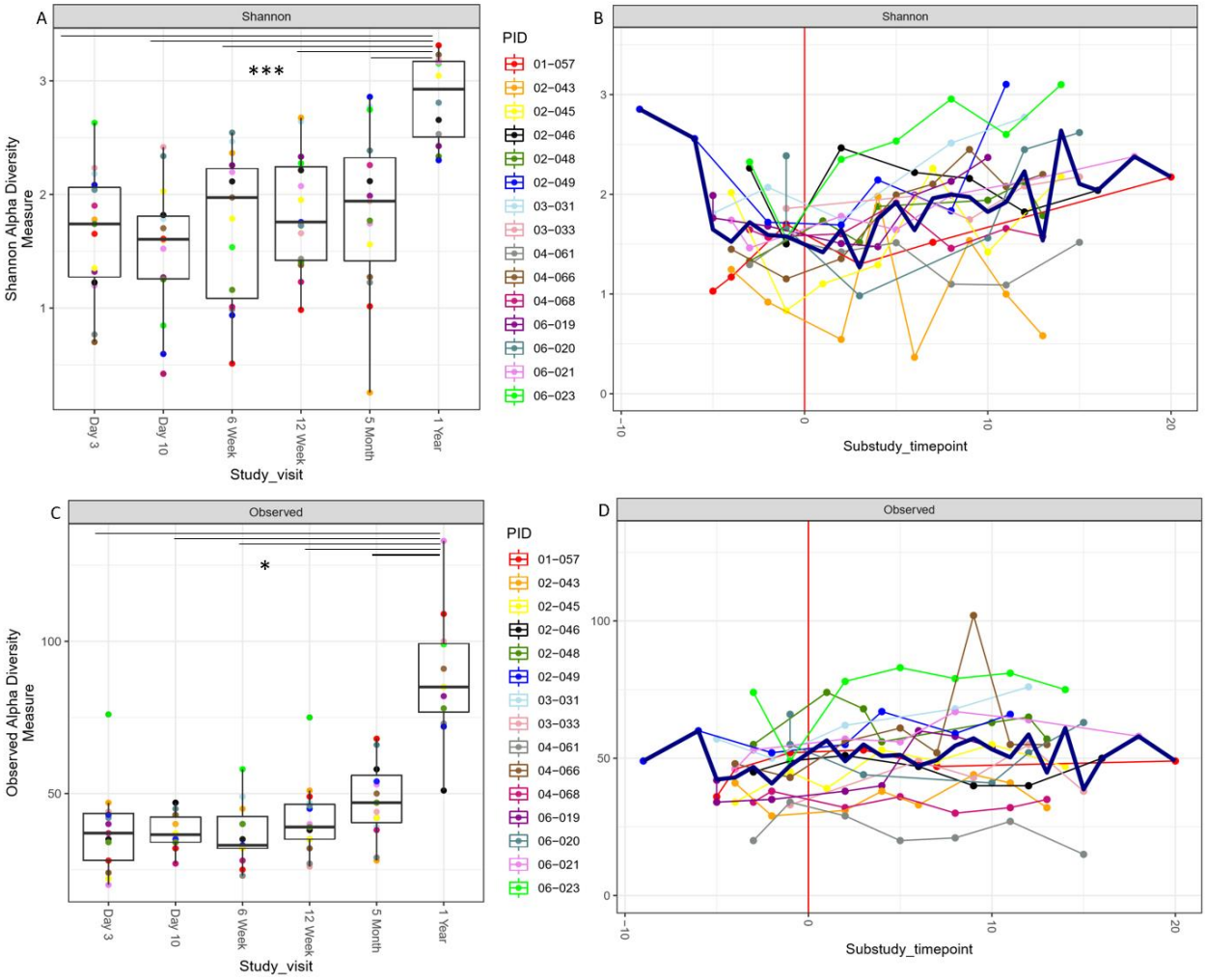


*Figure 18: **A.** Shannon alpha diversity for the study visits only, each point represents an individual infant, and the individual time points are represented as box plots. **B**: Shannon alpha diversity for the intensively sampled period, each color represents an individual infant. **C**: Observed alpha diversity for the study visits only, each point represents an individual infant, and the individual time points are represented as box plots. **D**: Observed alpha diversity for the intensively sampled period, each color represents an individual infant. The red line indicates the introduction of solid foods.*

visits. The Shannon index shows greater variation for the intensively sampled sub-study period than observed species richness, which makes sense as the Shannon index incorporates both evenness and richness of the gut microbiome, while observed species richness simply counts the number of bacterial ASVs.

For most infants, Shannon alpha diversity increases over the period of solid food introduction (n = 11), although 2 infants remained relatively stable during this period (n = 2) and 2 infants had a declined in alpha diversity. Looking at the gut microbiome over the first year of life, alpha diversity measured with the Shannon index, is higher at 1 year compared to previous visits ($p < 0.001$, Tukey test). The same relationship is visible for observed species richness ($p < 0.05$, Kruskal-Wallis Test).

Shannon diversity was related to age in days ($r = 0.55$, $p < 0.0001$) explaining 30% of total variation in the Shannon alpha diversity (adjusted R^2 – value): $Shannon\ alpha\ diversity: y = 0.0045x + 1.010$. Similarly, observed species richness was directly related to age ($r = 0.65$, $p < 0.0001$) and explained 42% of total variation of observed species richness (adjusted R^2 – value): $Observed\ alpha\ diversity: y = 0.174x + 18.7$ in linear regression models.

4.7.2. Beta Diversity

4.7.2.1. Individual Level

Individual plots of beta diversity over time are included in the appendix, *Figs. 45 – 59, appendix B*). Generally, the samples before the introduction of solid foods, i.e. day 3, day 10, 6-week and 12-week, are quite different from each other, which makes sense as the infant gut is highly dynamic, especially in early life. Additionally, the samples after introduction, i.e. the one-year sample and, if available, the two-year sample, are quite different as well. Beta diversity shows a higher degree of variation in trends for the intensively sampled time period than alpha diversity. Often, the later samples (the last one – two samples collected in the intensively sampled period) show greater differences between each other and the earlier samples, than the early samples collected. Within the intensively sampled time period, samples are closer together in terms of beta diversity, which makes sense as they were collected in a short time frame (ca. 17 days).

4.7.2.2. Population Level

Beta diversity trends for the sub-study population are shown in *Fig. 19.* There is considerable variation in beta diversity demonstrated, with less difference noted during the intensively samples period. This likely relates to the reduced number of days between sample collection as compared to the relatively longer time frame from 5 months to 1 year for example. However, even within this 2-week period, the intensively sampled late samples (collected over the last 2 days) are

considerably more diverse compared to both the before samples and the earliest samples after the introduction of solids. Changes are borderline significant, $p < 0.1$, between 1-year samples and the intensive before, early, and late samples. As well as day 10 samples and the intensive before, early, and late samples.

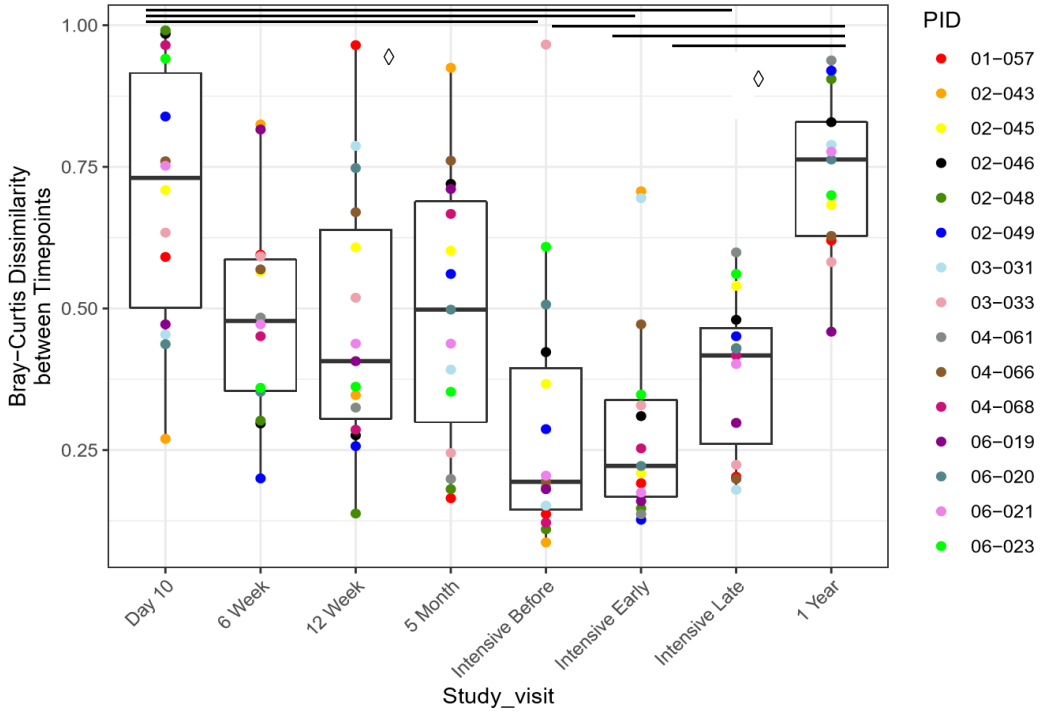


Figure 19: Changes in beta diversity for the study visit samples from the previous sample (i.e. change in beta diversity from day 3 to day 10, day 10 to 6 weeks, etc.) for all available samples and all participants. Each study visit is represented with a box plot, and individual points represent individual infants.

This PCoA plot (*Fig. 20*) of the samples collected during the introduction of solid foods, shows varying degrees of change over time. Some participants have samples that remain similar before and after introduction of solid foods, namely 02-046, 06-021, 03-031 and 01-057. Some participants mostly stay together but change a little: 02-045, 06-020, 02-048, 04-068, 04-066, 03-033 and 02-049.

Lastly, some samples show a larger amount of migration: 02-043, 06-023, 06-019 and 04-061.

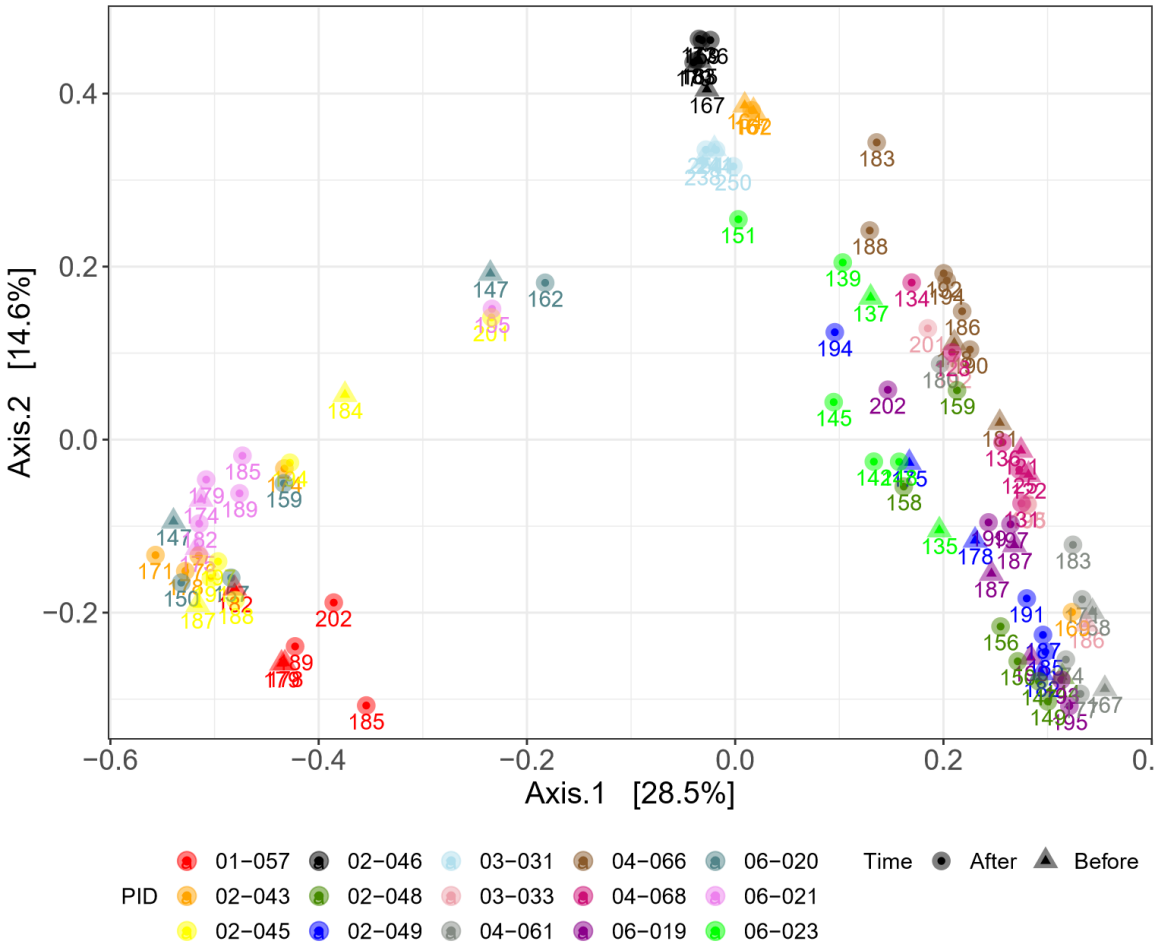


Figure 20: PCoA plot of samples for the sub-study only. The colors represent different participant IDs, and shape the time of the sample, i.e. before (triangle) or after (circle) the introduction of solid foods. Age in days of the sample can be seen underneath the plotted point.

The location on the PCoA plot for each participant is explained by their dominant bacterial ASV (see **4.7.3**.). The left side of the PCoA only has infants whose dominant bacterial ASV is *Bifidobacteriaceae_Bifidobacterium_3*, the top of the V-shape are the infants that have dominant bacterial ASVs that are not

members of the genus *Bifidobacterium*. The right side of the PCoA includes the infants that are dominated by bifidobacterial ASVs that are not *Bifidobacteriaceae_Bifidobacterium_3*, the exception here is that 06-023 is also represented on the right side of the PCoA, which is dominated by *Enterobacteriaceae_Escherichia/Shigella_1* (*Fig. 21*).

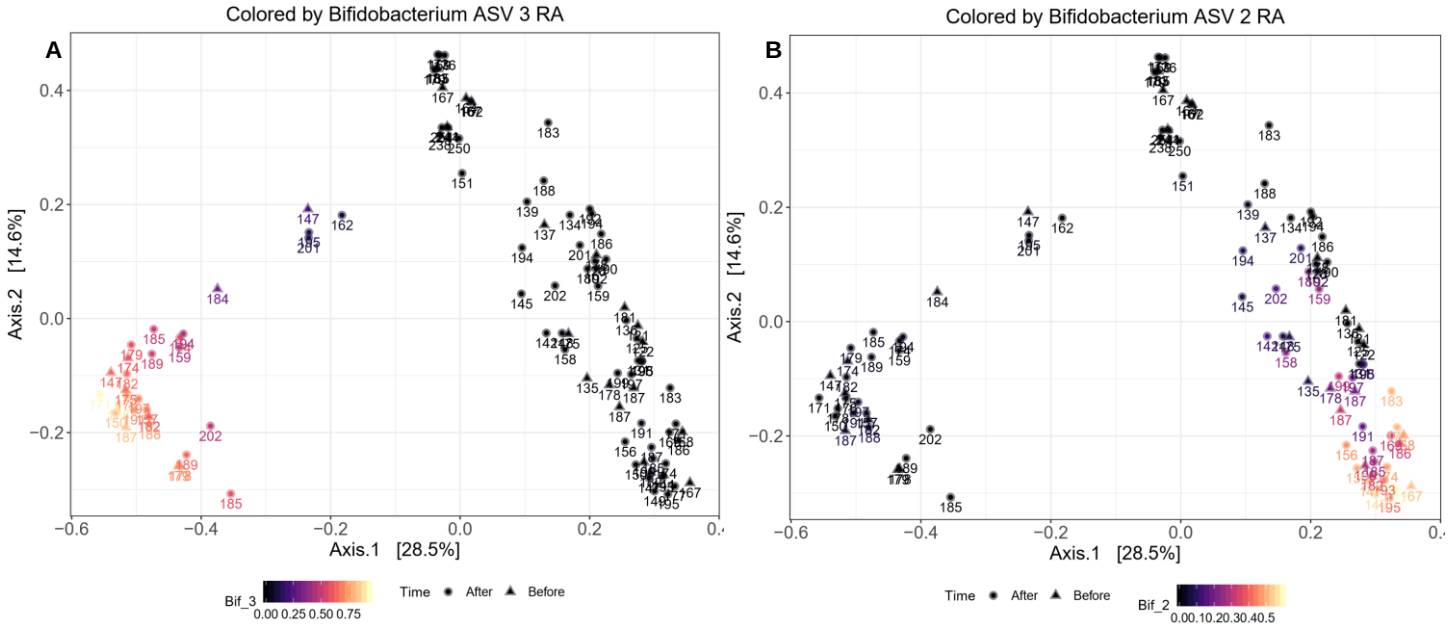


*Figure 21: PCoA plot of all the samples of the sub-study period, colored by the relative abundance of **A:** Bifidobacterium ASV 3 and **B**: Bifidobacterium ASV 2*

When looking at a PCoA plot with more samples (earlier and later, see *Fig. 90* in appendix), it is visible that later samples migrate toward the apex of the V-shape, indicating that there is a progression away from a Bifidobacterium-dominant gut as the infant ages and the infant gut matures.

4.7.3. Taxonomic Distribution

4.7.3.1. Individual Level

Overall, the infants can be sorted into three groups based on their dominant bacterial ASV throughout the study period. Group one is the *Bifiodobacteriaceae_Bifidobacterium_3* dominant infants (which seem to be mutually exclusive with *Bifiodobacteriaceae_Bifidobacterium_2*, see *Figure 21*): 01-057, 02-043[4], 02-045, 06-020 and 06-021. Group two is dominated by other Bifidobacterial ASVs: 02-048, 02-049, 04-061, 04-068, 03-033, 04-066 and 06-019. Group three is dominated by non-bifidobacterial ASVs (*_Lachnospiraceae_18, Bacteroidaceae_Bacteroides_22* and *Enterobacteriaceae_Escherichia/Shigella_1*): 02-046, 03-031, 02-043[5] and 06-023. *Bifiodobacteriaceae_Bifidobacterium_3* and *Bifiodobacteriaceae_ Bifidobacterium_2* are both suspected *B. longum* bacteria, which could indicate that these two bacteria are of different subspecies, namely ssp. *longum* and ssp. *infantis*. This is reflected in the PCoA plot shown above (**4.7.2.2.**). Additionally, over the ~ 14- day period following the introduction of solid foods dominant ASVs show decreasing abundance, bifidobacteria especially show a decrease over the study period. Previously less dominant ASVs show increasing abundance throughout the sub-study period. These results can be seen in more detail in the appendix: “Individual Analysis” (*Figs. 61 to 89*).

[4] After solid food introduction
[5] Before solid food introduction

4.7.3.2. Population Level

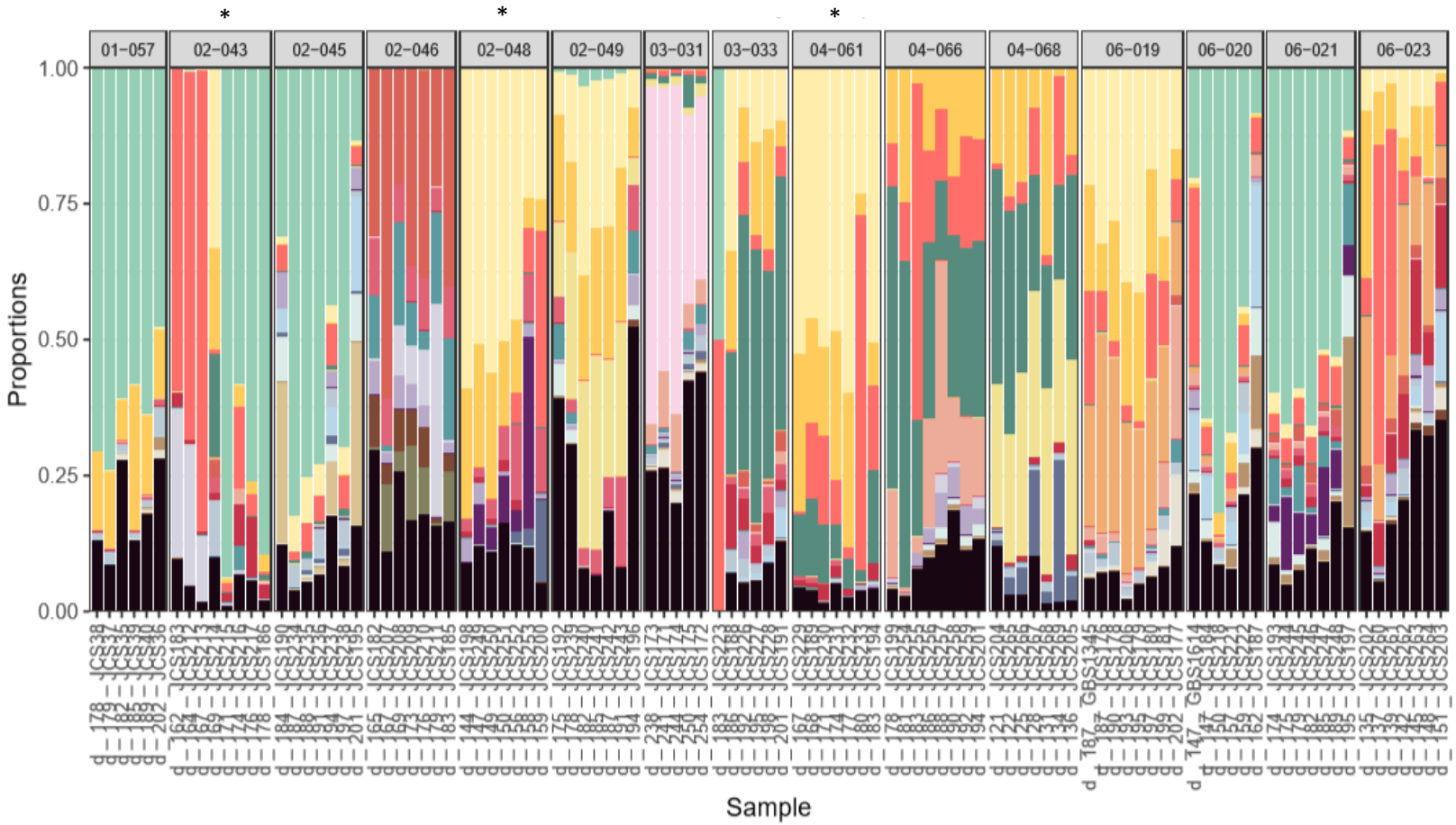


Figure 22: Taxa abundance bar chart, showing the top 25 ASVs for all sub-study samples, the samples are organized alphabetically, however this means they are generally organized by individual. Individuals annotated with an asterisk (*) have been exposed to GBS prophylaxis.

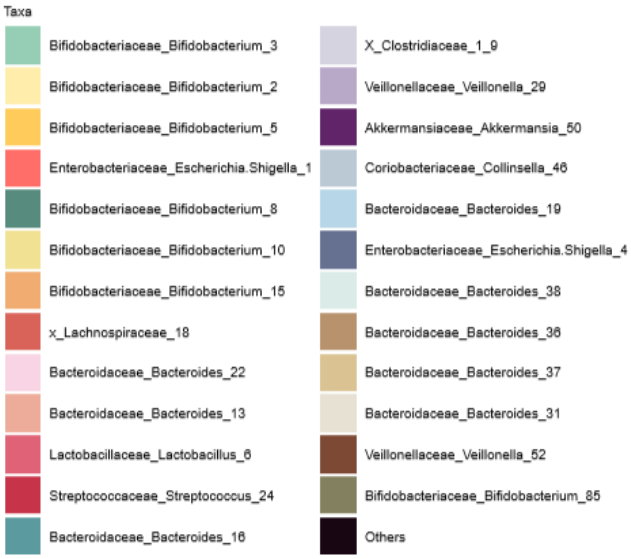


Figure 23: Legend for the taxa bar chart of Fig. 22

The key players identified from the heat maps and the taxa abundance bar chart (*Fig. 22*) are the following bacterial ASVs: *Bifidobacteriaceae_Bifidobacterium_5* (suspected *B. bifidum* using the Greengenes database), *Bifidobacteriaceae_Bifidobacterium_3* (suspected *B. longum*), *Bifidobacteriaceae_Bifidobacterium_2* (suspected *B. longum*), *Bifidobacteriaceae_Bifidobacterium_8* (suspected *B. breve*), *Bacteroidaceae_Bacteroides_22* (no match found on Greengenes), *_Lachnospiraceae_18* (suspected *R. gnavus*), *Lactobacillaceae_Lactobacillus_6* (suspected *L. zeae*), *Bacteroidaceae_Bacteroides_13* (no match found on Greengenes), *Bifidobacteriaceae_Bifidobacterium_10* (suspected *B. adolescentis*), *Bifidobacteriaceae_Bifidobacterium_15* (no match found on Greengenes), *Enterobacteriaceae_Escherichia/Shigella_1* (suspected *E. coli*) and *Enterococcaceae_Enterococcus_7* (no match found on Greengenes). These are the ten ASVs that had the highest abundance and three additional ASVs that had a relative abundance > 0.35 in more than one sample. These were used for further analyses: linear mixed effects analyses and negative binomial regressions.

As shown with the individual level analysis, infants are generally dominated by bifidobacterial ASVs, as expected from the literature. From *Fig. 24* it is visible that taxonomic distribution does not remain stable, there are changes that occur over the sub-study period. There seems to be a period of change and then a recovery towards the end of the sub-study period in terms of taxonomic distribution. This can be seen in more detail on the individual taxa bar charts (see appendix B, *Figs. 61 - 89*). The infant that shows the most diverse microbiome from bacterial ASV composition is 02-043, which shows a switch from *Enterobacteriacaea_Escherichia/Shigella_1* to *Bifidobacteriaceae_ Bifidobacterium_3* dominance, as solid food is introduced.

Previous analysis of taxonomic distribution focuses on abundance of bacterial ASVs, when looking at prevalence of bacterial ASVs for those present in at least 50% of the samples there is also a change in the most prevalent bacterial ASVs before and after the introduction of solid foods. The following ASVs are prevalent in more than 50% of the samples both before and after the introduction of solid foods: *Micrococcaceae_Rothia_243, Streptococcoaceae_Streptococcus_124, _95, _91, _24, _Lachnospiraceae_18, Enterococcaceae_Enterococcus_7, Veillonellaceae_Veillonella_29, Lactobacillaceae_Lactobacillus_6, Bifidobacteriaceae_Bifidobacterium_8, _3, _2, _5, Coriobacteriaceae_ Collinsella_46* and *Enterobacteriaceae_Escherichia/Shigella_1*. Some of these also overlap with the bacterial ASVs that have high abundance. Bacterial ASVs that are more prevalent

before the introduction of solid foods are *Streptococcaceae_Streptococcus_160, Bacteroidaceae_Bacteroides_38* and *Bifidobacteriaceae_Bifidobacterium_10*. Bacterial ASVs that are more prevalent after the introduction of solid foods are *Ruminococcaceae_Ruminoclostridium_5_206, _Clostridiaceae_1_9, Bacteroidacae_Bacteroides_36* and *_13*. This indicates that there is a decrease in prevalence of bifidobacteria and an increase in *Bacteroides*. *Figure 24* shows these bacterial ASVs in addition to their relative abundance.

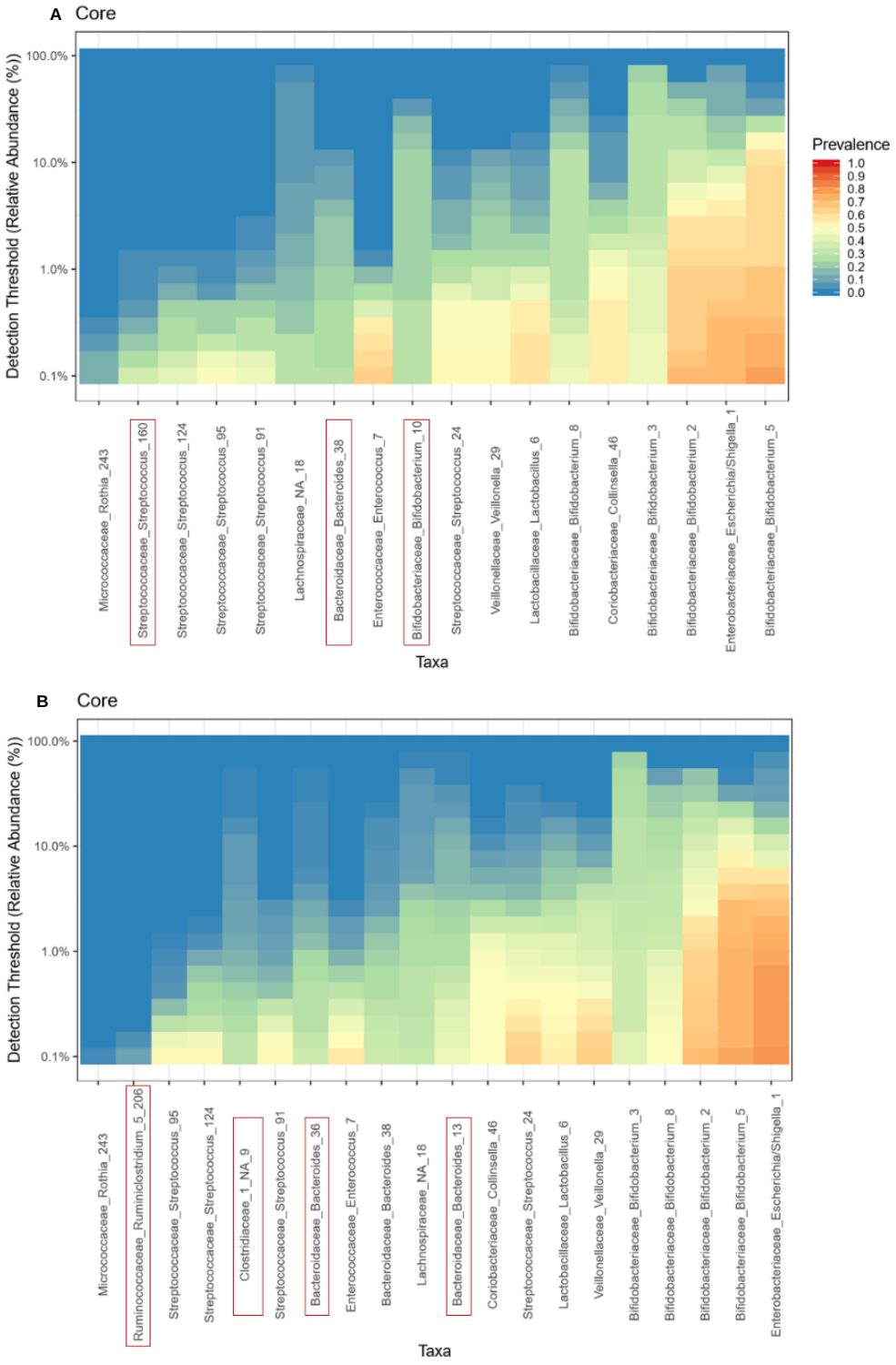


Figure 24: Heat maps showing the bacterial ASVs that are prevalent in more than 50% of the sub-study samples, in addition to their relative abundance. ASVs boxed in red differ between before and after the introduction of solid foods. ***A:*** *Before introduction of solid foods,* ***B:*** *After the introduction of solid foods.*

4.8. Integration of Nutrition into Microbiome Analysis

4.8.1. Alpha Diversity

4.8.1.1. Macronutrients

To investigate the effect of macronutrients on changes in alpha diversity, linear mixed effects analyses were performed.

Calories from carbohydrates are significantly associated with Shannon alpha diversity ($p < 0.01$, $\beta = 0.006$). This relationship is shown in *Fig. 25*. When adjusted for calories from protein and fat, calories from carbohydrates remain trending for significance ($p < 0.1$, $\beta = 0.005$). This model also includes adjustment for age in days, age at introduction, GBS prophylaxis. The unadjusted model has the best fit for the data with the smallest Akaike Information Criterion (AIC) value. Details for these models can be seen in *Table 9*:

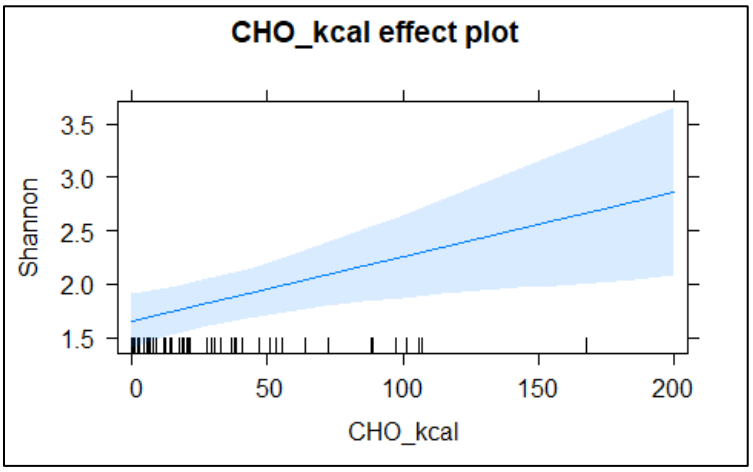


Figure 25: Effect plot showing the relationship between calories from carbohydrates and Shannon alpha diversity for the unadjusted model.

*Table 9: Output from the linear mixed effects analysis of calories from carbohydrates and Shannon alpha diversity. * = p <0.1, ** = p < 0.05, *** = p < 0.01. Adjustments are made for age in days, age at introduction and GBS prophylaxis as these were significantly associated with Shannon alpha diversity in univariate analysis.*

	Unadjusted Model: Estimate (SE)	Standard Multivariate 1: Estimate (SE)	Standard Multivariate 2: Estimate (SE)	Energy Partition Model 1: Estimate (SE)	Energy Partition Model 2: Estimate (SE)
CHO calories	0.006*** (0.002)	0.005 (0.004)	0.003 (0.004)	0.005* (0.003)	0.003 (0.003)
Intercept	1.658*** (0.124)	1.658*** (0.124)	3.232*** (0.876)	1.658*** (0.124)	3.232*** (0.876)
AIC	97.54	110.61	113.78	110.61	113.78

When delving deeper into the categories of carbohydrates, other carbohydrates (g), sugar (g) and fiber (g) were investigated. Fiber (g/d) was significantly associated with Shannon alpha diversity in the unadjusted model ($p < 0.05$, $\beta = 0.082$), and in the fully adjusted model ($p < 0.05$, $\beta = 0.113$, *Fig. 25*). The model that fits the dataset best is the unadjusted model, with the smallest AIC value of 92.28. This relationship was not seen for observed species richness. The models for other carbohydrates (g) and sugar (g) were non-significant for all models. Details about the linear mixed effects analyses can be seen in *Table 10*.

*Table 10: Output from the linear mixed effects analysis of fiber (g/d) and Shannon alpha diversity. * = p <0.1, ** = p < 0.05, *** = p < 0.01. Standard multivariate model 1 is adjusted for total energy intake and standard multivariate model two is adjusted for total calorie intake, age in days, age at introduction and GBS prophylaxis.*

	Unadjusted Model: Estimate (SE)	Standard Multivariate 1: Estimate (SE)	Standard Multivariate 2: Estimate (SE)
Fiber (g/d)	0.082** (0.034)	0.094 (0.062)	0.113** (0.058)
Intercept	1.726*** (0.115)	1.639*** (0.127)	3.330*** (0.867)
AIC	92.28	113.36	115.77

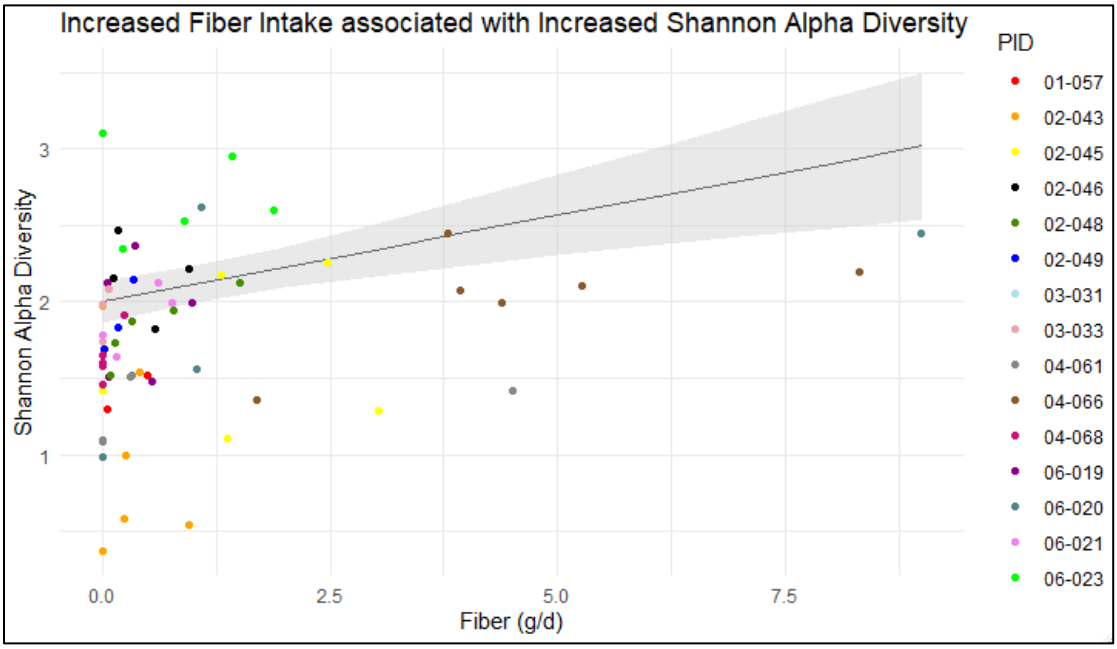


Figure 26: Relationship between fiber (g/d) and Shannon alpha diversity for the fully adjusted model, showing the individual participants and their samples.

Therefore, the driver of change for Shannon alpha diversity at the time of solid food introduction seems to be calories from carbohydrates, and especially the intake of fiber (g/d).

Calories from protein (β = 0.011) and fat (β = 0.003) trend towards significance ($p < 0.1$) in association with Shannon alpha diversity in unadjusted models. When adjusted for total energy intake and covariates are made, the relationships are not significant. The unadjusted models, however, have the lowest AIC values, which shows they have the best fit for the data. Calories from fat, protein and carbohydrates are all not significantly associated with observed species richness.

4.8.1.2. Grouped Characteristics of Dietary Intake

In addition to macronutrient intakes, the infants were grouped according to 4 characteristics or nutritional patterns: protein intake, dominant macronutrient, dominant food group, and vegetarian vs omnivorous. As demonstrated in *Figure 27A*, the mixed diet group had higher observed species richness ($p < 0.05$, Tukey test) in comparison to the fruit and vegetable group and the grain-based group had intermediate observed species richness. A similar trend was noted for Shannon alpha diversity ($p < 0.1$, Kruskal-Wallis test), it was not significant. The high protein intake group had higher observed species richness and Shannon index ($p < 0.05$, t-test) compared to the low protein group (*Figure 27B*).

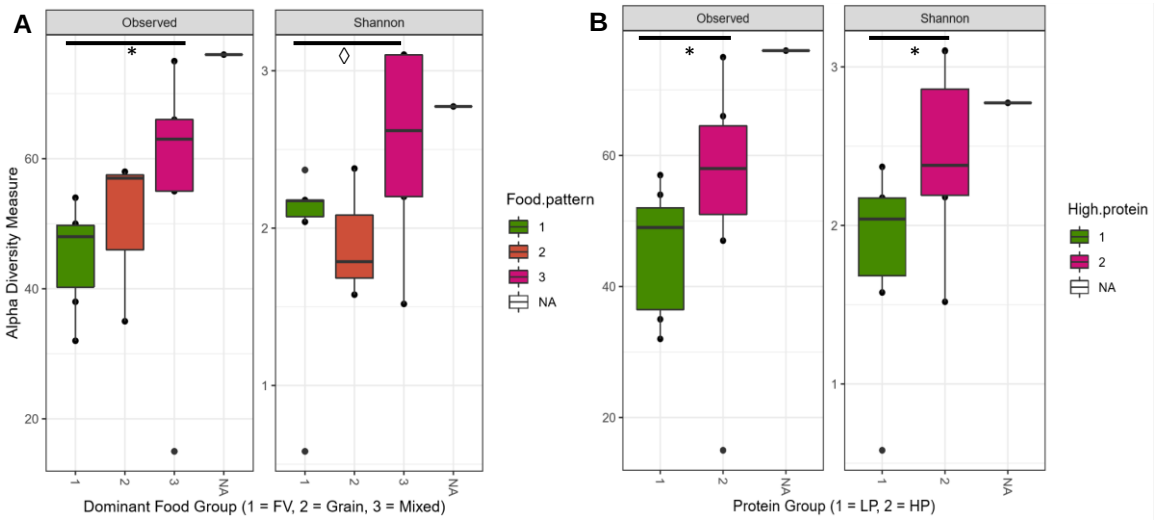


Figure 27: Alpha diversity boxplots for the "a priori" groupings. ***A:*** *dominant food group, 1 = fruit and vegetable based, 2 = grain based, 3 = mixed.* ***B:*** *protein group, 1 = low protein, 2 = high protein.*

As noted in *Figure 28*, alpha diversity by either measure did not differ based on vegetarian vs omnivorous diet (*Fig. 28, A*) or by dominant macronutrient (*Fig. 28, B*).

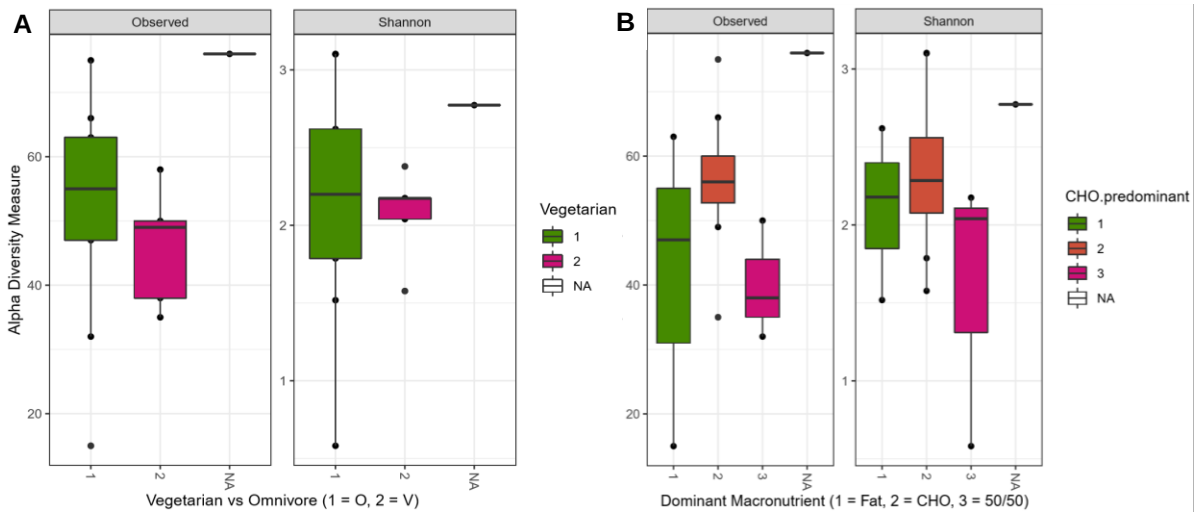


*Figure 28: Alpha diversity boxplot for the groupings according to vegetarianism (**A**), 1 = omnivorous, 2 = vegetarian; and dominant macronutrient (**B**) of the infant, 1 = carbohydrate dominant, 2 = fat dominants, 3 = 50/50 fat and carbohydrates.*

The differences in alpha diversity between groups before the introduction of solid foods were also investigated (figures not shown here, see *Fig. 92* in appendix). There were no statistically significant differences between groupings of protein, dominant macronutrient, dominant food group and vegetarianism before the introduction of solid foods, with p-values far greater than 0.05, when looking at the first sample only.

Linear mixed effects analyses that include all samples and control for participant ID do not support the results of the multiple comparisons for the dominant food groups, protein intake and vegetarianism. Though observed species richness was significantly positively associated with a carbohydrate dominant diet in unadjusted models and adjusted models that did not include GBS prophylaxis, the best fit for the model included GBS prophylaxis and this was not significant *(Table 11*). No models were significant for Shannon alpha diversity. Tables for the other food categories can be seen in *Appendix C: Tables 21 – 27*. Considering that the only models significantly associated with observed species richness are for the dominant macronutrient group, this underlines the impact of carbohydrates on alpha diversity.

*Table 11: Output from the linear mixed effects analysis of dominant macronutrient grouping and observed species richness. * = p <0.1, ** = p < 0.05, *** = p < 0.01. Adjustments are made for age in days, age at introduction and GBS prophylaxis as these were significant.*

	Model 1: Estimate (SE)	Model 2: Estimate (SE)	Model 3: Estimate (SE)	Model 4: Estimate (SE)
CHO predominant	17.465** (8.77)	17.510** (8.78)	17.626* (9.21)	15.735 (9.60)
50/50 Diet	2.463 (10.57)	3.416 (10.70)	2.990 (11.22)	3.099 (11.39)
Intercept	40.267*** (2.48)	39.711*** (7.64)	33.156 (32.09)	50.598 (32.39)
AIC	505.76	486.20	485.36	480.27

*Model 1 is unadjusted, Model 2 adjusts for total caloric intake, Model 3 adjusts for total caloric intake, age in days and age at introduction and model 4 adjusts for total caloric intake, age in days, age at introduction and GBS prophylaxis.

4.8.1.3. Dietary Diversity

Finally, we were interested in examining the relationship of the four derived dietary diversity scores with alpha diversity. Dietary diversity score 1, which included food groups and number of distinct items, was weakly related to the observed species richness ($\beta = 1.43$, $p < 0.1$), when GBS prophylaxis, total caloric intake, age in days and age at introduction to solid foods were included in the model of best fit ((AIC = 489.6) (shown in *Figure 29)*.

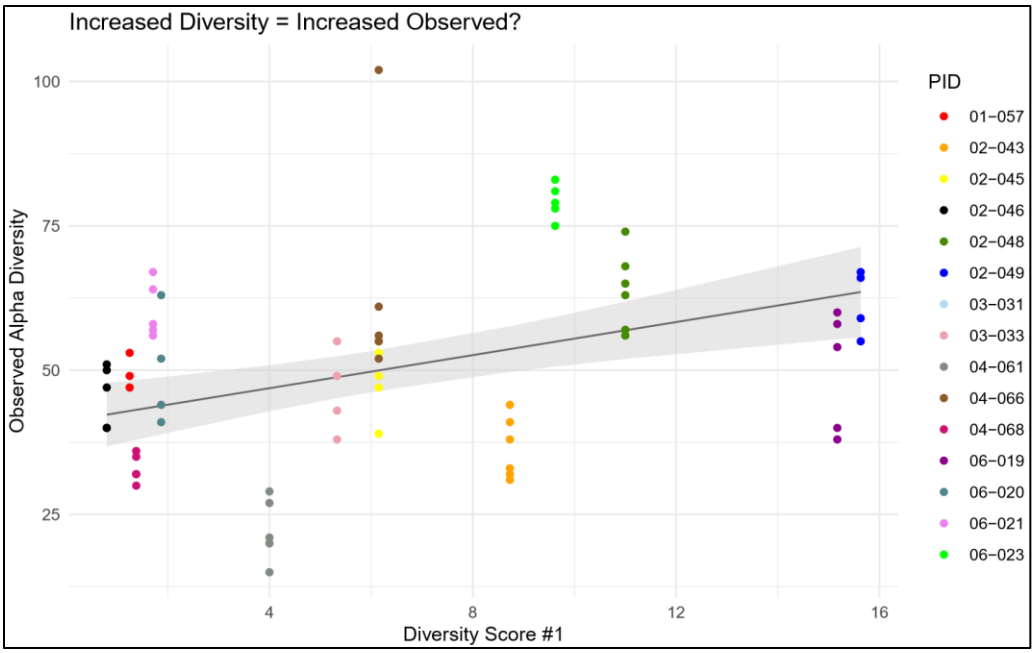


Figure 29: The relationship between dietary diversity score #1 and observed alpha diversity. Points are colored by PID. The shaded area is computed using standard error values.

Dietary diversity score 2, which included food groups only, was not related to either measure of alpha diversity.

Dietary diversity score 3, which included the number of distinct food items consumed daily over the period of solid food introduction was directly related to observed species richness ($\beta = 0.964$, $p < 0.05$). The full model, that best represented the data (AIC = 490.0) indicated this was independent of other factors influencing the alpha diversity including GBS prophylaxis, total caloric intake, age in days, age at introduction to solid foods (see *Figure 30*).

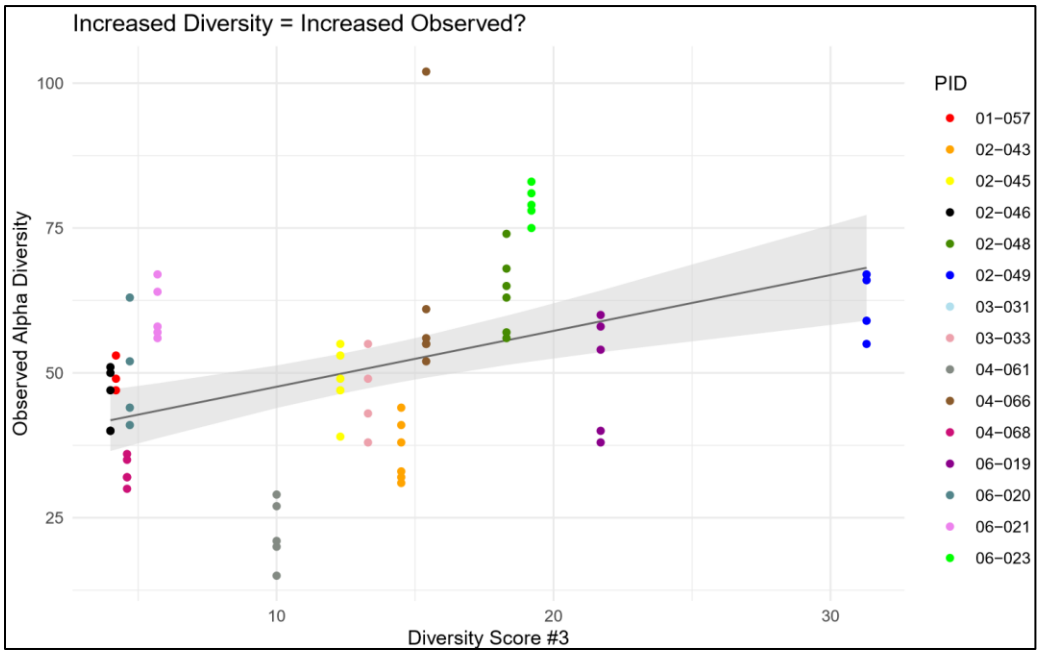


Figure 30: The relationship between dietary diversity score #3 and observed alpha diversity. Points are colored by PID. The shaded area is computed using standard error values.

The fourth diversity score considered foods that were either prebiotic or probiotic. This score was directly related to observed species richness ($\beta = 1.52$, $p < 0.05$) in the fully adjusted mixed effect model (included GBS prophylaxis, total caloric intake, age in days and age at introduction to solid foods) (see *Table 12, Figure 31*). There was a trend towards significance for the relationship between dietary diversity score 4 and Shannon alpha diversity ($\beta = 0.028$, $p < 0.1$).

*Table 12: Results of the linear mixed effects analyses for the relationship between the fourth dietary diversity score and Shannon/observed species richness. * = p <0.1, ** = p < 0.05, *** = p < 0.01. Adjustments are made for age in days, age at introduction and GBS prophylaxis as these were significant.*

Independent Variables	Shannon Alpha Diversity Unadjusted: Estimate (SE)	Shannon Alpha Diversity Standard Multivariate: Estimate (SE)	Observed species richness Unadjusted: Estimate (SE)	Observed species richness Standard Multivariate: Estimate (SE)
Diversity Score 4	0.017 (0.021)	0.028* (0.016)	1.194* (0.676)	1.515** (0.626)
Intercept	1.720*** (0.199)	3.122*** (0.785)	41.440*** (6.400)	74.665** (31.116)
AIC	110.25	108.39	517.07	488.33

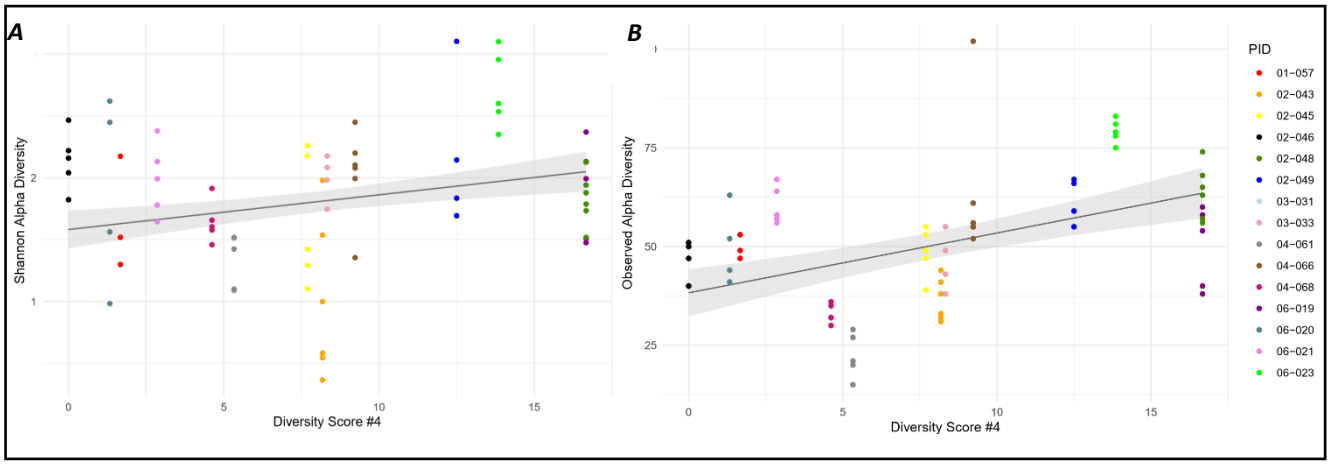


Figure 31: The relationship between dietary diversity score 4 and observed alpha diversity (A) and Shannon alpha diversity (B). Points are colored by PID. The shaded area is computed using standard error values.

The relationship between dietary diversity four and observed species richness is the strongest of the alpha diversity measures and dietary diversity

scores, however, dietary diversity scores one, three and four are all directly related to observed species richness. Only dietary diversity score four showed a trend towards significance with Shannon alpha diversity. Thus, indicating that a diet high in diversity is related to increased alpha diversity, especially if the diets have a high amount of prebiotic or probiotic foods.

4.8.2. Beta Diversity

We were also interested in examining how each of the nutritional variables might be related to the beta diversity of the gut microbial communities for these 14 participants during the introduction of solid foods. PERMANOVAs were performed for all sub-study samples, as well as the last sample of the sub-study period.

4.8.2.1. Macronutrients

In PERMANOVA analysis, the beta diversity of all samples collected during the study period, was related to calories from carbohydrates ($p = 0.009$), calories from protein ($p = 0.029$) and fiber intake ($p = 0.047$). These are relatively weak relationships, however, when considering that these p values are not corrected for repeated sampling or potential covariates. Furthermore, when we looked at the final sample only, considering that this may show the maximum influence of solid food introduction and avoids repeated sampling, beta diversity was unrelated to calories from any of the macronutrients and/or daily fiber intake (g/d). Thus,

calories from macronutrients do not seem to explain beta diversity, i.e. differences between individuals.

4.8.2.2. Grouped Characteristics of Dietary Intake

Similarly, when all sub-study samples were included in PERMANOVAs, groupings for dominant macronutrient ($p = 0.0005$), dominant food group ($p=0.006$) and vegetarianism ($p=0.0007$) were related to beta diversity of the community, i.e. differentiation on the PCoA plot (*Figure 95, appendix C*). However, none of these were significant when looking at the final sample only (*Figure 94, appendix C*). Beta diversity was not different by protein grouping for all samples or for the final sample alone. Thus, food categories do not seem to explain differences between individuals in this study.

4.8.2.3. Dietary Diversity

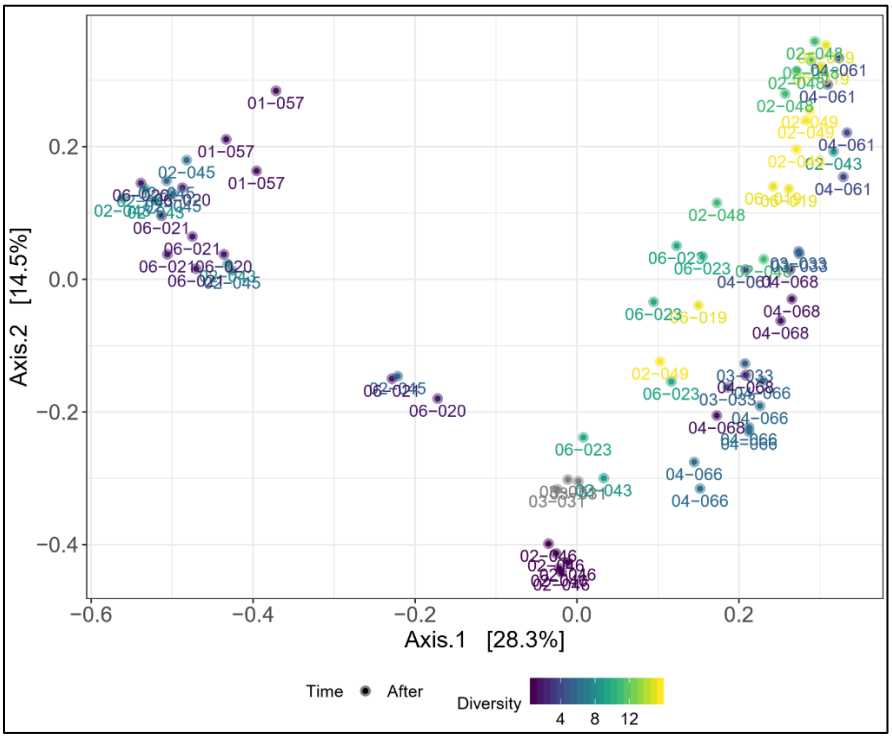


Figure 32: PCoA plot for the after samples of the sub-study period, colored according to the first dietary diversity score.

Dietary diversity score 1 was also related to beta diversity as demonstrated in *Fig. 32* and confirmed with PERMANOVA (p = 0.0002). Similarly, when using the final sample only, no differences were seen (p = 0.598).

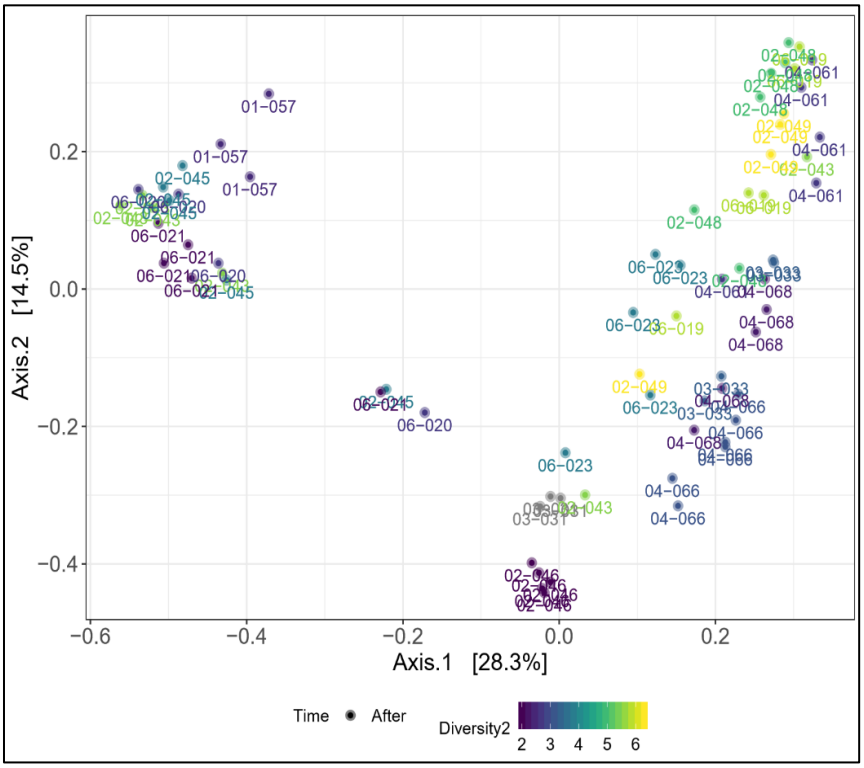


Figure 33: PCoA plot for the after samples of the sub-study period, colored according to the second dietary diversity score.

From the PCoA (*Fig. 33*) it is visible, that the samples with a higher value for diversity score 2 are generally localized more to the right side of the PCoA than the left, which is confirmed by a PERMANOVA ($p = 0.0002$). Performing a PERMANOVA using one sample only (i.e. the last sample) proves insignificant, however ($p = 0.598$).

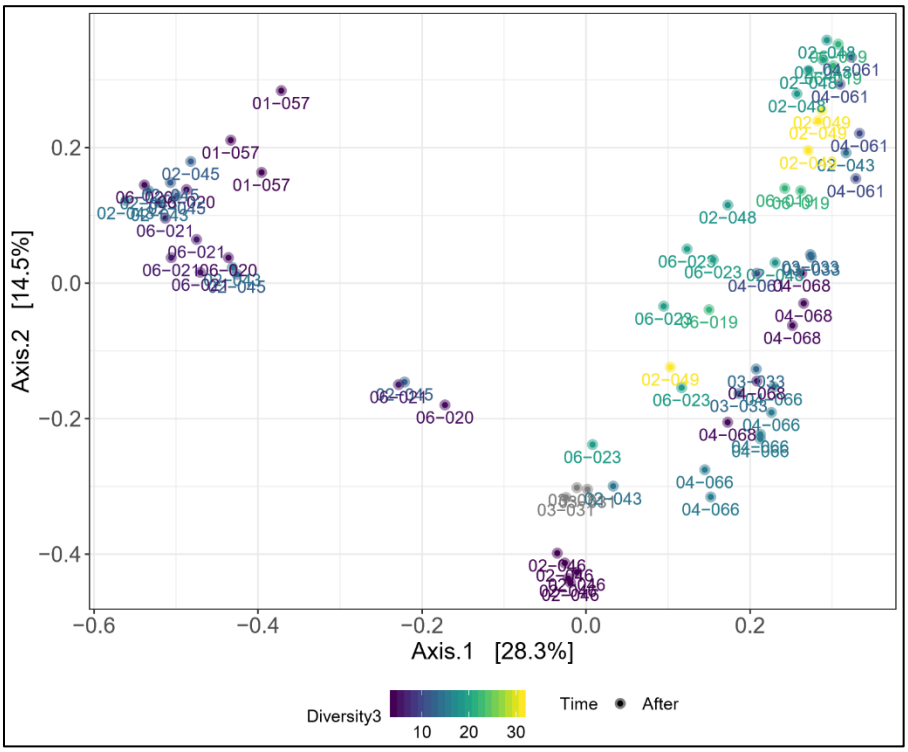


Figure 34: PCoA plot for the after samples of the sub-study period, colored according to the third dietary diversity score.

Similarly, in the PCoA plot in *Fig. 34*, those with highest diversity score 3, based on food items, are different than those with the lowest number of items (p = 0.0001, PERMANOVA). No significance was seen when using the final sample only (p = 1).

From the PCoA (*Fig. 35*) it is visible, that when all samples are included, those consuming more prebiotic and probiotic foods, i.e. with a higher dietary diversity score 4, are generally localized more to the right side of the PCoA than the left, which is confirmed by a PERMANOVA (p = 0.0001). No significance was identified with only the final sample (p = 0.126).

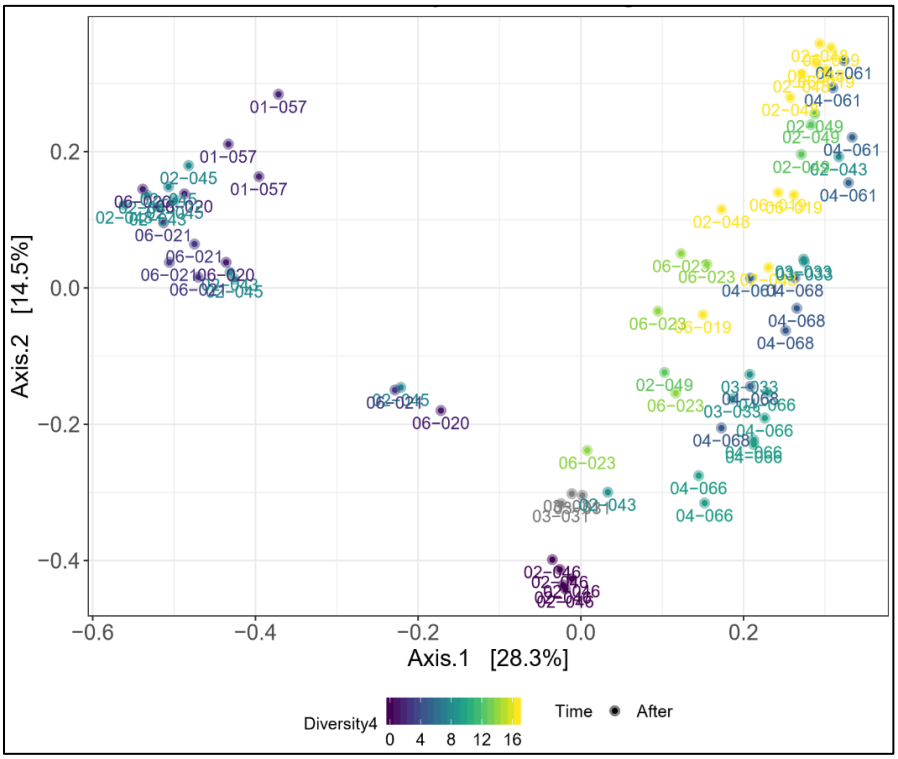


Figure 35: PCoA plot for the after samples of the sub-study period, colored according to the fourth dietary diversity score.

In addition to looking at the separation of samples on the PCoA plot based on dietary diversity scores, dietary diversity in relation to the degree of change during the sub-study period was investigated. After visually inspecting the PCoA plots once colored by PID and those colored by dietary diversity scores, a potential association between magnitude of dietary diversity and the magnitude of movement throughout the sub-study period became apparent. This was tested using a Spearman correlation and a univariate linear regression model. All the dietary diversity scores are statistically significantly correlated with the degree of movement in the PCoA plot. The results of the Spearman correlation tests can be seen in *Table 13,* and the corresponding plots in *Fig. 36*.

Table 13: Results of the Spearman correlation tests for the relationship between the degree of movement on the PCoA plot and the dietary diversity scores.

Dietary Diversity Score	Correlation Coefficient (rho)	p-value
1	0.63	0.015
2	0.64	0.015
3	0.61	0.021
4	0.60	0.024

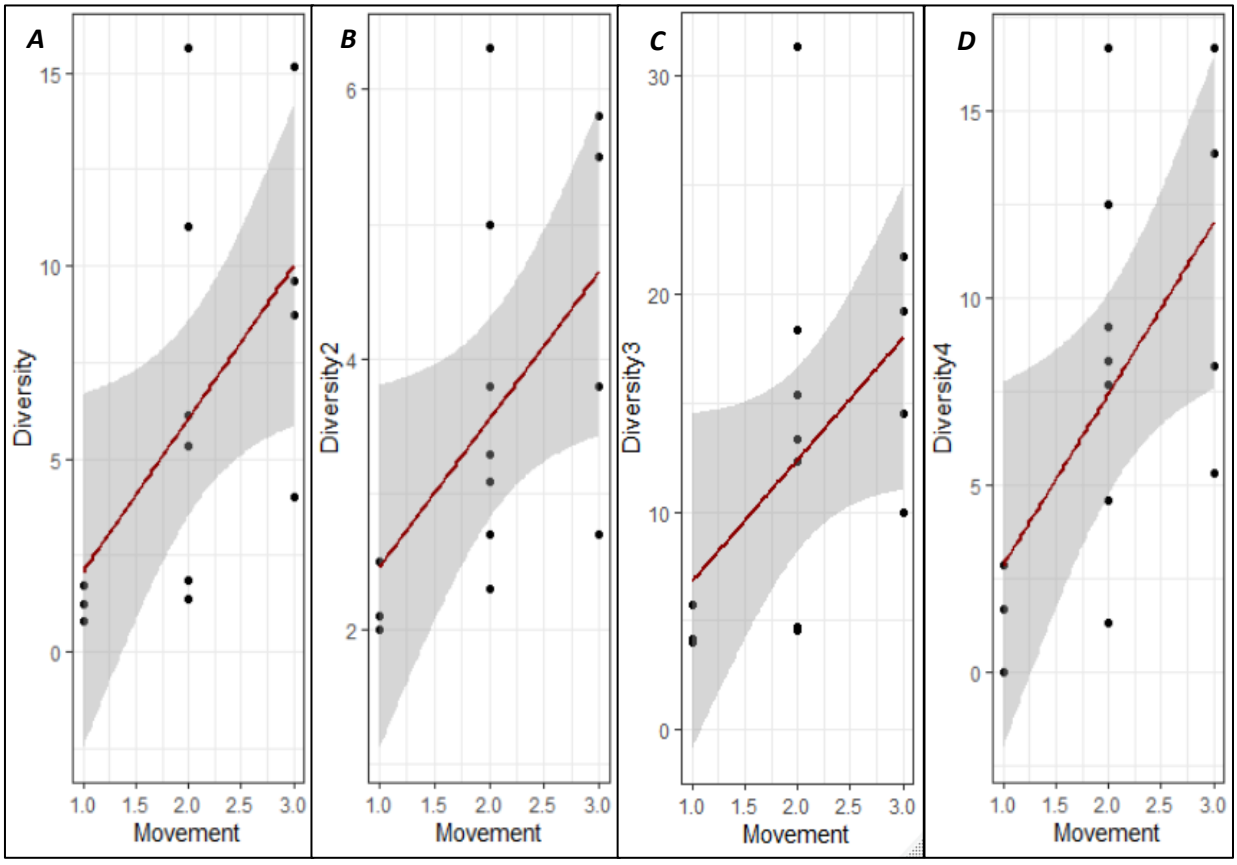


Figure 36: Plots showing the relationship between the degree of movement in the population PCoA plot and the different diversity scores. A: score 1, B: score 2, C: score 3, D: score 4.

In addition to Spearman correlations investigating the relationship between the dietary diversity scores and degree of change over the intensively sampled period, simple linear regressions were performed to determine how much variance of the movement on the PCoA plot is explained by the dietary diversity scores. The simple linear regressions were only significant for dietary diversity scores 1 (β = 3.96, p = 0.0326, R^2 = 0.27), 2 (β = 1.09, p = 0.0422, R^2 = 0.24) and 4 (β = 4.58, p = 0.0225, R^2 = 0.31) (*Table 31, appendix C*), and assumptions for performing a linear regression were met. As with the separation on the PCoA, the strongest relationship can be seen for the fourth dietary diversity score and the degree of change over the sub-study period.

4.8.3. Taxonomic Distribution

4.8.3.1. Macronutrients

Macronutrients and their relationship to counts of specific bacterial taxa were examined in the models seen in Methodology: *Table 5*. Unadjusted models of calories of the specific macronutrient with total caloric intake from solid foods or calories of the specific macronutrient with calories from the remaining macronutrients (energy partition model) were first studied. Further, these models were adjusted for age (days). Calories from fat were inversely related to the presence of *Enterococcaceae_Enterococcus_7* in both adjusted models (multivariate model 1 (partially adjusted): β = -0.040, p = 0.029; multivariate model 2 (fully adjusted): β = -0.042, p = 0.026). Calories from fat trended to significance in the energy partition models (model 1 (unadjusted): β = -0.015, p = 0.073; model 2 (fully adjusted): β = -0.015, p = 0.01). The partially adjusted multivariate model and energy partition model 1 had the best fit for the data, with the smallest AIC value. Calories from fat were also negatively associated with the presence of *Bacteroidaceae_Bacteroides_22*, in the first energy partition model (β = -0.080, p = 0.045), and trended towards significance in the second

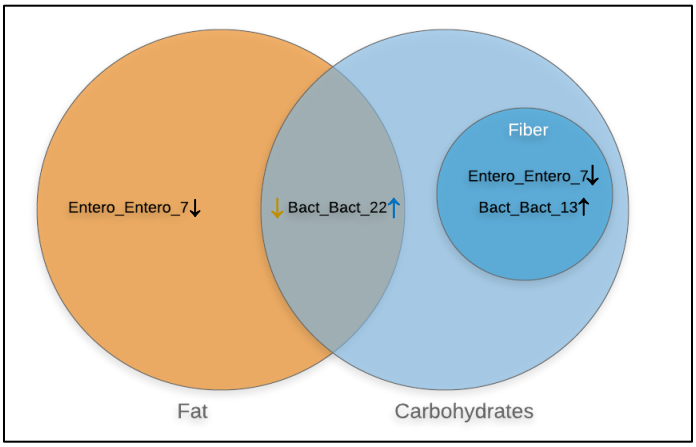


Figure 37: Key bacterial ASVs and their association with the calories from macronutrients (significant associations only). Arrows indicate the direction of the relationship.

energy partition model (β = -0.064, p = 0.083). These two models also have the best fit for the data. This suggests that infants with a higher intake of fat had reduced counts of *Enterococcaceae_Enterococcus_7* and *Bacteroidaceae_Bacteroides_22*.

Calories from carbohydrates were weakly associated with the presence of *Bacteroidaceae_Bacteroides_22* in standard multivariate model 1 (β = 0.078, p = 0.078) with best fit (AIC = 149.85).

Fiber (g/d) intake was positively associated with the presence of *Bacteroidaceae_Bacteroides_13* in the unadjusted energy partition model (β = 0.52, p = 0.042; best fit model). It was also negatively associated with *Enterococcaceae_Enterococcus_7* in both the partially adjusted multivariate model (β = -0.52, p = 0.009) and the fully adjusted model (β = -0.56, p = 0.006). Thus, higher carbohydrate intake and higher fiber intake are associated with higher counts of Bacteroides ASVs. Full tables for the regression analyses can be seen in *appendix C: Tables 32 - 36*.

In summary, even over the first days of solid food introduction, evidence of the influence of fat and carbohydrate intake on taxonomic abundance in the gut microbiome is seen. Further, fiber has opposite influence of fat on the abundance of *Enterococcaceae_Enterococcus_7* and both fiber and carbohydrate intake are associated with higher counts of Bacteroides ASVs.

4.8.3.2. Grouped Characteristics of Dietary Intake

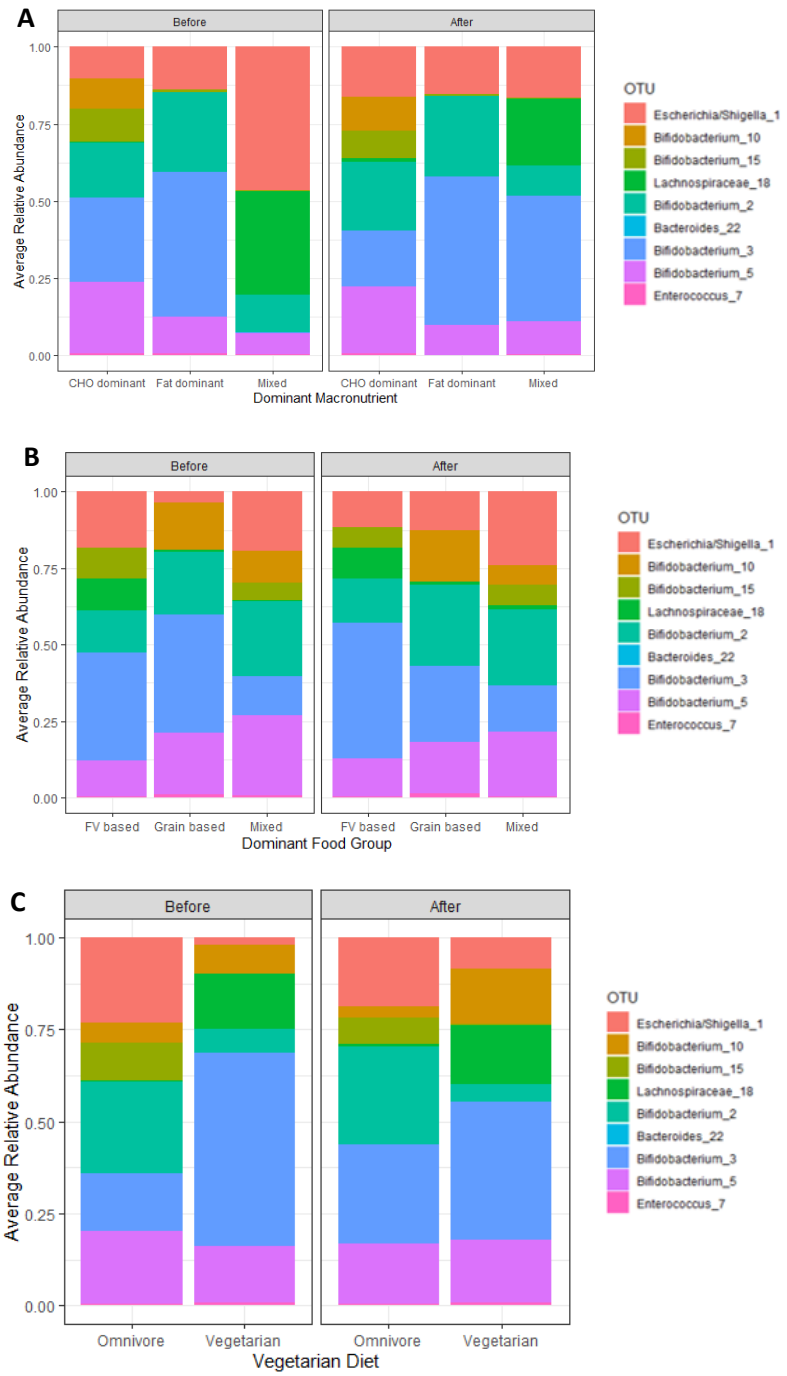


*Figure 38: ASV level bar graphs for the food categories, where there is a significant effect in the negative binomial regression analysis. **A**: Dominant macronutrient, **B:** Dominant food group, **C:** Vegetarian diet. No significant associations were found for protein grouping. These taxa bar charts exclude abundances of non-significant ASVs.*

There were no taxonomic differences noted in comparing the high and low protein intake groups. However, when grouped by dominant macronutrient (fat vs carbohydrate or mixed), a diet that is carbohydrate-based is weakly associated with the presence of *Bifidobacteriaceae _ Bifidobacterium _5* (suspected *B. breve)* (β = 2.6, p = 0.053), and is also associated with *Bifidobacteriaceae _ Bifidobacterium_15* (β = 2.6, p = 0.002), but these

relationships are no longer significant when covariate age (days) is included in the model (*Figure 38, A*).

A diet with mixed fat and carbohydrate intake was positively associated with *_Lachnospiraceae_18* (suspected *R. gnavus*) presence in unadjusted (β = 2.6, p = 0.064) and all adjusted models (1: β = 5.7, p = 0.003, 2: β = 5.3, p = 0.001, 3: β = 5.7, p = 0.003).

In *Figure 38, B*, the influence of food groupings is illustrated. *Enterococcaceae_Enterococcus_7* is positively associated with both a grain-based diet and a mixed diet. For the grain-based diet the significant associations are in the unadjusted model (β = 3.1, p = 0.035) and the first standard multivariate model (β = 3.2, p = 0.032). The mixed diet is positively associated with *Enterococcaceae_Enterococcus_7* counts in the unadjusted model (β = 2.6, p = 0.042) and the first standard multivariate models (β = 2.6, p = 0.042). Other more prevalent ASVs trended towards significant differences in the grain-based diet compared to the other groups, but none were significantly different in either unadjusted or adjusted models. These ASVs are *Bifidobacteriaceae_Bifidobacterium_10* (suspected *B. adolescentis*), *Bifidobacteriaceae_Bifidobacterium_*15 and *Bacteroidaceae_Bacteroides_*22.

The infants consuming a vegetarian diet had less *Bifidobacteriaceae_Bifidobacterium_2* (suspected *B. longum*) prevalence in unadjusted and the multivariate model with the best fit (unadjusted: β = -2.5, p =

0.023, standard multivariate 1 (partially adjusted): β = -3.4, p = 0.034.) The relationship with *Enterobacteriaceae_Escherichia/Shigella_1* is trending towards significance for the unadjusted model (β = -1.5, p = 0.102), the first standard multivariate model (β = -3.0, p = 0.059) and the fully adjusted model (β = -3.0, p = 0.053). Full tables for the regression analyses can be seen in *appendix C*: *Tables 37 - 47*.

4.8.3.3. Dietary Diversity

All four dietary diversity scores were significantly (or trending towards significantly) positively associated with *Bifidobacteriaceae_Bifidobacterium_2, _5,* _15 and *Bacteroidaceae_Bacteroides_22,* in negative binomial regression analyses. Interestingly, although the scores are calculated differently to reflect diversity in food groups, food items or in pre- and probiotic foods, infants with all higher dietary diversity scores had higher counts of the more abundant bifidobacterial ASVs. Full tables for the regression analyses can be seen in the appendix, part C: *Tables 48 - 64*.

5. Discussion

This research presents the relationship between dietary intake at the introduction of solid foods and the gut microbiome in a small group of healthy infants, who were breastfed and were born at full-term by vaginal delivery. The primary objective of this project was to analyze the nutritional patterns in this cohort at the time of the introduction of solid foods. The second objective was to characterize gut bacterial dynamics during this period, and its relationship with the foods introduced, in order to understand the impact of solid foods on gut microbiota development.

In this study we observed a change in the gut microbiome during the introduction of solid foods, including alpha diversity, beta diversity and community We observed a change in the characteristics of the gut microbiome during the period of solid food introduction. Changes in alpha diversity, beta diversity and taxonomic distribution were noted. Of the three macronutrients, carbohydrate intake seems to be the driver of change for Shannon alpha diversity, with fiber intake (g/d) specifically acting as the significant factor increasing alpha diversity. We developed several dietary diversity scores based on food groups, food items and inclusion of pre-probiotic foods and noted that increased dietary diversity was associated with increased observed species richness, as well as the differences observed in beta diversity between individuals and the degree of change occurring in the gut microbiome over the 2 week period after the introduction of solid foods. Of the most abundant ASVs, or key bacterial ASVs, in this cohort:

_Enterococcus_7, *Bifidobacterium_15, _2, _5, _Lachnospiraceae_18, Bacteroides_22 and _13* were significantly related to nutritional variables, which underlines the changing bacterial community of the infant gut microbiome as new foods are introduced.

Nutrition. The first step of this thesis was to investigate patterns in the introduction of solid foods in terms of calories from macronutrients, food groups introduced and other nutritional variables, including the developed dietary diversity scores. The median age at introduction of solids in this cohort was 5.59 months; this is marginally lower than 6 months as recommended by the WHO (Organization, 2018). This timing of solid food introduction is consistent with multiple studies in the Netherlands and the U.S.A., where solid foods were introduced between 4 and 6 months of age (Kuo *et al.*, 2011; Wang *et al.*, 2019). In the Epifane study from France however, only 50% of infants were introduced to solid foods between 4 and 6 months of age, 13% were introduced to solid foods before 4 months, and 33% after (Boudet-Berquier *et al.*, 2017).

The energy intake from solid foods over this initial two-week period was low – estimated to be approximately 4.7% of daily energy requirements only. In contrast, Friel and colleagues described an average caloric intake of 118 kcal/d, an average over a period of four days, at four months of age, compared to 50 kcal/day in our study. Friel et al. noted higher intakes of protein (3g vs 1.9 g) and carbohydrate (22 g vs 6.7 g) and, interestingly, lower fat intake (2 g vs 5.6 g) than participants in this study (Friel *et al.*, 2010). While neither study meets the

recommendations provided by WHO for slightly older infants, which stipulates that infants aged 6 – 8 months consume 200 kcal per day from solid foods (Organization, 2002), the *Baby, Food and Mi* study showed much lower intakes of calories in comparison to the other Canadian study. The differences in macronutrient intakes may be due to the fact that the study by Friel et al. was performed in 2003, and most caregivers seemed to follow a more traditional approach to infant feeding, which has shown to decrease the amount of fat ingested (Erickson, 2015). Caloric intake increased over the sub-study period, with carbohydrates and fat making up most of the diet, which is expected as requirements of carbohydrates and fats exceed those for protein.

More infants consumed a fruit and vegetable dominant diet (6/14) than a grain-based diet (3/14) and 5/14 consumed a mixed diet. Most infants were following an omnivorous diet at the introduction of solid foods (9/14), rather than a vegetarian or vegan diet (5/14). Many studies have shown that in developed and developing countries, the first foods introduced to infants are cereals, fruit and vegetables (Boudet-Berquier *et al.*, 2017; Lopes *et al.*, 2018; Wang *et al.*, 2019). Differences between studies were more apparent with the introduction of animal products. In France, fish and meat were introduced after six months and eggs after one year, while in Brazil introduction of protein-rich foods, meat and meat alternatives occurred earlier (Boudet-Berquier *et al.*, 2017; Lopes *et al.*, 2018). While the succession of foods introduced in the *Baby, Food and Mi* study was similar to other studies, it most closely resembled the findings from the Brazilian

population. This is especially interesting, because the Brazilian infants were closer to fulfilling the recommendations from the WHO, as they introduced protein-rich foods more frequently. This also provides iron, an important nutrient for infant development (USDA, 2018). Overall, the nutritional patterns of the *Baby, Food and Mi* study resemble those of previously published cross-sectional and short-term studies. The cross-sectional studies and a single longitudinal study were conducted throughout the first year of life, often based on caregiver recall, and data for the 4 – 6 month time points were assessed in the discussion above, as the infants in our study were introduced to solids around this time, however, energy intake in this study was much lower than in others.

Infant gut microbiome. The changes in the gut microbiome during the introduction of solid foods were first investigated independent of nutritional intake as the first objective of this thesis. In this study, alpha diversity increased over time from the first days after birth to one year of age. This is expected from the literature, as alpha diversity increases over the first three years of life as the infant gut microbiota transitions to a more adult-like composition (Arrieta *et al.*, 2014; Backhed *et al.*, 2015). In the two-week sub-study period, Shannon diversity increased, while observed richness remained relatively stable. Shannon alpha diversity incorporates both ASV richness and evenness, while observed species richness only calculates the richness of bacterial ASVs. Generally high diversity is considered “good” (Sprague, 2018). Observed richness and phylogenetic diversity (PD) are both richness estimates, so these are comparable. PD has been studied

in two studies at the introduction of solid foods. In one study, PD increased throughout the introductory period in one infant at four months. In the second study, no change in PD was noted between the age of four and six months in a cohort of 120 infants (Koenig *et al.*, 2011; Pannaraj *et al.*, 2017). Interestingly, in a study of 13 infants followed over one year, a decline in bacterial richness throughout the introductory period was noted and the authors attributed this change to a loss of rare bacterial taxa (Valles *et al.*, 2014). This could also explain the stability of observed richness that was seen in the *Baby, Food and Mi* study, as rare taxa may have been lost, while other counts of bacterial phylotypes (ASVs) increased. The results from this study may also be more similar to the Pannaraj study as both studied a larger cohort of infants ranging in age from four to six months. Although alpha diversity is utilized often in microbiome studies, it is prone to error and usually underestimates true alpha diversity in a sample, due to the limitations of sample depth leading to decreased detection limits of bacterial ASVs. In spite of this, it is a widely used metric that provides a general indication of bacterial diversity in the gut (Willis, 2019).

Beta diversity was observed across the first year of life in these 15 infants. Samples collected during the two-week period of intensive sampling were less different from one another than the one-year samples were from samples collected at 5 months and earlier. The PCoA plot with samples over the first year illustrated that later samples were closer to the apex of the plot with increasing age, suggesting a shift to a more similar microbial community. This is expected from the

literature, as beta diversity decreases with age, reducing the inter-individual variability observed in the gut microbiota composition of infants (Backhed *et al.*, 2015). The placement of the infant samples on the PCoA was dependent on the dominant bacterial ASVs, which could be grouped as *Bifidobacterium* ASVs and non-*Bifidobacterium* ASVs. *Bifidobacterium* dominance was common in infants (12/15), and the samples were largely dominated by ASV 3 or other bifidobacterial ASVs. *Bifidobacterium* ASVs 2 and 3 are among the most abundant ASVs of the infant gut in this cohort. Both ASVs are suspected *B. longum*, which suggests that they are different sub-species, namely ssp. *longum* and ssp. *infantis*. Interestingly, *Bifidobacterium* ASVs 2 and 3 did not co-exist in the infants of this cohort, suggesting that these are competitive sub-species. However, previous studies have suggested that these *B. longum* sub-species digest different substrates, with *B. longum* ssp. *longum* better suited to digest plant oligosaccharides, and *B. longum* ssp. *infantis* better suited to digest HMOs (Lee and O'Sullivan, 2010). A dominance of *Bifidobacteria* is expected in these infants, as they were breast fed. HMOs in breast milk support the growth of *Bifidobacteria*, and breastfed infants are generally dominated by Actinobacteria and Firmicutes (Azad *et al.*, 2013). As infants aged, there was a transition away from a *Bifidobacterium*-dominant gut microbiome in our cohort. This is also consistent with previous studies (Cresci and Bawden, 2015), as the presence of *Bifidobacteria* is lower in adults.

In this study, key bacterial ASVs were identified and these included ASVs from the following genera: *Bifidobacterium, Bacteroides, Lachnospiraceae,*

Lactobacillus, Escherichia/Shigella and *Enterococcus*. Bacteria from these families have been shown to be highly abundant in other studies looking at infants aged around 4 months (Azad *et al.*, 2013; Backhed *et al.*, 2015), which confirms the reproducibility of the results shown here. Previous studies have also shown a high abundance of other bacterial families, including *Veillonellaceae*, *Clostridiaceae*, *Erysipelotrichacea*e and *Streptococcaceae*, as well as the genera of *Ruminococcus,* that were not among the key bacterial ASVs identified in our cohort. In this study, however, they were among the most prevalent ASVs (in >50% of the sub-study samples – before and after the introduction of solid foods). We identified an inverse relationship between the phyla Actinobacteria and Firmicutes, as well as Actinobacteria and Bacteroidetes over time in the infants of this study. The abundance of Bifidobacterial ASVs decreased and were replaced by other ASVs, usually of the phylum Firmicutes/Bacteroidetes. Conversely other studies have reported increases of Verrucomicrobia (Azad *et al.*, 2013), which we did not. Different positions for the same infant over time on the PCoA plot were identified during the intensively sampled period, consistent with changes occurring in the community structure of the infant microbiota. This has been demonstrated by other studies examining the introduction of solid foods and the gut microbiome (Pannaraj *et al.*, 2017; Valles *et al.*, 2014). These studies, however, did not examine which solid foods were being introduced as was done in objective 2 of this study.

Nutrition and the infant gut microbiome. The second objective was to evaluate the relationship between characteristics of nutritional intake (caloric value

of macronutrient intakes, descriptive characteristics relating to macronutrient or food groups and dietary diversity scores) and the microbiome during the introduction of solid foods.

In the *Baby, Food and Mi* study, carbohydrates and more specifically fiber were significantly associated with increased alpha diversity. In a study performed in mice, a diet with low fiber led to reductions in alpha diversity (Sonnenburg *et al.*, 2016), and another study found that a diet high in fruit, vegetables and fiber resulted in increased bacterial richness and diversity in adult humans (Jandhyala *et al.*, 2015). A study looking at general dietary quality, based on the Healthy Eating Index (HEI) also demonstrated that a high-quality diet, especially the consumption of whole grains and vegetables, increased Shannon alpha diversity (Laitinen and Mokkala, 2019). Interestingly, this suggests that fiber influences the infant gut in a similar manner to its effect on adults – even during the earliest exposure to solid foods. Few studies have looked at the impact of dietary diversity and the gut microbiome in both the adult and the pediatric population. Increased dietary diversity in the *Baby, Food and Mi* study was associated with increased alpha diversity, as well as the degree of change during the sub-study period and the placement on the PCoA plot for beta diversity. Previous studies in adults also showed that alpha diversity was positively associated with dietary diversity (Claesson *et al.*, 2012; Heiman and Greenway, 2016), suggesting that the infant gut microbiome reacts in a similar way to the adult gut microbiome. In infants, it seems that increased dietary diversity is related to less stability, i.e. more change

during the introductory period. The opposite, however, was seen in a study in human adults, where increased dietary diversity is correlated with higher stability of the microbiome (Johnson *et al.*, 2019). This may be due to the fact that the infant gut microbiome is still developing as new foods are introduced that the gut bacteria have not been confronted with previously. In contrast, in the adult population the stability noted in the adult gut microbiome was attributed to bacteria being accustomed to exposure to a variety of nutrients.

In previous research, introduction of solid foods was associated with increased *Bacteroides, Ruminococcus* and decreased *Escherichia* abundances (Valles et al., 2014). Koenig and colleagues reported an increase of Bacteroidetes, known to digest plant polysaccharides, after the introduction of peas (Koenig *et al.*, 2011). Apart from these two studies, little is known about the impact of dietary choices on gut bacterial dynamics at the introduction of solid foods.

We observed greater increases in *Bacteroides/*Bacteroidetes with higher carbohydrate intake in the *Baby, Food and Mi* study as both carbohydrate intake and fiber intake were positively associated with *Bacteroides* ASVs. The consumption of fat had an inverse relationship with the consumption of carbohydrates for the *Bacteroides* ASV 22, and fat calories and fiber intake had a negative association with *Enterococcus* ASV 7. In the dominant macronutrient category, a diet high in carbohydrates was positively and significantly associated with two *Bifidobacterium* ASVs, which could be due to the increased fiber intake, as fiber is known to act as a prebiotic (Holscher, 2017). A diet with equal

percentages of carbohydrates and fat was positively associated with *_Lachnospiraceae_18*. Interestingly, protein intake, evaluated as either caloric amount or category (high vs. low), was not significantly associated with any of the key bacterial ASVs. This may be because protein intake remained low throughout the study period. In adults, dietary protein was associated with altered gut microbiome composition, especially in terms of metabolite concentrations from protein degradation (Beaumont *et al.*, 2017; Davila *et al.*, 2013). A vegetarian diet in this study was significantly negatively associated with *Bifidobacterium* ASV 2 and trending for *Escherichia/Shigella* ASV 1, which is similar to the findings of another study, that reported lower counts of *Bifidobacteria* and *Enterobacteriaceae* in adults consuming vegetarian diets compared to adults consuming omnivorous diets (Zimmer et al., 2012). However, the impact of a vegetarian diet on *Bifidobacterium* abundance remains controversial, since other studies have shown higher counts of Actinobacteria in vegetarians (De Filippo *et al.*, 2010).

Dietary diversity scores were calculated based on number of food groupings, number of food items and number of pre- and pro-biotic foods. High dietary diversity scores were associated positively with *Bifidobacterium_15, _2, _5* and *_3* (trending), as well as *Bacteroides_22*. It is interesting that the same bacterial ASVs are affected but may reflect that larger number of items for example is likely to also be associated with more food groups and with pre- and pro-biotic foods. Regardless, this finding suggests that dietary diversity in any form, whether due to food groups or food items, is beneficial to a healthy gut microbiome, as

Bifidobacteria have many beneficial effects on the human host and are therefore desired in the gut community. They are also often constituents of probiotics, indicating that they have positive effects on the human gut (O'Callaghan and Van Sinderen, 2016).

Long-term implications. The introduction of solid foods is one factor impacting the development of the gut microbiome during infancy, in addition to mode of delivery, early infant feeding, antibiotic use and the environment, amongst other factors mentioned earlier. These early-life influences could have long-term health implications, as the infant microbiome is developing at this time to a more permanent adult-like state by three years of age. An individual's microbiome is then particular to the individual and generally stays similar into adulthood, i.e. the introduction of solid foods starts the trajectory toward the diversity and composition of the gut microbiota in the adult. A healthy gut microbiota has been associated with beneficial effect on the immune system and the metabolism of the host (Jandhyala *et al.*, 2015; Young, 2012). High alpha diversity has been shown to be good for health, as high alpha diversity is seen less in disease states, while lower levels of diversity have been seen in pathological/inflammatory states, for example in obesity (Turnbaugh and Gordon, 2009). Therefore, establishing a healthy gut microbiota in early life may prove important to combat diseases that are more prevalent later in life. Thus, our findings suggest that a diet high in fiber and with high dietary diversity can increase alpha diversity at a young age; additionally, microbes associated with these nutritional patterns could be important for long-

term health. This may be because a high carbohydrate, plant-based diet increases the production of short-chain fatty acids (SCFAs) (De Filippo *et al.*, 2010). SCFAs have a range of functions in the human body, including reduction of gut pH, improvement of mineral absorption, central appetite regulation, maintenance of gut barrier integrity and reduction of inflammation (Alexander *et al.*, 2019; Chambers *et al.*, 2018). Also, interestingly, the infant gut microbiome reacts similarly to the adult gut microbiome in terms of alpha diversity and what is considered healthy for adults (high fiber and high dietary diversity) leads to increased alpha diversity in infants. Only the influence of this pattern on the stability of the microbiome differs in adults and infants and this may be due to the early stage of infant microbiome development.

Strengths and Limitations. This study has a few limitations that should be acknowledged. Firstly, the study population is relatively small with only 15 infants, and 14 with dietary information. This means that the statistical power for the analyses is lower than it would have been with more infants and suggests are findings are exploratory. Since numerous results still reach statistical significance this suggests that the associations seen here are quite strong. Another limitation is that only a period of two weeks is examined after the introduction of solid foods, so changes in diet beyond this and how these changes impact the gut microbiome over the long term have not been analyzed. This could be of importance as the contribution of solid foods to total energy intake increases and may therefore have greater impacts on the gut microbiome than are captured in this study. Considering

the participant burden of completing daily food diaries for two weeks, exceeding this time period may however have not been feasible. Another limitation is that in this study the highest reliable taxonomic rank attained is the ASV, so the species or strain level of these bacteria are unknown, which limits the predictability of the functions of the ASVs that are changing throughout the sub-study period.

One of the greatest strengths of this study is that it is of a longitudinal nature, with intensive sampling. This means that changes occurring can be seen on a day-to-day basis, giving a comprehensive understanding of the changing infant gut microbiome at this time allowing for the exploratory nature of this project. Additionally, the study population is very homogeneous, enabling consideration of the exposures from solid foods with a relatively small sample size. The caregivers in the study are very engaged, which means that the food diary entry is probably accurate, even over a two-week period. Lastly, the *Baby, Food and Mi* study is part of a larger research consortium, namely “The intersection of gastrointestinal microbial communities, diet, and health (GI-MDH Study)”, and a sister study is being performed in the Netherlands. Findings in that population will be compared to these, enabling further understanding of the influence of solid food introduction on the infant gut microbiome.

6. Conclusion

In conclusion, this study demonstrates that the introduction of solid foods has an impact on the developing infant gut microbiome and that nutritional choices influence the changes occurring. A high intake of fiber and high dietary diversity are associated with higher alpha diversity, and dietary diversity increases the degree of change occurring over the sub-study period. Certain nutritional decisions also impact the community structure of the healthy infant gut. Interestingly, these findings are similar to observations in adults, underlining the importance of a healthy diet throughout the life-course, especially as reduced diversity has been linked to diseases and conditions later in life, such as obesity. Additionally, this study shows how susceptible the infant gut microbiome is to change, even with low amounts of available substrate. As little research has investigated the impact of dietary choices at the time of introduction of solid foods, this study highlights the contribution of another factor impacting the development of the gut microbiome in early life.

Further research in the form of metagenomics and metabolomics is needed to understand the whole ecosystem of the infant gut microbiome. Metagenomic analysis of the bacterial ASVs in this study would enable improved understanding of the functions of the bacteria, explaining why some of the ASVs are increasing, while others decrease during the introduction of solid foods, as the available substrates for bacteria are changing. Metabolomic analysis could be used to confirm which substrates are being digested, as bacterial metabolites are specific

to certain nutrients. This would contribute to a bigger picture of the changes occurring in the infant gut at the time of introduction of solid foods. It would also be prudent to examine the impact of solid food introduction on the gut microbiome in a larger population to determine if the relationships seen here are reproducible on a larger scale. This might also allow for other statistical approaches that require larger sample sizes, for example network analysis, which facilitates the understanding of community dynamics.

Overall, this study contributes new knowledge to the research topic of the infant gut microbiome and early dietary choices, which is insufficiently researched in the literature.

Works Cited

(ECHO), C. o. E. C. O. (2017). Facts and figures on childhood obesity, Vol. 2018, World Health Organization.

Alexander, C., Swanson, K., Fahey, G. and Garleb, K. (2019). Perspective: Physiologic Importance of Short-Chain Fatty Acids from Nondigestible Carbohydrate Fermentation. *American Society for Nutrition,* **10,** 576 - 589.

Arboleya, S., Binetti, A., Salazar, N., Fernandez, N., Solis, G., Hernandez-Barranco, A., Margolles, A., de los Reyes-Gavilan, C. and Gueimonde, M. (2012). Establishment and development of intestinal microbiota in preterm neonates. *FEMS Microbiol Ecol,* **79,** 763-772.

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

Azad, M., Konya, T., Maughan, H., Guttman, D., Field, C., Chari, R., Sears, M., Becker, A., Scott, J. and Kozyrskyj, A. (2013). Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. *CMAJ,* **185,** 385-394.

Backhed, F., Ding, H., Wang, T., Hooper, L., Koh, G., Nagy, A., Semenkovich, C. and Gordon, J. (2004). The gut microbiota as an environmental factor that regulates fat storage. *PNAS,* **101,** 15718-15723.

Backhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Li, Y., Xia, Y., Xie, H., Zhong, H., Khan, M. T., Zhang, J., Li, J., Xiao, L., Al-Aama, J., Zhang, D., Lee, Y. S., Kotowska, D., Colding, C., Tremaroli, V., Yin, Y., Bergman, S., Xu, X., Madsen, L., Kristiansen, K., Dahlgren, J. and Wang, J. (2015). Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host Microbe,* **17,** 690-703.

Ballard, O. and Morrow, A. (2013). Human milk composition: nutrients and bioactive factors. *Pediatr Clin North Am,* **60,** 49-74.

Bartram, A., Lynch, M., Stearns, J., Moreno-Hagelsieb, G. and Neufeld, J. (2011). Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. *Appl Environ Microbiol. Am Soc Microbiol,* **77,** 3846-3852.

Bazanella, M., Maier, T., Clavel, T., Lagkouvardos, I., Lucio, M., Maldonado-Gomez, M., Autran, C., Walter, J., Bode, L., Schmitt-Kopplin, P. and Haller, D. (2017). Randomized controlled trial on the impact of early-life intervention with bifidobacteria on the healthy infant fecal microbiota and metabolome. *Am J Clin Nutr,* **106,** 1274-1286.

Beaumont, M., Portune, K., Steuer, N., Lan, A. C., V., Audebert, M., Dumont, F., Mancano, G., Khodorova, N., Andriamihaja, M., Airinei, G., Tome, D., Benamouzig, R., Davila, A., Claus, S., Sanz, Y. and Blachier, F. (2017). Quantity and source of dietary protein influence metabolite production by gut microbiota and rectal mucosa gene expression: a randomized, parallel, double-blind trial in overweight humans. *Am J Clin Nutr***,** 1005-1019.

Betoko, A., Charles, M., Hankard, R., Forhan, A., Bonet, M., Saurel-Cubizolles, M., Heude, B. and Lauzon-Guillain, D. (2013). Infant feeding patterns over the first year of life: influence of family characteristics. *Eur J Clin Nutr,* **67,** 631-637.

Blanton, L., Charbonneau, M., Salih, T., Barratt, M., Venkatesh, S., Ilkaveya, O., Subramanian, S., Manary, M., Trehan, I., Jorgensen, J., Fan, Y.-M., Henrissat, B., Leyn, S., Rodionov, D., Osterman, A., Maleta, K., Newgard, C., Ashorn, P., Dewey, K. and Gordon, J. (2016). Gut bacteria that rescue growth impairments transmitted by immature microbiota from undernourished children. *Science,* **351**.

Boudet-Berquier, J., Salanave, B., de Launay, C. and Castetbon, K. (2017). Introduction of complementary foods with respect to French guidelines: description and associated socioeconomic factors in a nationwide birth cohort (Epifane survey). *Maternal & Child Nutrition,* **13**.

Brown, A., Jones, S. and Rowan, H. (2017). Baby-Led Weaning: The Evidence to Date. *Curr Nutr Rep,* **6,** 148-156.

Callahan, B., McMurdie, P. and Holmes, S. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene analysis. *ISMEJ***,** 2639-2643.

Callahan, B., McMurdie, P., Rosen, M., Han, A., Johnson, A. and Holmes, S. (2016). DADA2: High resolution sample inference from Illumina amplicon data. *Nat Methods,* **13,** 581-583.

Canada, H. (2006). Dietary Reference Intakes.

Canadian_Government. (2019). Canada's food guide, Vol. 2019.

Candido, F., Valente, F., Grzeskowiak, L., Moreira, A., Rocha, D. and Alfenas, R. (2017). Impact of dietary fat on gut microbiota and low-grade systemic inflammation: mechanisms and clinical implications on obesity. *International Journal of Food Sciences and Nutrition*.

Chambers, E., Preston, T., Frost, G. and Morrison, G. (2018). Role of Gut Microbiota-Generated Short-Chain Fatty Acids in Metabolic and Cardiovascular Health. *Current Nutrition Reports,* **7,** 198-206.

Chang, L. and Neu, J. (2015). Early Factors Leading to Later Obesity: Interactions of the Microbiome, Epigenome, and Nutrition. *Curr Probl Pediatr Adolesc Health Care,* **45,** 134-142.

Claesson, M., Jeffery, I., Conde, S., Power, S., O'Connor, E., Cusack, S., Harris, H., Coakley, M., Lakshminarayanan, B., O'Sullivan, O., Fitzgerald, G., Deane, J., O'Connor, M., Harnedy, N., O'Connor, K., O'Mahony, D., Van Sinderen, D., Wallace, M., Brennan, L., Stanton, C., Marchesi, J., Fitzgerald, A., Shanahan, F., Hill, C., Ross, P. and O'Toole, P. (2012). Gut microbiota composition correlates with diet and health in the elderly. *Nature*.

Cresci, G. and Bawden, E. (2015). The Gut Microbiome: What we do and don't know. *Nutr Clin Pract,* **30,** 734-746.

Davani-Davari, D., Negahdaripour, M., Karimzadeh, I., Seifan, M., Mohkam, M., Masoumi, S., Berenjian, A. and Ghasemi, Y. (2019). Prebiotics: Definition, Types, Sources, Mechanisms, and Clinical Applications. *Foods,* **8,** 92.

David, L., Maurice, C., Carmody, R., Gootenberg, D., Button, J., Wolfe, B., Ling, A., Devlin, A., Varma, Y., Fischbach, M., Biddinger, S., Dutton, R. and Turnbaugh, P. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature,* **505,** 559-563.

Davila, A., Blachier, F., Gotteland, M., Andriamihaja, M., Benetti, P., Sanz, Y. and Tome, D. (2013). Intestinal luminal nitrogen metabolism: role of the gut microbiota and consequences for the host. *Pharmacological Research,* **68,** 95-107.

De Filippo, C., Cavalieri, D., Di Paolo, M., Ramazzotti, M., Poullet, J., Massart, S., Collini, S., Pieraccini, G. and Lionetti, P. (2010). Impact of diet in shaping gut

microbiota revealed by a comparative study in children from Europe and rural Africa. *PNAS,* **107,** 14691-14696.

DeSantis, T., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E., Keller, K., Huber, T., Dalevi, D., Hu, P. and Andersen, G. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol,* **72,** 5069-5072.

de Vrese, M. and Schrezenmeir, J. (2008). *Probiotics, prebiotics, and synbiotics*. Springer-Verlag, Berlin, Heidelberg.

Erickson, L. (2015). A baby-led approach to complementary feeding: adherence and infant food and nutrient intakes at seven months of age, Vol. Master of Science, University of Otago, Dunedin, New Zealand, pp. 140.

Experts, F. W. U. (1985). Energy and protein requirements. In: *World Health Organization Technical Report Series*, World Health Organization, Geneva, Switzerland.

Fallani, M., Amarri, S., Uusijarvi, A., Adam, R., Khanna, S., Aguilera, M., Gil, A., Vieites, J., Norin, E., Young, D., Scott, J., Dore, J. and Edwards, C. (2011). Determinants of the human infant intestinal microbiota after the introduction of first complementary foods in infant samples from five European centres. *Microbiology,* **157,** 1385-1392.

Fewtrell, M., Bronskz, J., Campoy, C., Domellöf, M., Embleton, N., Mis, N., Hojsak, I., Hulst, J., Indrio, F., Lapillone, A. and Molgaard, C. (2017). Complementary Feeding: A Position Paper by the European Society for Paediatric Gastroenterology, Hepatology, and Nutrition (ESPGHAN) Committee on Nutrition. *JPGN,* **64,** 119-132.

Friel, J., Hanning, R., Isaak, C., Prowse, D. and Miller, A. (2010). Canadian infants' nutrient intakes form complementary foods during the first year of life. *BMC Pediatrics,* **10**.

Garlick, P. (2006). Protein Requirements of Infants and Children. *Nestle Nutr Inst Workshop Ser,* **58,** 39-50.

Group, U. S. C. (2020). Negative Binomial Regression | STATA data analysis examples

Haschke, F., Grathwohl, D., Detzel, P., Steenhout, P., Wagemans, N. and Erdmann, P. (2016). Postnatal high protein intake can contribute to accelerated weight gain of infants and increased obesity risk. *Nestle Nutr Inst Workshop Ser,* **85,** 101-109.

Heiman, M. and Greenway, F. (2016). A healthy gastrointestinal microbiome is dependent on dietary diversity. *Molecular Metabolism,* **5,** 317-320.

Hellmuth, C., Uhl, O., Kirchberg, F., Grote, V., Weber, M., Rzehak, P., Carlier, C., Ferre, N., Verduci, E., Gruszfeld, D., Socha, P. and Koletzko, B. (2016). Effects of early nutrition on the infant metabolome. *Nestle Nutr Inst Workshop Ser,* **85,** 89-100.

Holscher, H. (2017). Dietary fiber and prebiotics and the gastrointestinal microbiota. *Gut Microbes,* **8,** 172 - 184.

Jandhyala, S., Talukdar, R., Subramanyam, C., Vuyyuru, H., Sasikala, M. and Reddy, D. (2015). Role of the normal gut microbiota. *Worl J Gastroenterol,* **21,** 8787-8803.

Johnson, A., Vangay, P., Al-Ghalith, G., Hillmann, B., Ward, T., Shields-Cutler, R., Kim, A., Shmagel, A., Syed, A., Students, P. M. C., Walter, J., Menon, R., Koecher, K. and Knights, D. (2019). Daily Sampling Reveals Personalized Diet-Microbiome Associations in Humans. *Cell Host Microbe,* **25**.

Johnson, C. L. and Versalovic, J. (2012). The human microbiome and its potential importance to pediatrics. *Pediatrics,* **129,** 950-960.

Kalliomaeki, M., Collado, M., Salmien, S. and Isolauri, E. (2008). Early differences in fecal microbiota composition in children may predict overweight. *Am J Clin Nutr,* **87,** 534-538.

Koenig, J., Spor, A., Scalfone, N., Fricker, A., Stombaugh, J., Knight, R., Angenent, L. and Ley, R. (2011). Succession of microbial consortia in the developing infant gut microbiome. *PNAS,* **108,** 4578-4585.

Koletzko, B., Demmelmair, H., Grote, V., Prell, C. and Weber, M. (2016). High protein intake in young children and increased weight gain and obesity risk. *Am J Clin Nutr,* **103,** 303-304.

Koletzko, B. and Hermoso, M. (2008). Reference Nutrient Intakes for Infants, Children and Adolescents. In: *Pediatric Nutrition in Practice*, Karger, Basel, pp. 285-292.

Kumar, S. and Kelly, A. (2017). Review of Childhood Obesity: From Epidemiology, Etiology, and Comorbidities to Clinical Assessment and Treatment. *Mayo Clin Proc.,* **92,** 251-265.

Kuo, A., Inkelas, M., Slusser, W., Maidenberg, M. and Halfon, N. (2011). Introduction of Solid Food to Young Infants. *Matern Child Health J,* **15,** 1185-1194.

Lahti, L. and Shetty, S. (2012-2019). microbiome R package.

Laitinen, K. and Mokkala, K. (2019). Overall Dietary Quality Relates to Gut Microbiota Diversity and Abundance. *Int J Mol Sci,* **20**.

Lanigan, J., Wells, J., Lawson, M. and Lucas, A. (2001). Validation of food diary method for assessment of dietary energy and macronutrient intake in infants and children aged 6-24 months. *Eur J Clin Nutr,* **55,** 124-129.

Lee, J. and O'Sullivan, D. (2010). Genomic Insights into Bifidobacteria. *Microbiol Mol Biol Rev,* **74,** 378 - 416.

Ley, R., Turnbaugh, P., Klein, S. and Gordon, J. (2006). Human gut microbes associated with obesity. *Nature,* **444,** 1022-1023.

Lopes, W., Marques, F., de Oliveira, C., Rodrigues, J., Silveira, M., Caldeira, A. and de Pinho, L. (2018). Infant feeding in the first two years of life. *Rev Paul Pediatr.,* **36,** 164-170.

Marco, M., Heeney, D., Binda, S., Cifelli, C., Cotter, P., Foligne, B., Gaenzle, M., Kort, R., Pasin, G., OPihlanto, A., Smid, E. and Hutkins, R. (2017). Health benefits of fermented foods: microbiota and beyond. *Current Opinion in Biotechnology,* **44,** 94-102.

Marks, K. (2015). Infant and Toddler Nutrition. *AFP,* **44,** 886-889.

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

Matamoros, S., Gras-Leguen, C., Le Vacon, F., Potel, G. and de La Cochetiere, M. (2013). Development of intestinal microbiota in infants and its impact on health *Trends in Microbiology,* **21,** 167-173.

McDonald, D., Price, M., Goodrich, J., Nawrocki, E., DeSantis, T., Probst, A., Andersen, G., Knight, R. and Hugenholtz, P. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J,* **6,** 610-618.

Morelli, L. and Capurso, L. (2012). FAO/WHO Guidelines on Probiotics - 10 Years Later. *J Clin Gastroenterol,* **46,** S1 - S2.

Morgan, X. and Huttenhower, C. (2012). Chapter 12: Human Microbiome Analysis. *PLoS Computational Biology,* **8**.

Mueller, N., Bakacs, E., Combellick, J., Grigoryan, Z. and Dominguez-Bello, M. (2015). The infant microbiome development: mom matters. *Trends Mol Med,* **21,** 109-117.

O'Callaghan, A. and Van Sinderen, D. (2016). Bifidobacteria and Their Role as Members of the Human Gut Microbiota. *Front Microbiol,* **7**.

Öhlund, I., Hernell, O., Hörnell, A., Stenlund, H. and Lind, T. (2010). BMI at 4 years of age is associated with previous and current protein intake and with paternal BMI. *European Journal of Clinical Nutrition,* **64,** 138-145.

Organization, W. H. (2002). *Complementary feeding -- Report of the global consultation -- summary of guiding principles*. WHO, Geneva, Switzerland.

Organization, W. H. (2018). Infant and young child feeding, Vol. 2019.

Ortega, R., Perez-Rodrigo, C. and Lopez-Sobaler, A. (2015). Dietary assessment methods: dietary records. *Nutr Hosp,* **31,** 38-45.

Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A. and Brown, P. O. (2007). Development of the human infant intestinal microbiota. *PLoS Biol,* **5,** e177.

Pannaraj, P., Li, F., Cerini, C., Bender, J., Yang, S., Rollie, A., Adisetiyo, H., Zabih, S., Lincez, P., Bittinger, K., Bailey, A., Bushman, F., Sleasman, J. and Aldrovandi, G. (2017). Association Between Breast Milk Bacterial Communities and Establishment and Development of the Infant Gut Microbiome. *JAMA Pediatr.,* **171,** 647 - 654.

Penders, J., Thijs, C., Vink, C., Stelma, F., Snijders, B., Kummeling, I., van den Brandt, P. and Stobberingh, E. (2006). Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics,* **118,** 511-521.

Perez-Munoz, M., Arrieta, M.-C., Ramer-Tait, A. and Walter, J. (2017). A critical assessment of the "sterile womb" and "in utero colonization" hypotheses: implications for research on the pioneer infant microbiome. *Microbiome,* **5**.

Quin, C., Estaki, M., Vollman, D., J., B., Gill, S. and Gibson, D. (2018). Probiotic supplementation and associated infant gut microbiome and health: a cautionary retrospective clinical comparison. *Scientific Reports,* **8**.

Schwiertz, A., Taras, D., Schaefer, K., Beijer, S., Bos, N., Donus, C. and Hardt, P. (2010). Microbiota and SCFA in lean and overweight healthy subjects. *Obesity,* **18,** 190-195.

Sonnenburg, E., Smits, S., Tikhonov, M., Higginbottom, S., Wingreen, N. and Sonnenburg, J. (2016). Diet-induced extinction in the gut microbiota compounds over generations. *Nature,* **529,** 212-215.

Sprague, R. (2018). How to think about microbiome diversity.

Stearns, J., Davidson, C., McKeon, S., Whelan, F., Fontes, M. and Schryvers, A. (2015). Culture and molecular-based profiles show shifts in bacterial communities of the upper respiratory tract that occur with age. *ISMEJ,* **9,** 1256-1259.

Stearns, J., Simioni, J., Gunn, E., McDonald, H., Holloway, A., Thabane, L., Mousseau, A., Schertzer, J., Ratcliffe, E., Rossi, L., Surette, M., Morrison, K. and Hutton, E. (2017). Intrapartum antibiotics for GBS prophylaxis alter colonization patterns in the early infant gut microbiome of low risk infants. *Scientific Reports,* **7**.

Stephen, A., Alles, M., de Graaf, C., Fleith, M., Hadjilucas, E., Isaacs, E., Maffeis, C., Zeinstra, G., Matthys, C. and Gil, A. (2012). The role and requirements of digestible dietary carbohydrates in infants and toddlers. *Eur J Clin Nutr,* **66,** 765-779.

Tandoi, F., Morlacchi, L., Bossi, A. and Agosti, M. (2017). Introducing complementary foods in the first year of life. *Medical and Surgical Pediatrics,* **39,** 128-130.

Tapiainen, T., Koivusaari, P., Nrinkac, L., Lorenzi, H., Salo, J., Renko, M., Pruikkonen, H., Pokka, T., Li, W., Nelson, K., Pirttila, A. and Tejesvi, M. (2019). Impact of intrapartum and postnatal antibiotics on the gut microbiome and emergence of antimicrobial resistance in infants. *Scientific Reports,* **9**.

Turnbaugh, P. and Gordon, J. (2009). The core gut microbiome, energy balance and obesity. *J Physiol.,* **587,** 4153-4158.

Turnbaugh, P., Hamady, M., Yatsunenko, T., Cantarel, B., Duncan, A., Ley, R., Sogin, M., Jones, W., Roe, B., Affourtit, J., Egholm, M., Henrissat, B., Heath, A., Knight, R. and Gordon, J. (2009). A core gut microbiome in obese and lean twins. *Nature,* **457,** 480-484.

Turnbaugh, P., Ridaura, V., Faith, J., Rey, F., Knight, R. and Gordon, J. (2010). The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Scie Transl Med,* **1**.

Turroni, F., Peano, C., Pass, D., Foroni, E., Severegnini, M., Claesson, M., Kerr, C., Hourihane, J., Murray, D., Fuligni, F., Gueimonde, M., Margolles, A., De Bellis, G., O'Toole, P., van Sinderen, D., Marchesi, J. and Ventura, M. (2012). Diversity of bifidobacteria within the infant gut microbiota. *PLos ONE,* **7**.

Uauy, R. and Dangour, A. (2009). Fat and Fatty Acid Requirements and Recommendations for Infants of 0-2 Years and Children of 2-18 Years. *Ann Nutr Metab,* **55,** 76-96.

Unicef. (2019). Infant and young child feeding.

USDA. (1992). The Food Guide Pyramid, Washington, DC.

USDA. (2018). National Nutrient Database for Standard Reference Legacy Release.

Vael, C., Verhulst, S., Nelen, V., Goossens, H. and Desager, K. (2011). Intestinal microflora and body mass index during the first three years of life: an observational study. *Gut Pathog,* **3**.

Valles, Y., Artacho, A., Pascual-Garcia, A., Ferrus, M., Gosalbes, M., Abellan, J. and Francino, M. (2014). Microbial Succession in the Gut: Directional Trends of Taxonomic and Functional Change in a Birth Cohort of Spanish Infants. *PLoS Genet,* **10**.

van Leeuwen, M., Louwerse, M., Opmeer, B., Limpens, J., Serlie, M., Reitsma, J. and Mol, B. (2012). Glucose challenge test for detecting gestational diabetes mellitus: a systematic review. *BJOG An International Journal of Obstetrics and Gynaecology,* **119,** 393-401.

Voortman, T., Braun, K., Kiefte-de Jong, J., Jaddoe, V., Franco, O. and van den Hooven, E. (2016). Protein intake in early childhood and body composition at the age of 6 years: The Generation R Study. *International Journal of Obesity,* **40,** 1018-1025.

Wang, L., van Grieken, A., van der Velde, L., Vlasbom, E., Beltman, M., L'Hoir, M., Boere-Boonekamp, M. and Raat, H. (2019). Factors associated with early introduction of complementary feeding and consumption of non-recommended foods among Dutch infants: the BeeBOFT study. *BMC Public Health,* **19**.

Weber, M., Grote, V., Closa-Monasterolo, R., Escribano, J., Langhendries, J.-P., Dain, E., Giovannini, M., Verduci, E., Gruszfeld, D., Socha, P. and Koletzko, B. (2014). Lower protein content in infant formula reduces BMI and obesity risk at school age: follow-up of a randomized trial. *Am J Clin Nutr,* **99,** 1041-1051.

Weir, C. and Jan, A. (2019). StatPearls. In: *BMI Classification Percentile and Cut Off Points*.

White, J., Begin, F., Kumapley, R., Murray, C. and Krasevec, J. (2017). Complementary feeding practices: Current global and regional estimates. *Matern Child Nutr,* **13**.

Williams, S. (2014). Gnotobiotics. *Proc Natl Acad Sci USA,* **11**.

Willis, A. (2019). Rarefaction, Alpha Diversity, and Statistics. *Front Microbiol,* **10**.

Yang, B., Hur, K. and Lee, M. (2017). Alterations in Gut Microbiota and Immunity by Dietary Fat. *YMJ,* **58,** 1083-1091.

Yatsunenko, T., Rey, F., Manary, M., Trehan, I., Dominguez-Bello, M., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R., Anokhin, A., Heath, A., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J., Lozupone, C., Lauber, C., Clemente, J., Knights, D., Knight, R. and Gordon, J. (2012). Human gut microbiome viewed across age and geography. *Nature,* **486,** 222-228.

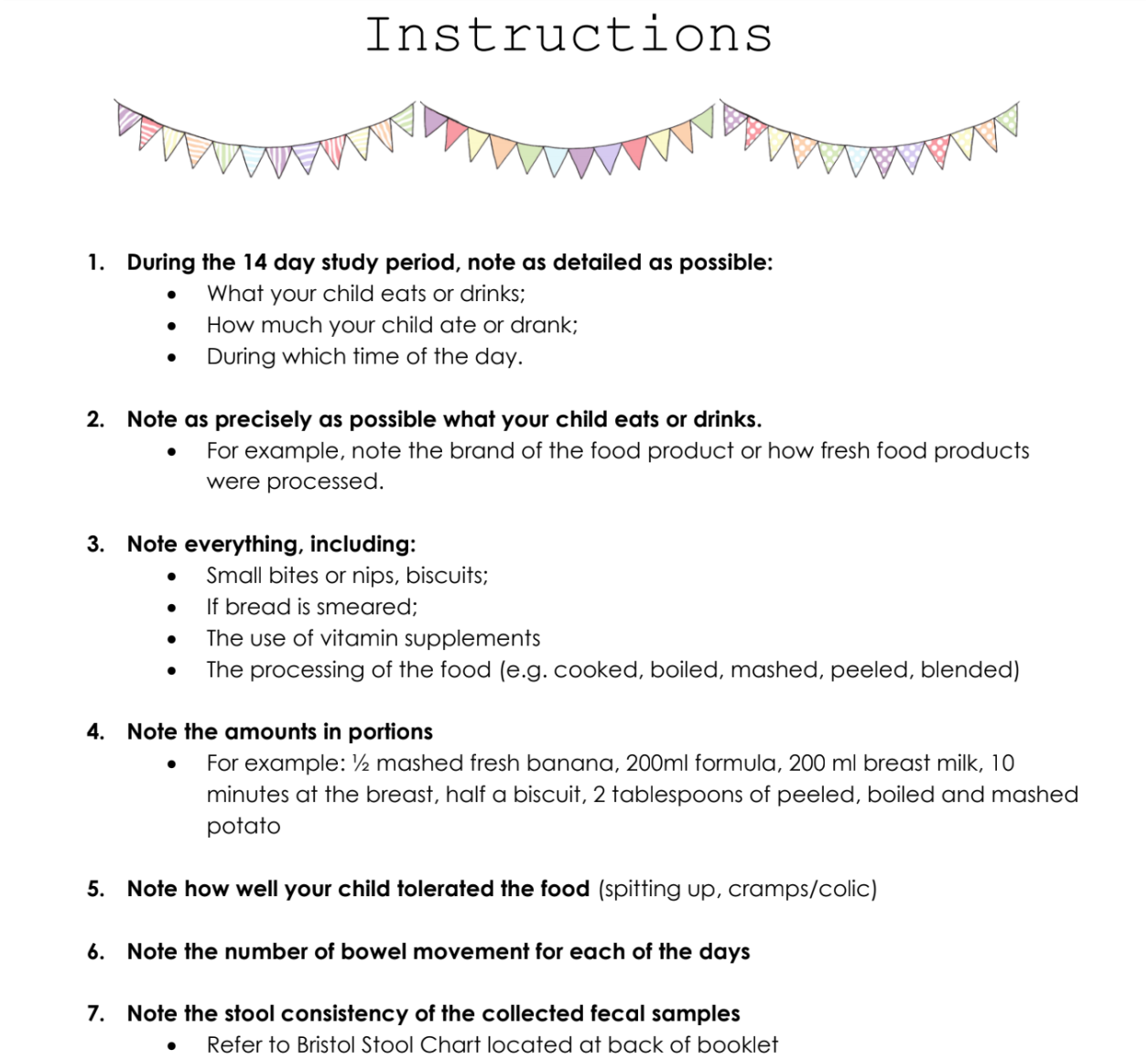
Young, V. (2012). The Intestinal Microbiota in Health and Disease. *Curr Opin Gastroenterol,* **28,** 63-69.

Zimmer, J., Lange, N., Frick, J., Sauer, H., Zimmermann, K., Schwiertz, A., Rusch, K., Klosterhalfen, S. and Enck, P. (2012). A vegan or vegetarian diet substantially alters the human colonic faecal microbiota. *Eur J Clin Nutr,* **66,** 53-60.

Zivkovic, A., German, J., Lebrilla, C. and Mills, D. (2011). Human milk glycobiome and its impact on the infant gastrointestinal microbiota. *PNAS,* **108,** 4653-4658.

Appendix A: **Nutrition Analysis**

Fig. 39 shows the instructions for completion of the food diary given to caregivers for the sub-study period.



Instructions

1. **During the 14 day study period, note as detailed as possible:**
   - What your child eats or drinks;
   - How much your child ate or drank;
   - During which time of the day.

2. **Note as precisely as possible what your child eats or drinks.**
   - For example, note the brand of the food product or how fresh food products were processed.

3. **Note everything, including:**
   - Small bites or nips, biscuits;
   - If bread is smeared;
   - The use of vitamin supplements
   - The processing of the food (e.g. cooked, boiled, mashed, peeled, blended)

4. **Note the amounts in portions**
   - For example: ½ mashed fresh banana, 200ml formula, 200 ml breast milk, 10 minutes at the breast, half a biscuit, 2 tablespoons of peeled, boiled and mashed potato

5. **Note how well your child tolerated the food** (spitting up, cramps/colic)

6. **Note the number of bowel movement for each of the days**

7. **Note the stool consistency of the collected fecal samples**
   - Refer to Bristol Stool Chart located at back of booklet

Figure 39: Instructions given to mothers of the participants for completion of the food diary for the intensively sampled cohort

Figure 40 shows the completion rates of the food diaries of the Baby, Food & Mi cohort.

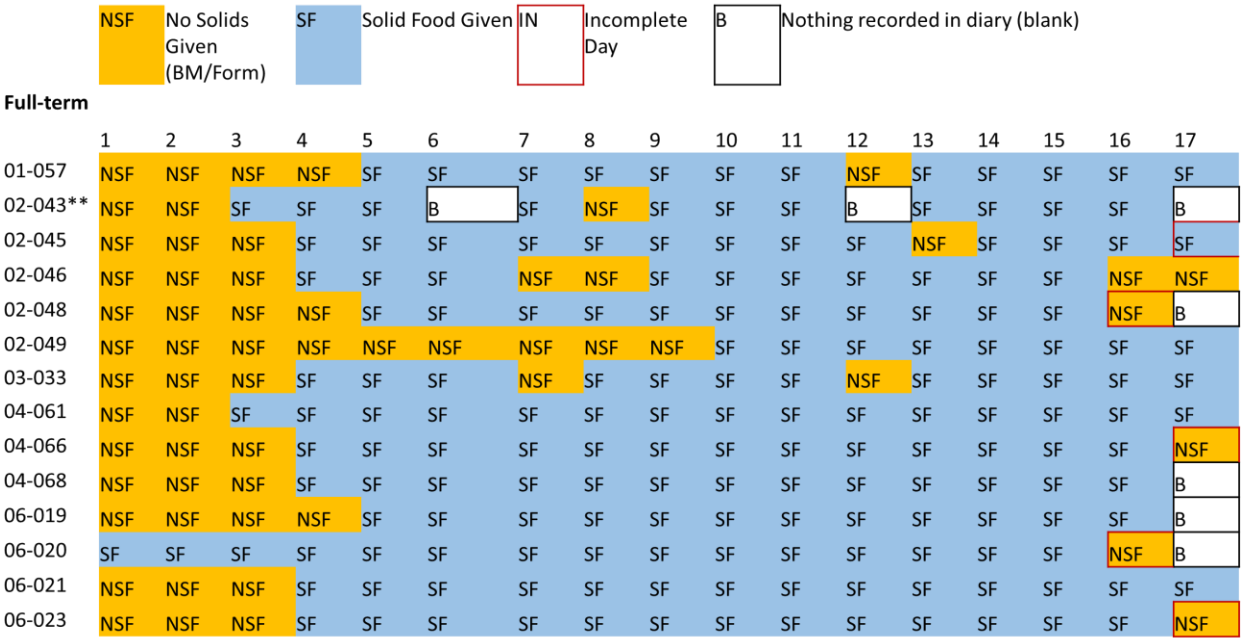


Figure 40: Overview of the completion of the food diaries

Figure 41 shows the proportion of energy from solid foods on the last day that the study diary was completed fully. The majority of the infants receive less than 10% of their energy from solid foods at the time of introduction in this study.

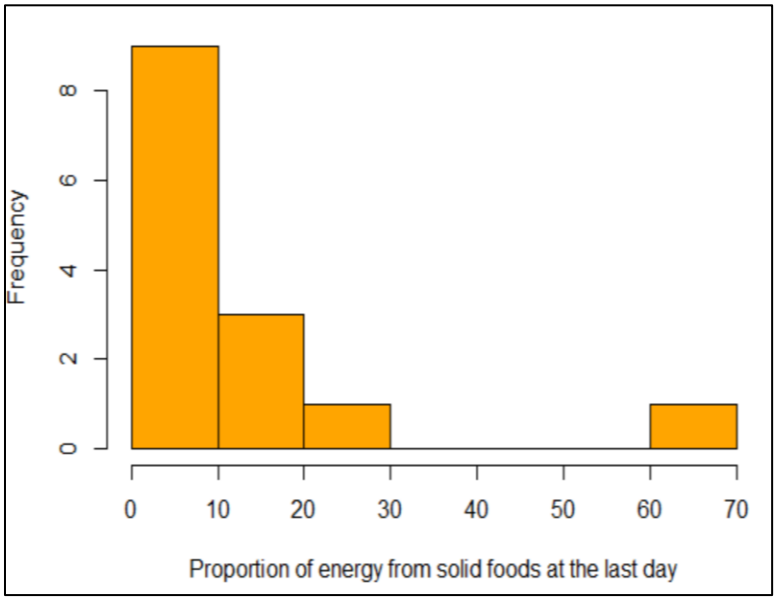


Figure 41: Proportion of calories from solid foods on the last filled out day of the study diary

Fig. 42 shows macronutrient intake by participant, in this image macronutrient calories are shown as percentages of total energy from solid foods.

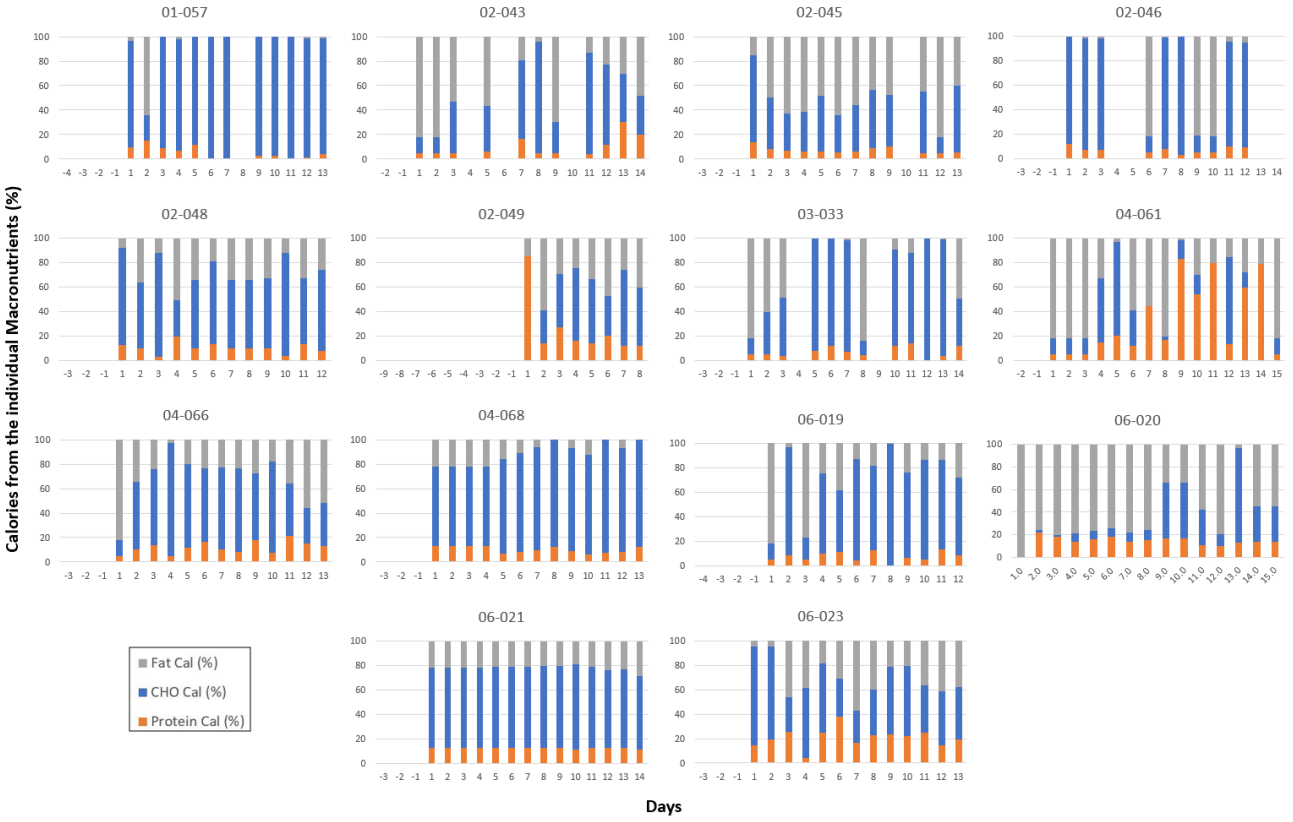


Figure 42: Calories from the macronutrients (in % of total calories).

Fig. 43 shows the individual intakes of fiber (g/d) for the sub-study period.

Figure 43: Individual intakes of fiber (g/d) by participant ID over the sub-study period.

Food Groups and Foods in the Food Diaries

Table 14: Items from the food diaries, classified into pre-chosen food groups. This is an extensive list of all the foods described in the food diaries.

Fruit	**Vegetables**	**Grains/Beans/Legumes/Nuts**	**Dairy**	**Meat**	**Confections**	**Oils**
Orange	Squash	Peanut Butter	Yogurt	Meatballs	Carrot cake	Avocado oil
Banana	Broccoli	Tortilla, white	Cheese	Salmon	Barbecue sauce[6]	Flax seeds
Pear	Sweet potato	Tortellini	Milk	Corned Beef	Doughnut	Olive oil
Avocado	Spinach	Potatoes	Cream cheese	Chicken breast	Croissant	Mayonnaise
Blueberries	Carrots	Rice		Eggs	Chocolate cookie	Butter
Raspberries	Cauliflower	Bread, white		Beef		Cod liver oil
Mango	Zucchini	Macaroni		Kangaroo steak		Margarine
Peaches	Onion	Oats		Turkey breast		
Strawberries	Asparagus	Polenta		Pork		

[6] Barbecue sauce is classified as a confection due to its high sugar content.

Apricots (dried)	Bell pepper	Bread, rye		Bone broth		
Marinara sauce (mostly tomato)	PC organic vegetable puffs	Oatmeal				
Apple sauce	Green peas	Baby rice rusks				
Natural preserved lemons	Green beans	Beans				
Persimmon	Creamed corn puree	Bread, pita				
Prune puree	Brussel sprouts	Hummus				
Watermelon		Bread, sourdough				
Acerola powder		Spaghetti				
Kiwi		Bagel				
		Arrowroot cookie				
		Rice cereal				
		Lentils				
		Almond milk				
		Coconut milk				

		Almond butter				

Fig. 44 shows the classification of the dominant food group in more detail. Percentages of the food groups from cumulative days of food data entered in the food diaries.

PID	%Dairy	%Fruit	%Grain	%Meat	%Confection	%Veg	%Oil	%Fruit+Veg
01-057	0.0	59.1	4.5	0.0	0.0	36.4	0.0	95.5
02-043	10.3	51.7	17.2	10.3	6.9	3.4	0.0	55.2
02-045	0.0	35.6	8.5	6.8	0.0	27.1	22.0	62.7
02-046	0.0	40.0	0.0	0.0	0.0	60.0	0.0	100.0
02-048	4.3	15.2	52.2	4.3	0.0	21.7	2.2	37.0
02-049	26.8	14.6	39.0	17.1	2.4	0.0	0.0	14.6
03-033	0.0	46.4	7.1	0.0	0.0	32.1	14.3	78.6
04-061	17.9	28.6	0.0	25.0	0.0	28.6	0.0	57.1
04-066	0.0	21.7	33.3	16.7	0.0	28.3	0.0	50.0
04-068	0.0	18.8	68.8	0.0	0.0	12.5	0.0	31.3
06-019	8.9	55.4	10.7	3.6	1.8	16.1	3.6	71.4
06-020	33.3	31.1	0.0	31.1	0.0	0.0	4.4	31.1
06-021	0.0	6.5	72.6	0.0	0.0	21.0	0.0	27.4
06-023	0.0	29.6	22.5	18.3	0.0	25.4	4.2	54.9

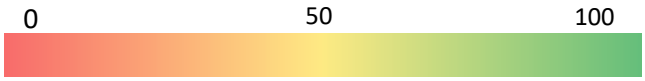


Figure 44: Percentages of the food groups from the cumulative days of food data entered into the food diaries

Appendix B: Microbiome Individual Analysis

The following figures show the alpha diversity plots for Shannon and Observed species richness, as well as changes in beta diversity between samples and PCoA plots. These were used to analyze changes on the individual basis.

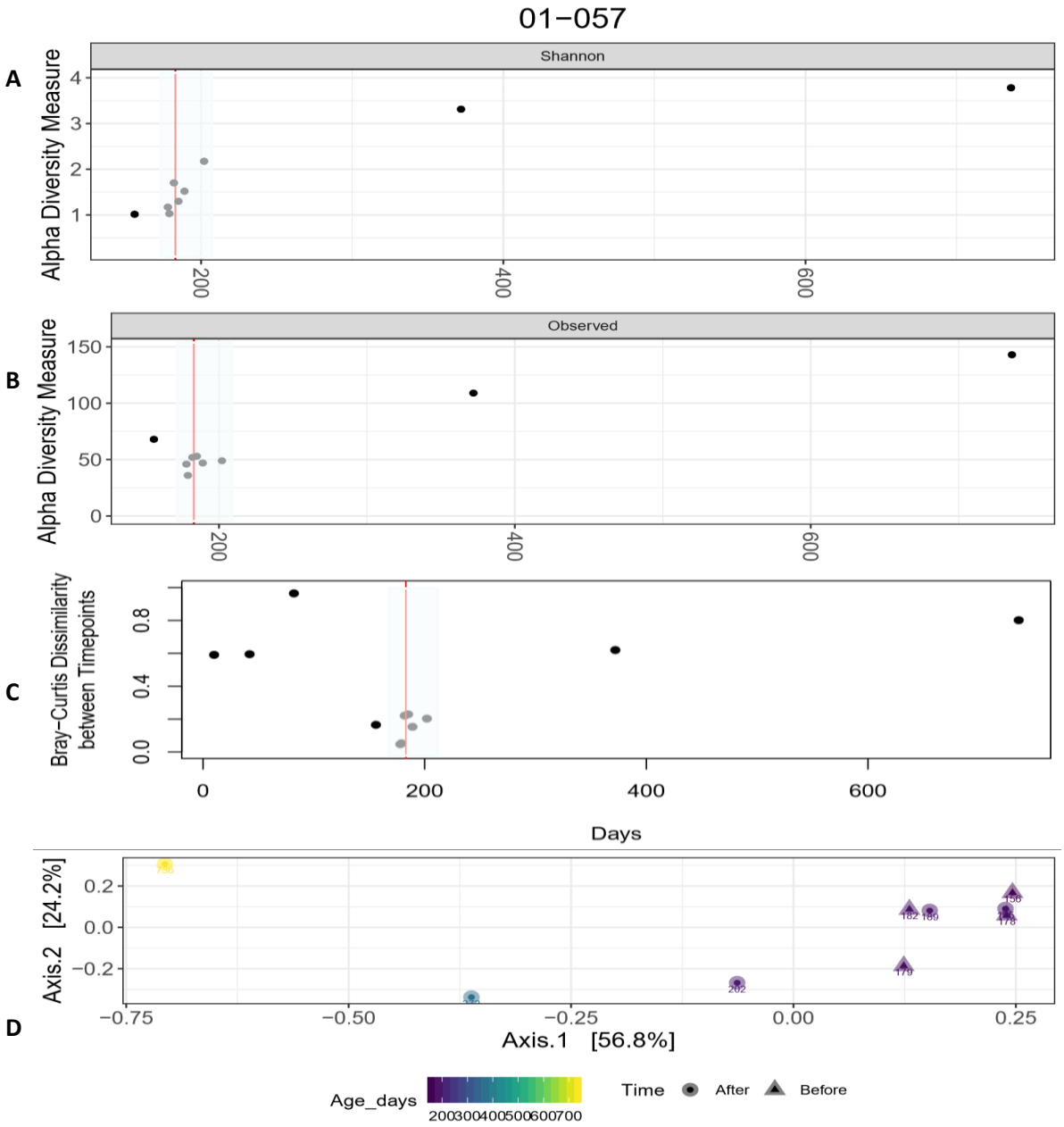


Figure 45: A: Shannon alpha diversity for samples > 90 days old, B: Observed alpha diversity for samples > 90 days, C: Change in Beta - Diversity from sample to sample, D: PCoA plots for samples > 90 days old. All for participant 01-057.

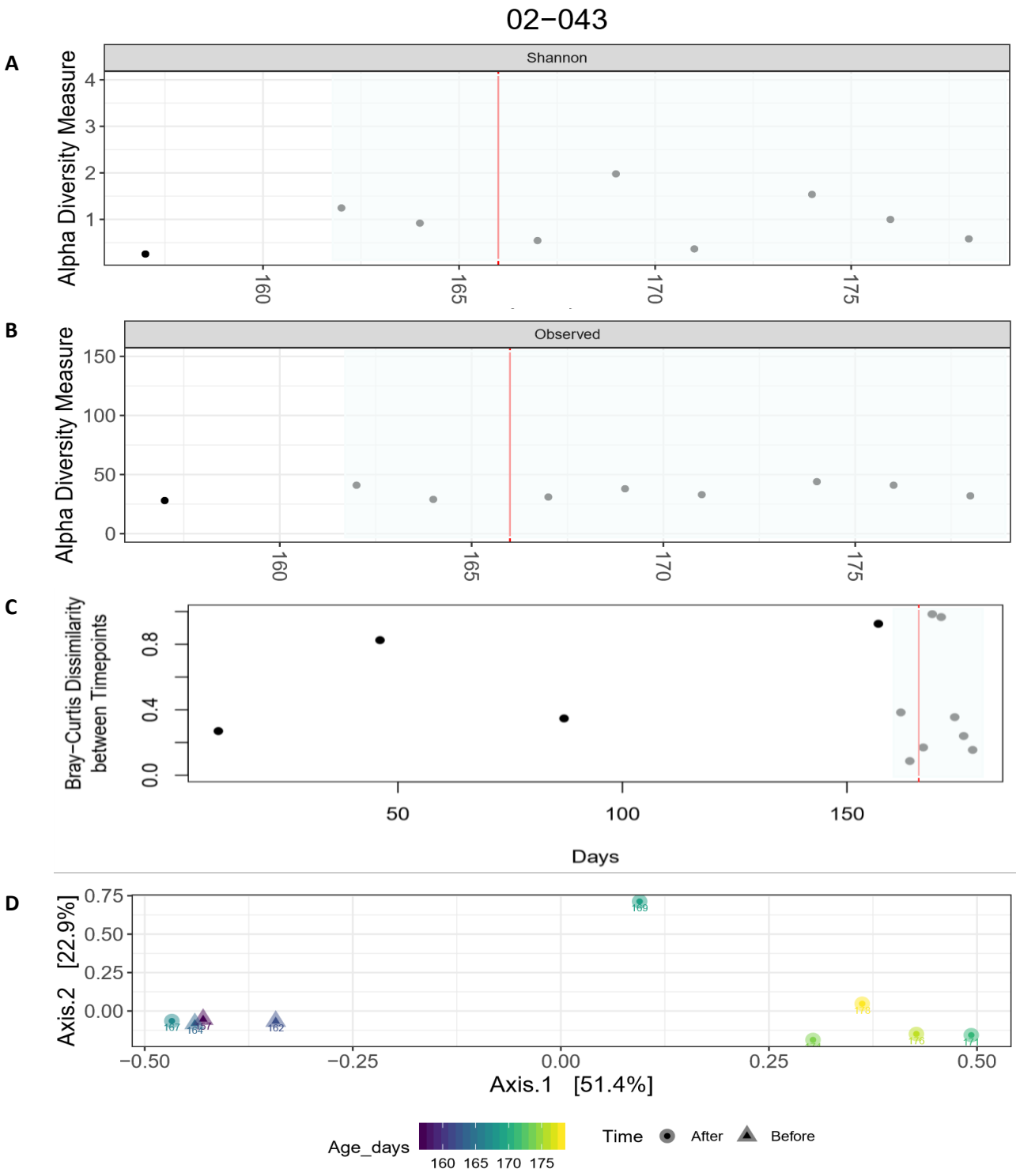


Figure 46: A: Shannon alpha diversity for samples > 90 days old, B: Observed alpha diversity for samples > 90 days, C: Change in Beta - Diversity from sample to sample, D: PCoA plots for samples > 90 days old. All for participant 02-043.

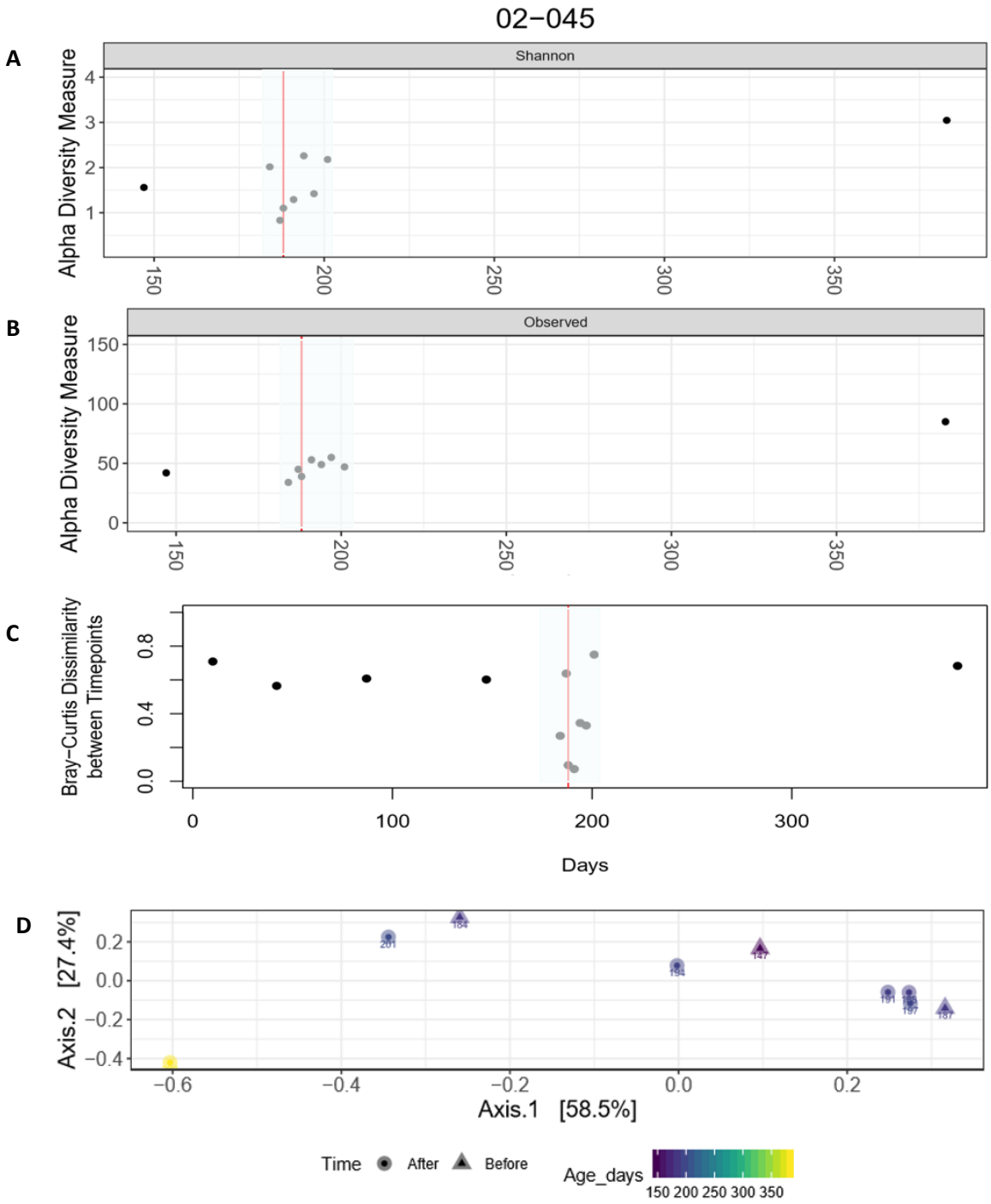


Figure 47: A: Shannon alpha diversity for samples > 90 days old, B: Observed alpha diversity for samples > 90 days, C: Change in Beta - Diversity from sample to sample, D: PCoA plots for samples > 90 days old. All for participant 02-045.

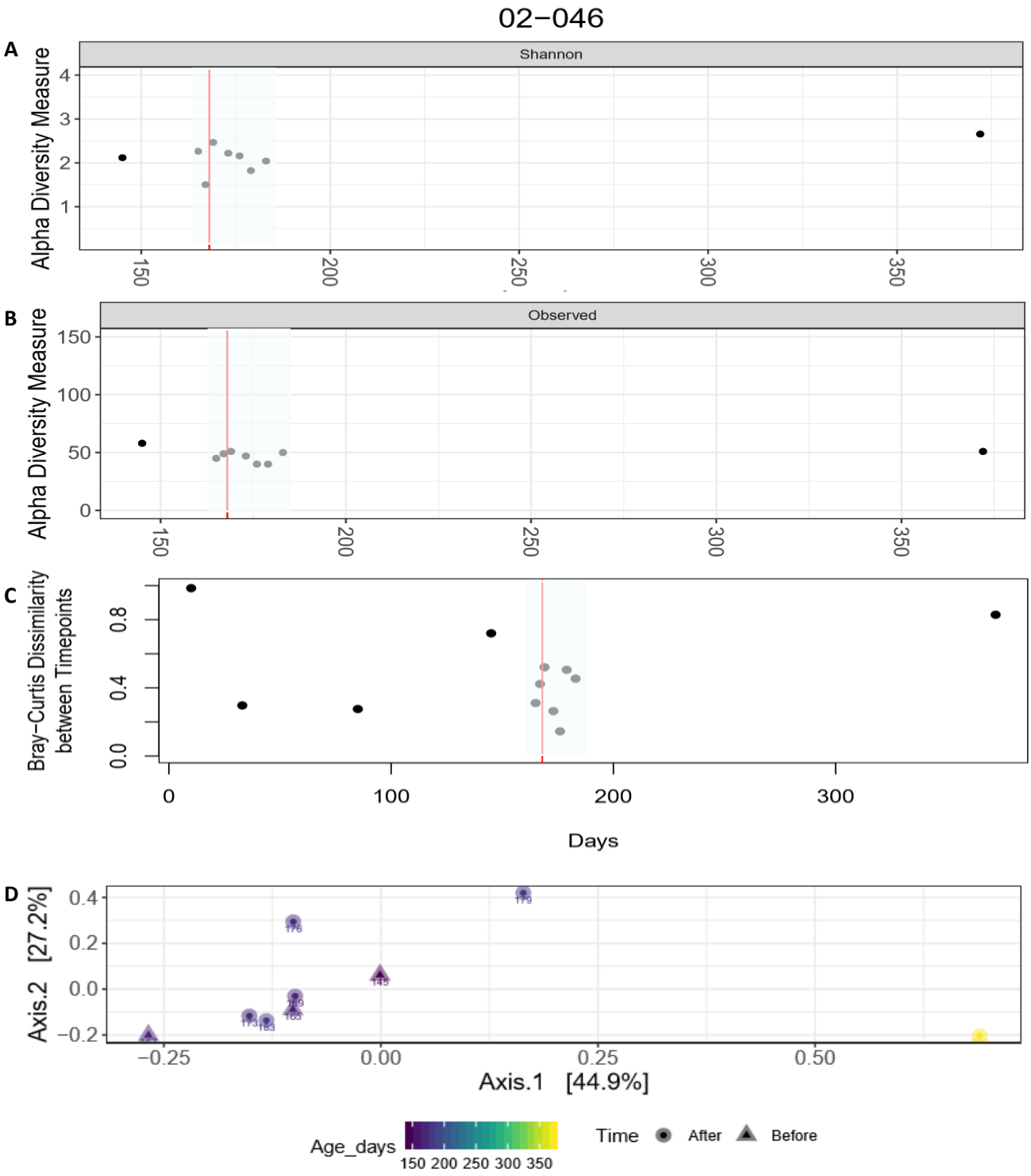


Figure 48: A: Shannon alpha diversity for samples > 90 days old, B: Observed alpha diversity for samples > 90 days, C: Change in Beta - Diversity from sample to sample, D: PCoA plots for samples > 90 days old. All for participant 02-046.

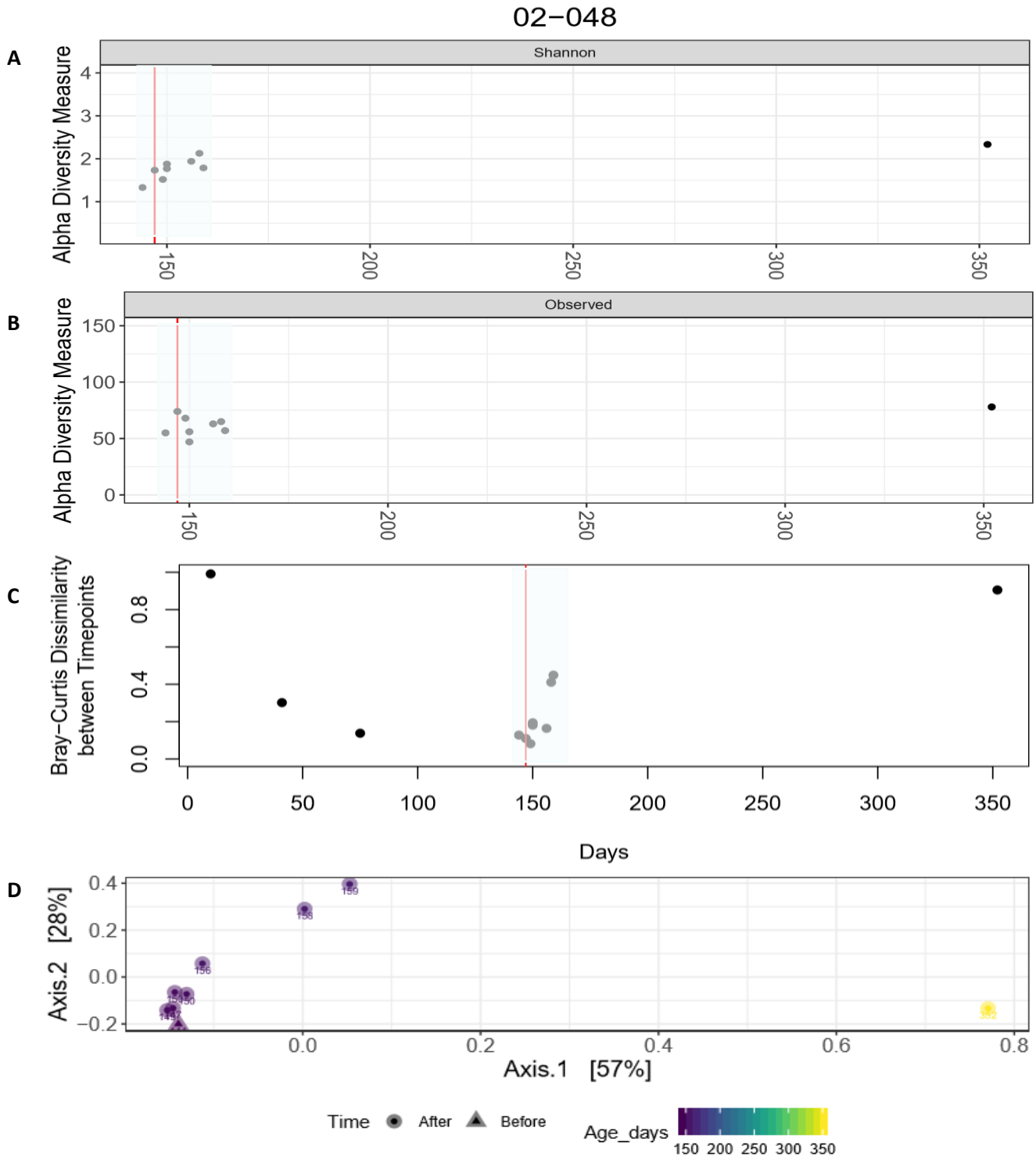


Figure 49: A: Shannon alpha diversity for samples > 90 days old, B: Observed alpha diversity for samples > 90 days, C: Change in Beta - Diversity from sample to sample, D: PCoA plots for samples > 90 days old. All for participant 02-048.

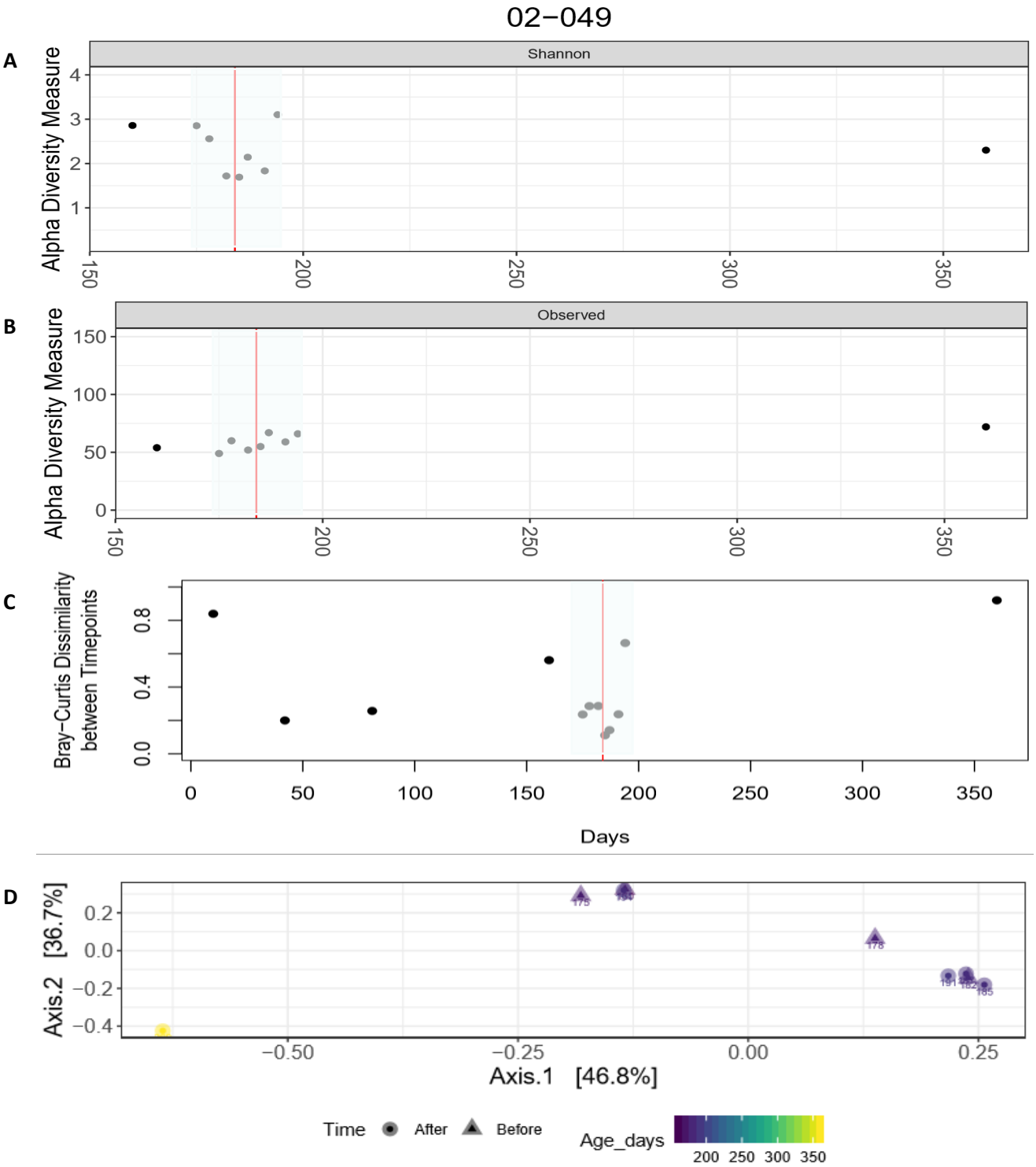


Figure 50: A: Shannon alpha diversity for samples > 90 days old, B: Observed alpha diversity for samples > 90 days, C: Change in Beta - Diversity from sample to sample, D: PCoA plots for samples > 90 days old. All for participant 02-049.

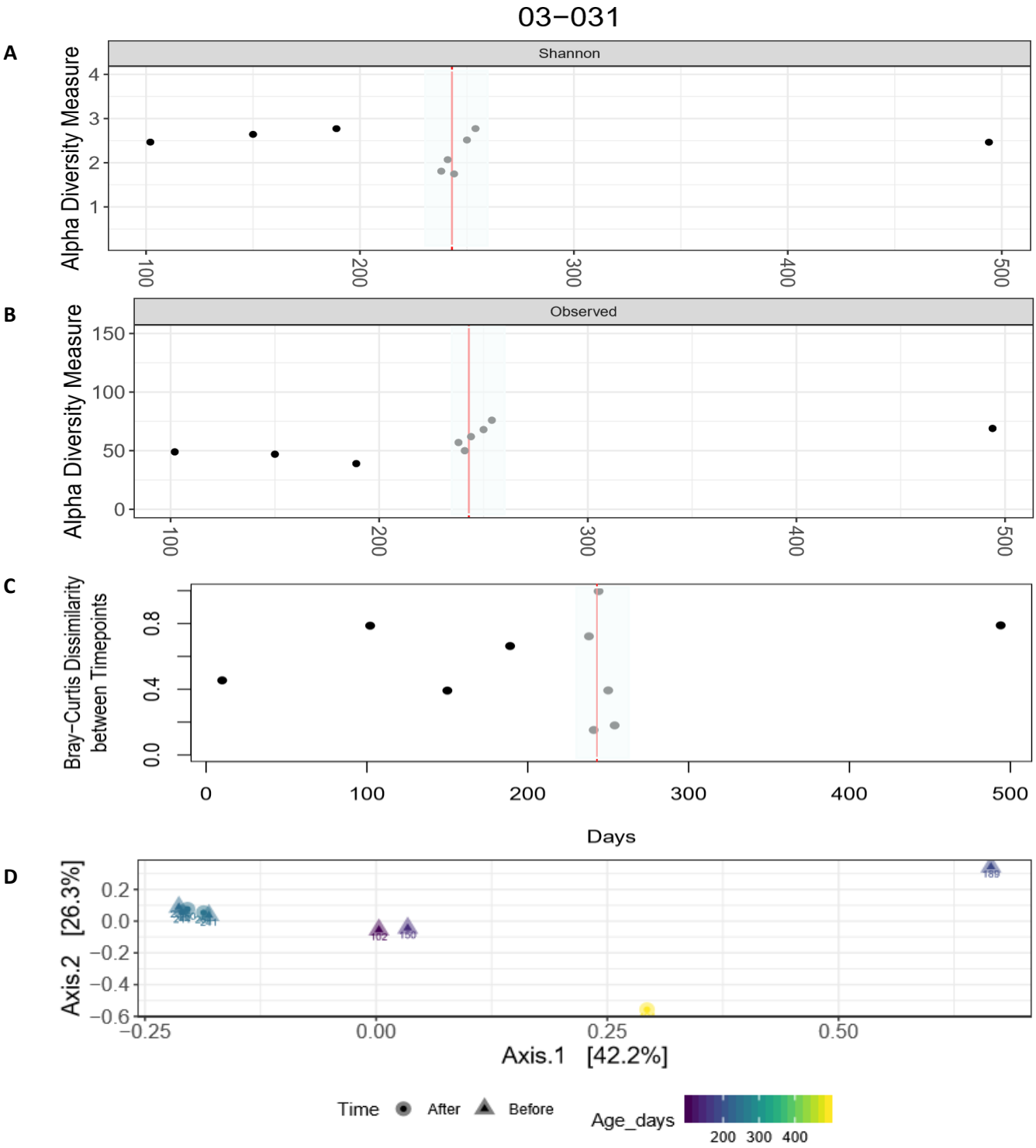


Figure 51: A: Shannon alpha diversity for samples > 90 days old, B: Observed alpha diversity for samples > 90 days, C: Change in Beta - Diversity from sample to sample, D: PCoA plots for samples > 90 days old. All for participant 03-031.

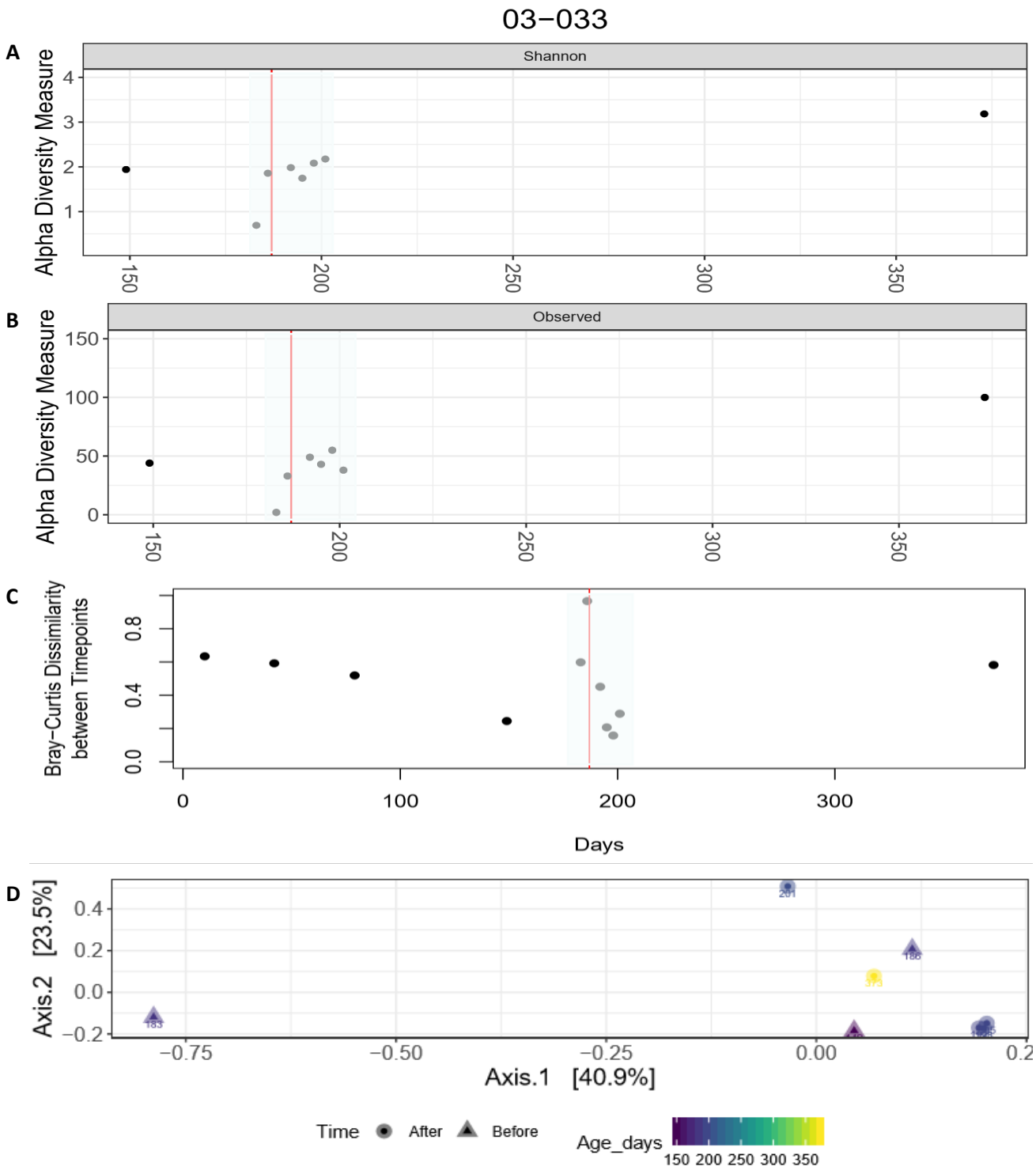


Figure 52: A: Shannon alpha diversity for samples > 90 days old, B: Observed alpha diversity for samples > 90 days, C: Change in Beta - Diversity from sample to sample, D: PCoA plots for samples > 90 days old. All for participant 03-033.

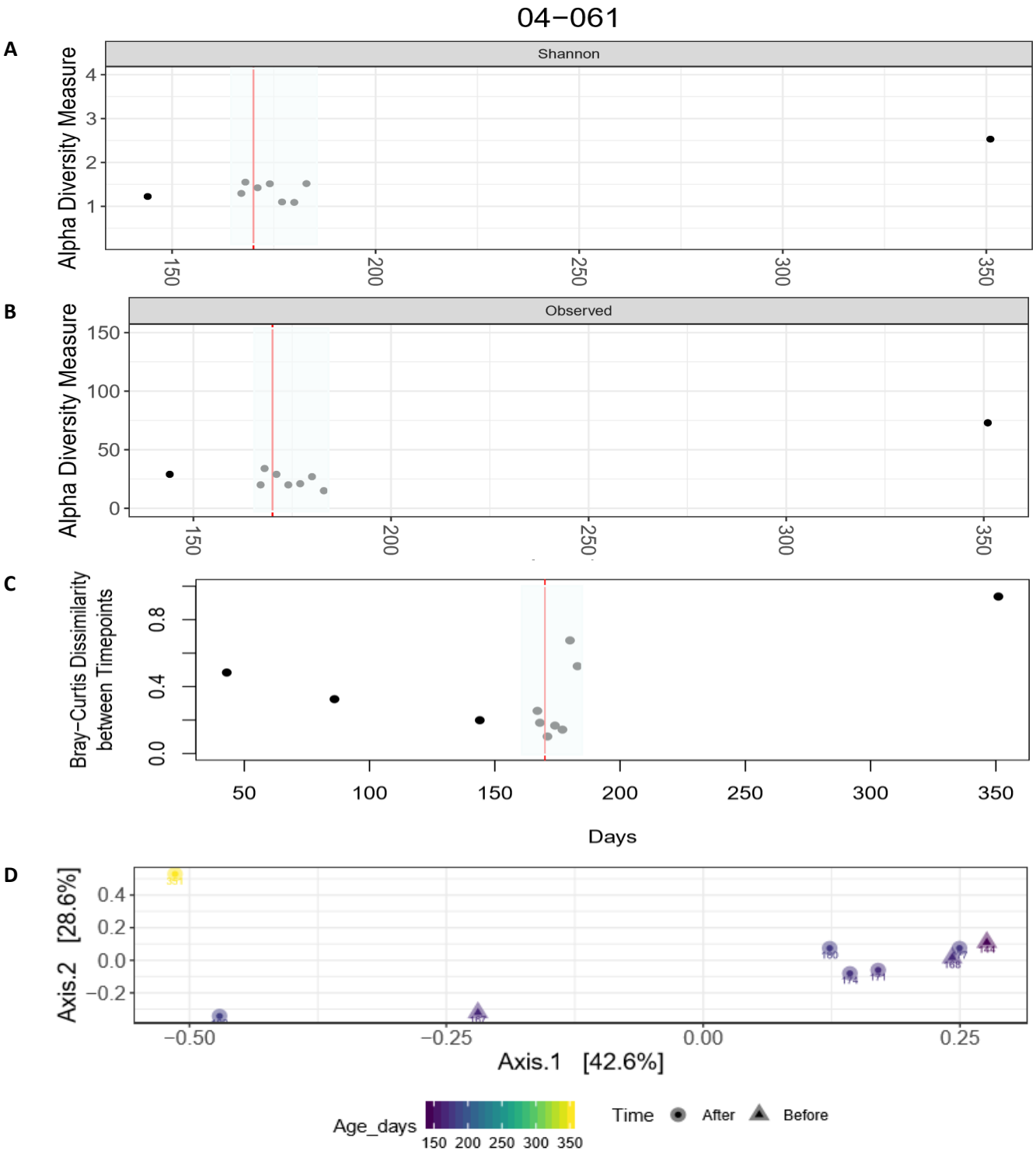


Figure 53: A: Shannon alpha diversity for samples > 90 days old, B: Observed alpha diversity for samples > 90 days, C: Change in Beta - Diversity from sample to sample, D: PCoA plots for samples > 90 days old. All for participant 04-061.

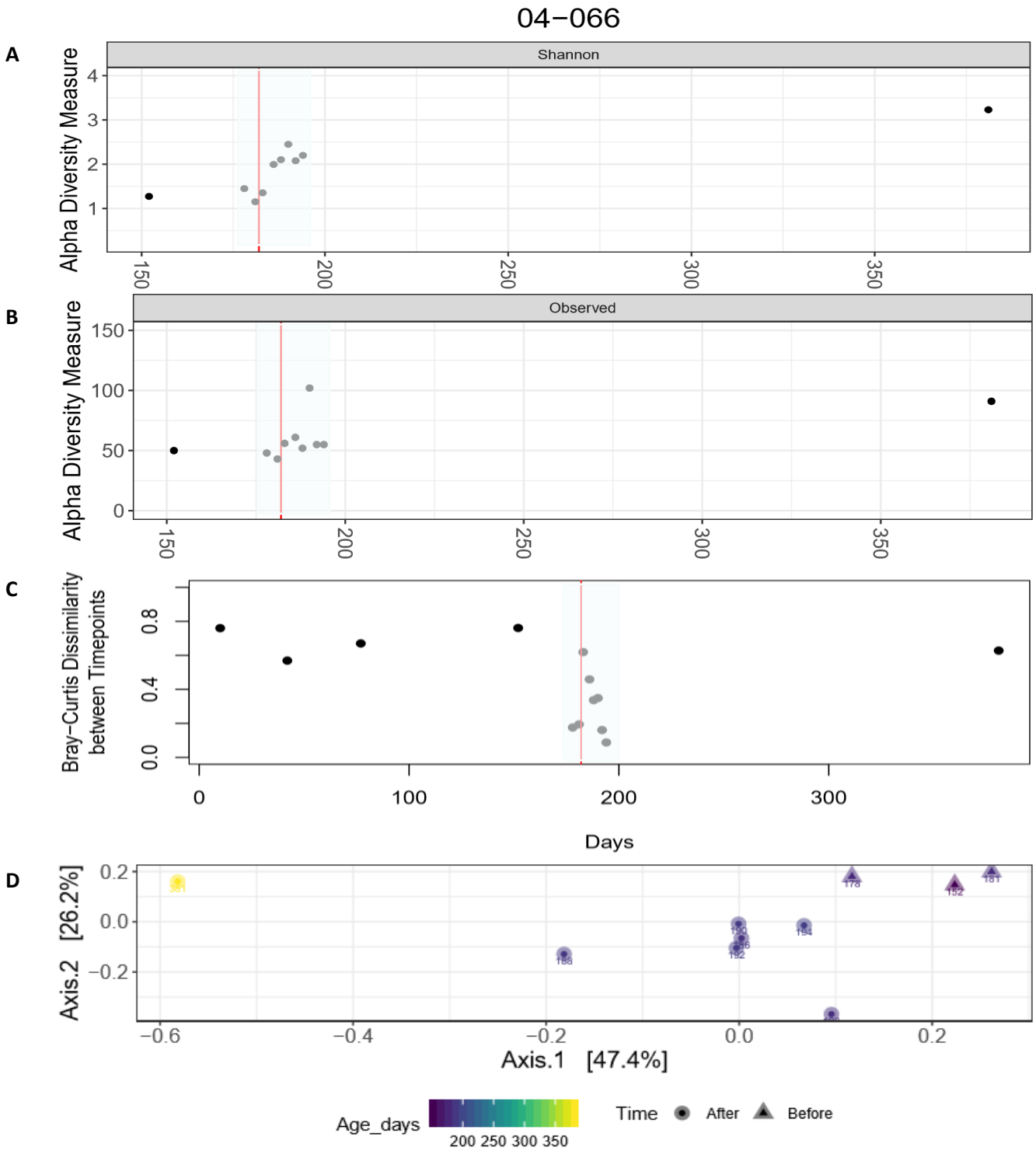


Figure 54: A: Shannon alpha diversity for samples > 90 days old, B: Observed alpha diversity for samples > 90 days, C: Change in Beta - Diversity from sample to sample, D: PCoA plots for samples > 90 days old. All for participant 04-066.

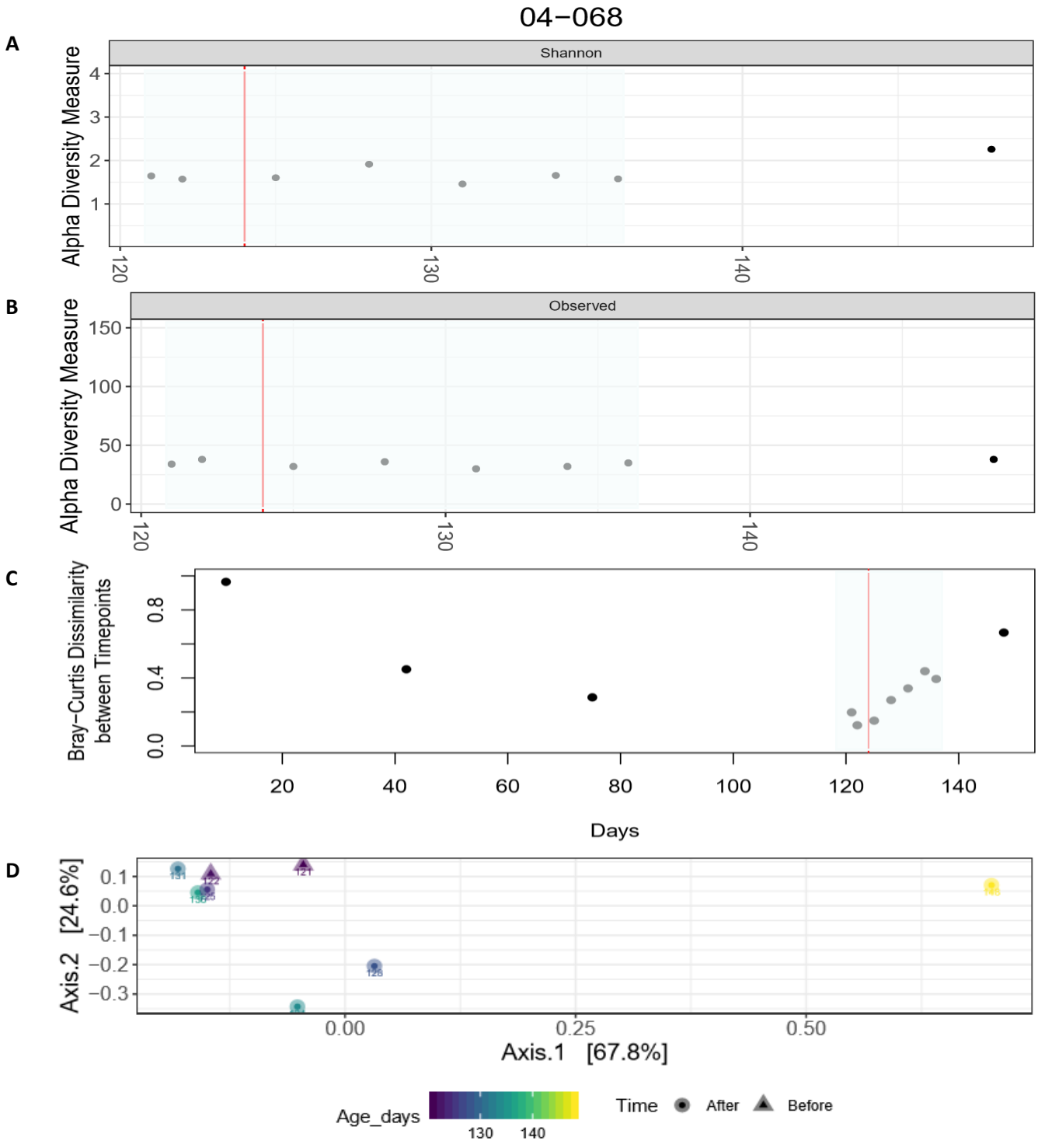


Figure 55: A: Shannon alpha diversity for samples > 90 days old, B: Observed alpha diversity for samples > 90 days, C: Change in Beta - Diversity from sample to sample, D: PCoA plots for samples > 90 days old. All for participant 04-068.

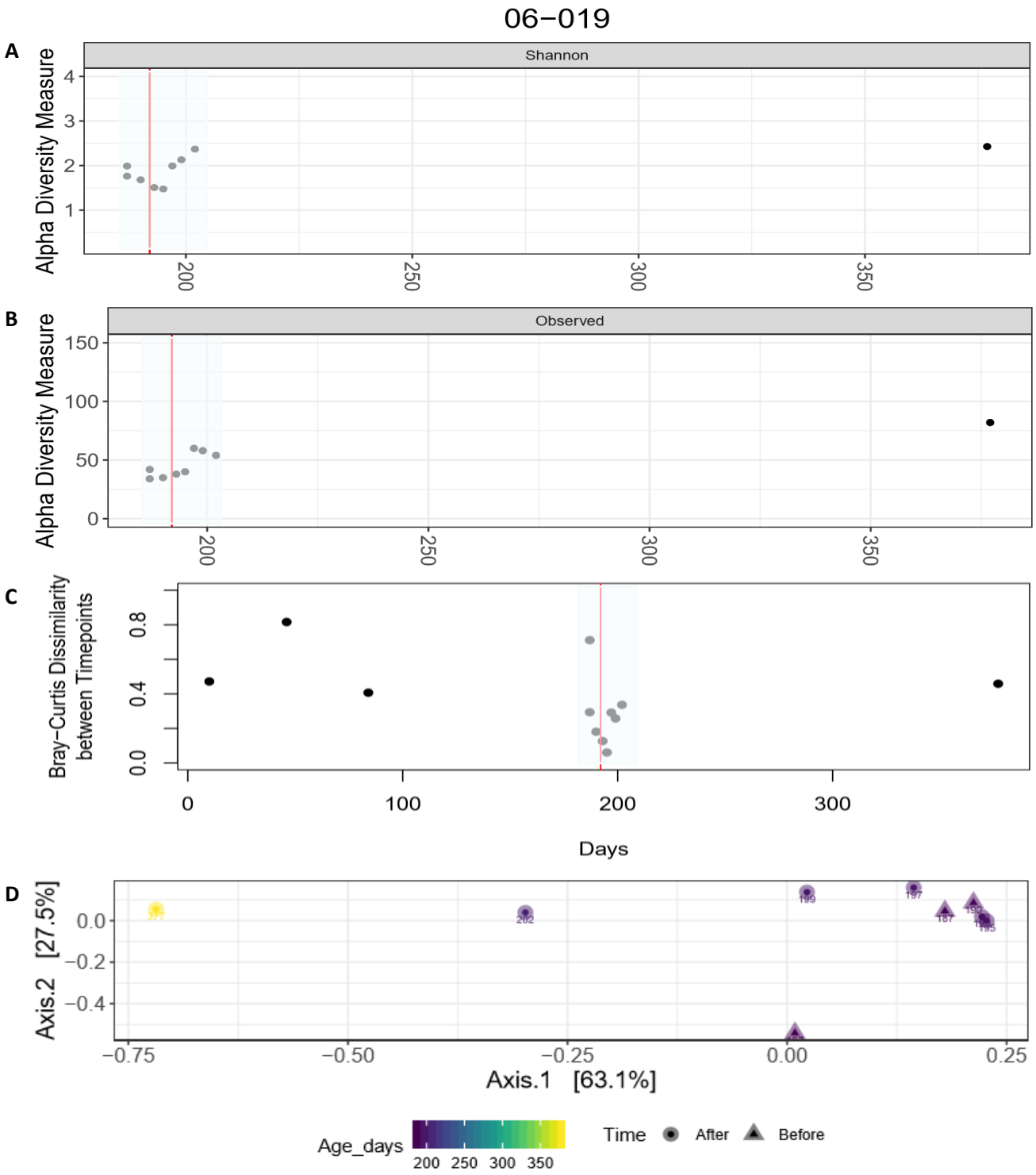


Figure 56: A: Shannon alpha diversity for samples > 90 days old, B: Observed alpha diversity for samples > 90 days, C: Change in Beta - Diversity from sample to sample, D: PCoA plots for samples > 90 days old. All for participant 06-021.

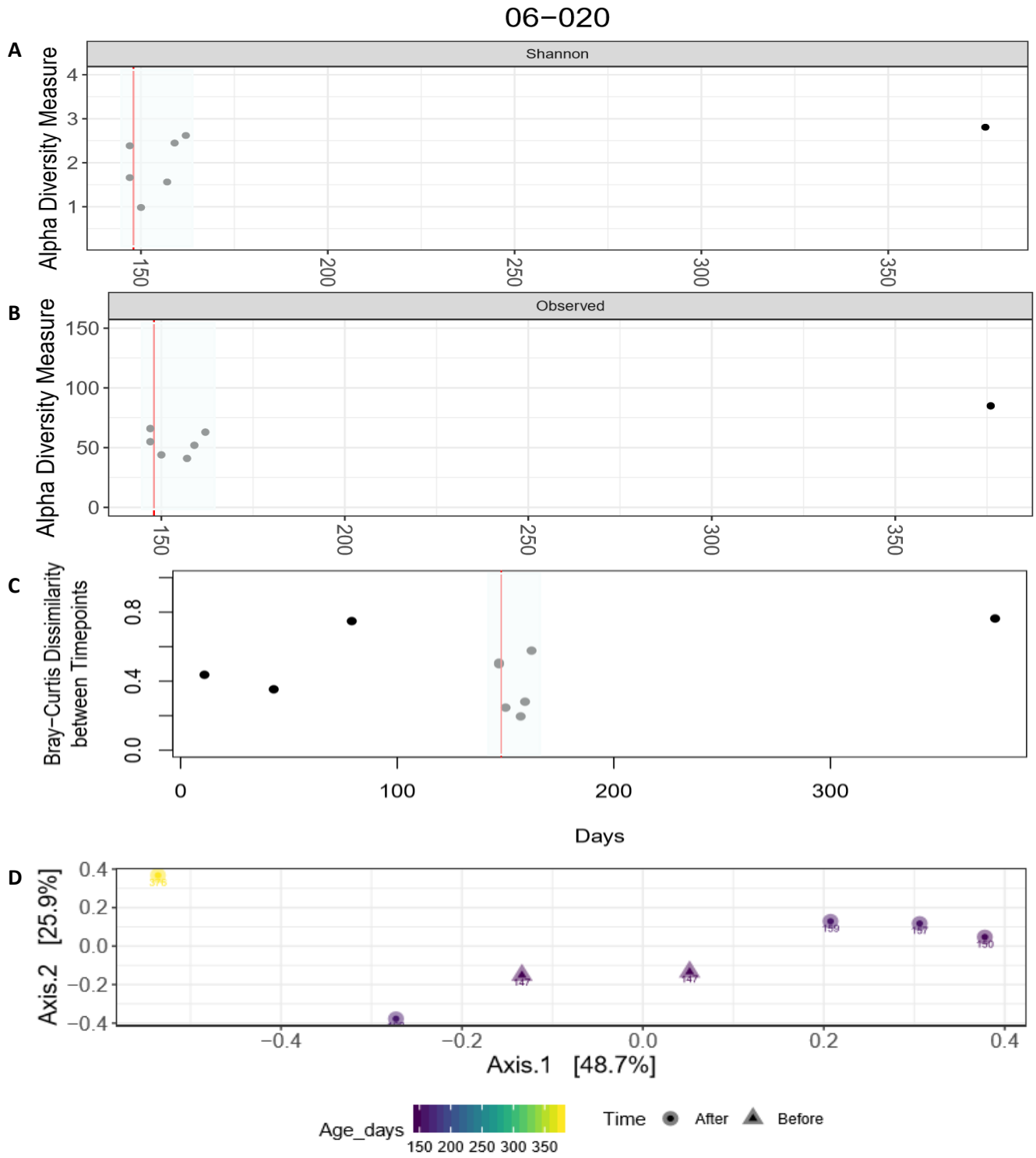


Figure 57: A: Shannon alpha diversity for samples > 90 days old, B: Observed alpha diversity for samples > 90 days, C: Change in Beta - Diversity from sample to sample, D: PCoA plots for samples > 90 days old. All for participant 06-020.

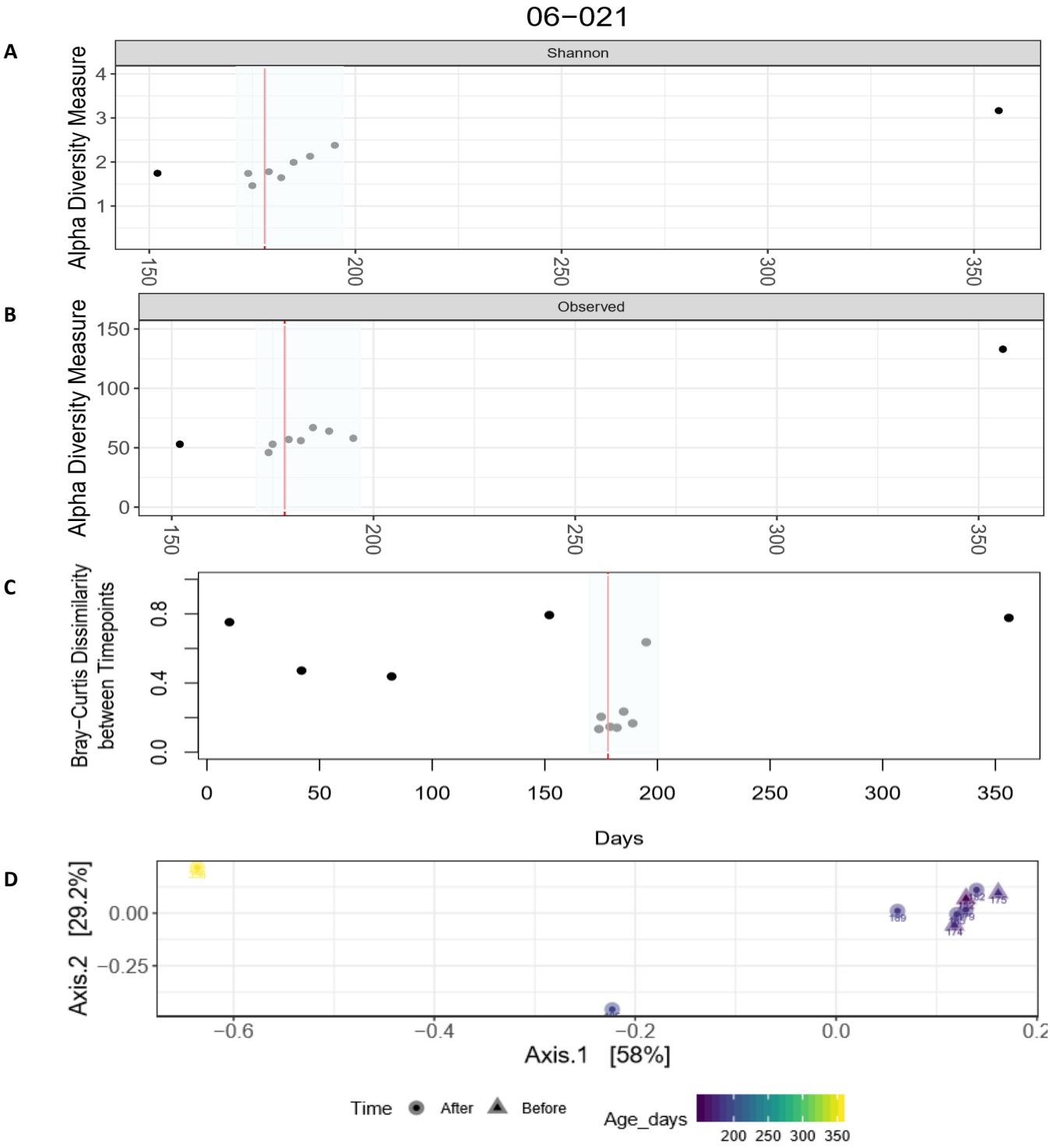


Figure 58: A: Shannon alpha diversity for samples > 90 days old, B: Observed alpha diversity for samples > 90 days, C: Change in Beta - Diversity from sample to sample, D: PCoA plots for samples > 90 days old. All for participant 06-021.

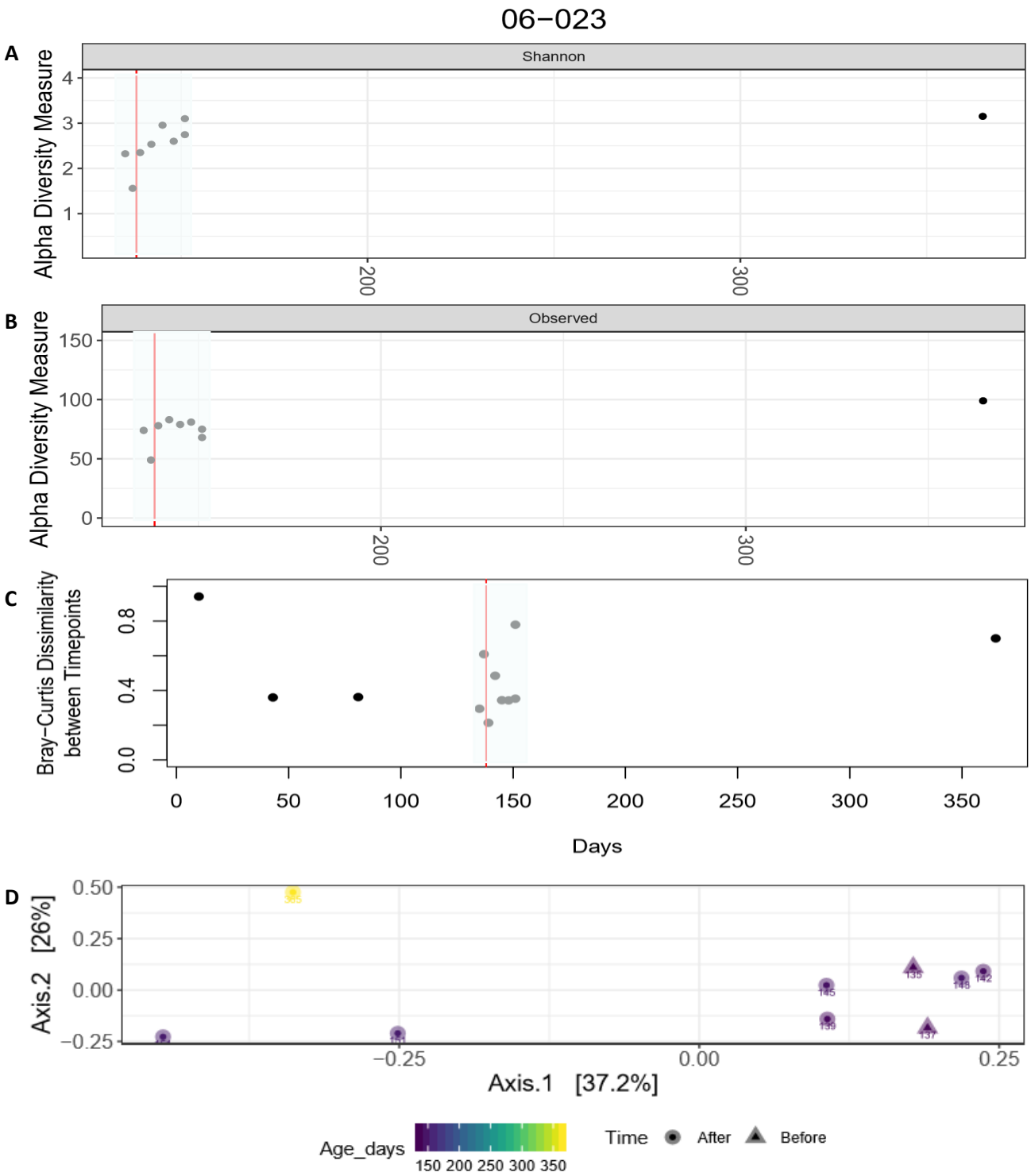


Figure 59: A: Shannon alpha diversity for samples > 90 days old, B: Observed alpha diversity for samples > 90 days, C: Change in Beta - Diversity from sample to sample, D: PCoA plots for samples > 90 days old. All for participant 06-023.

Individual Analysis – Heat Maps (Relative Abundance)

Heat maps showing the relative abundances of the bacteria, on the phylum, genus and ASV level.

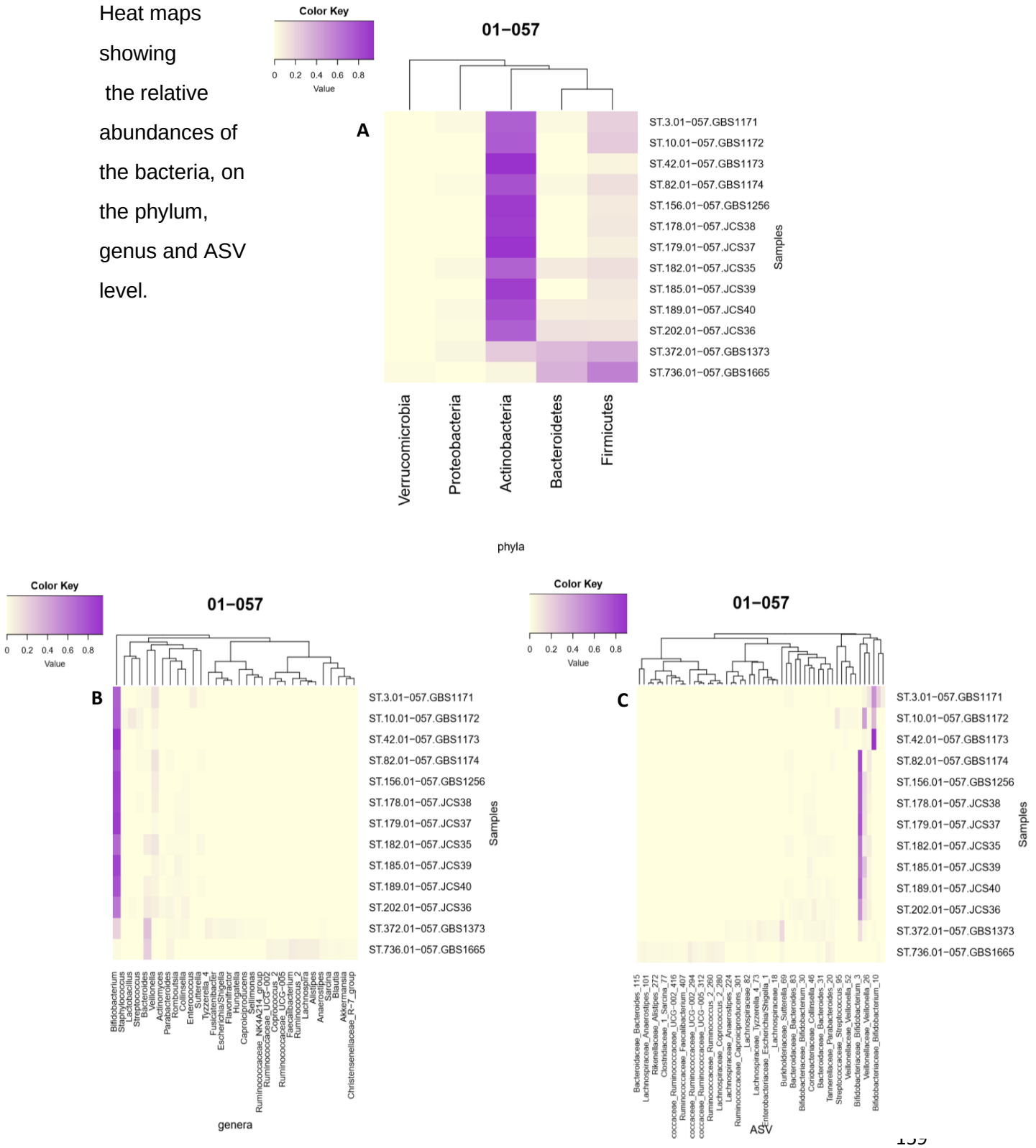


Figure 60: Heat maps of relative abundance for 01-057. A: Phylum level, B: Genus level, C: ASV level

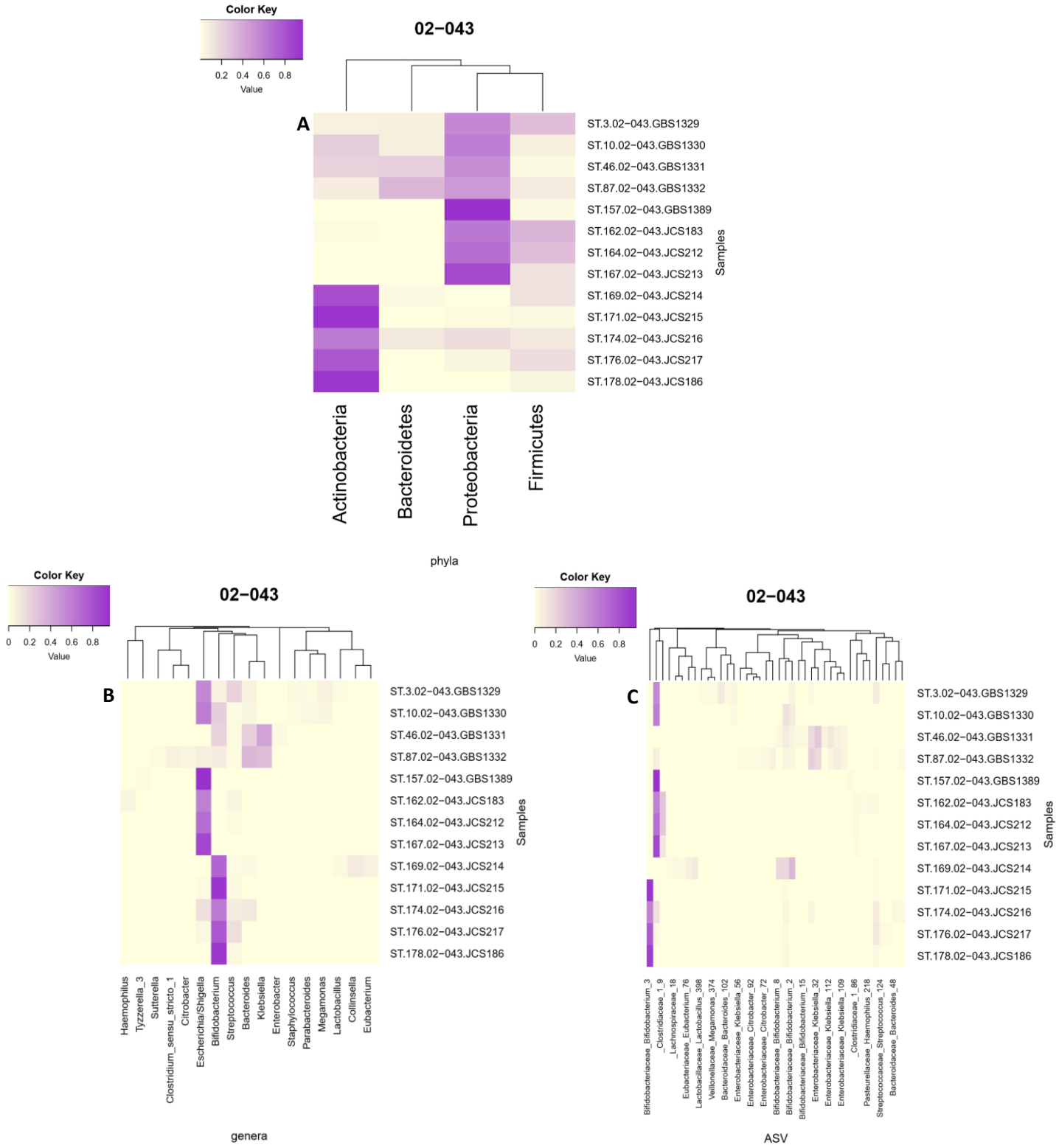


Figure 61: Heat maps of relative abundance for 02-043. A: Phylum level, B: Genus level, C: ASV level

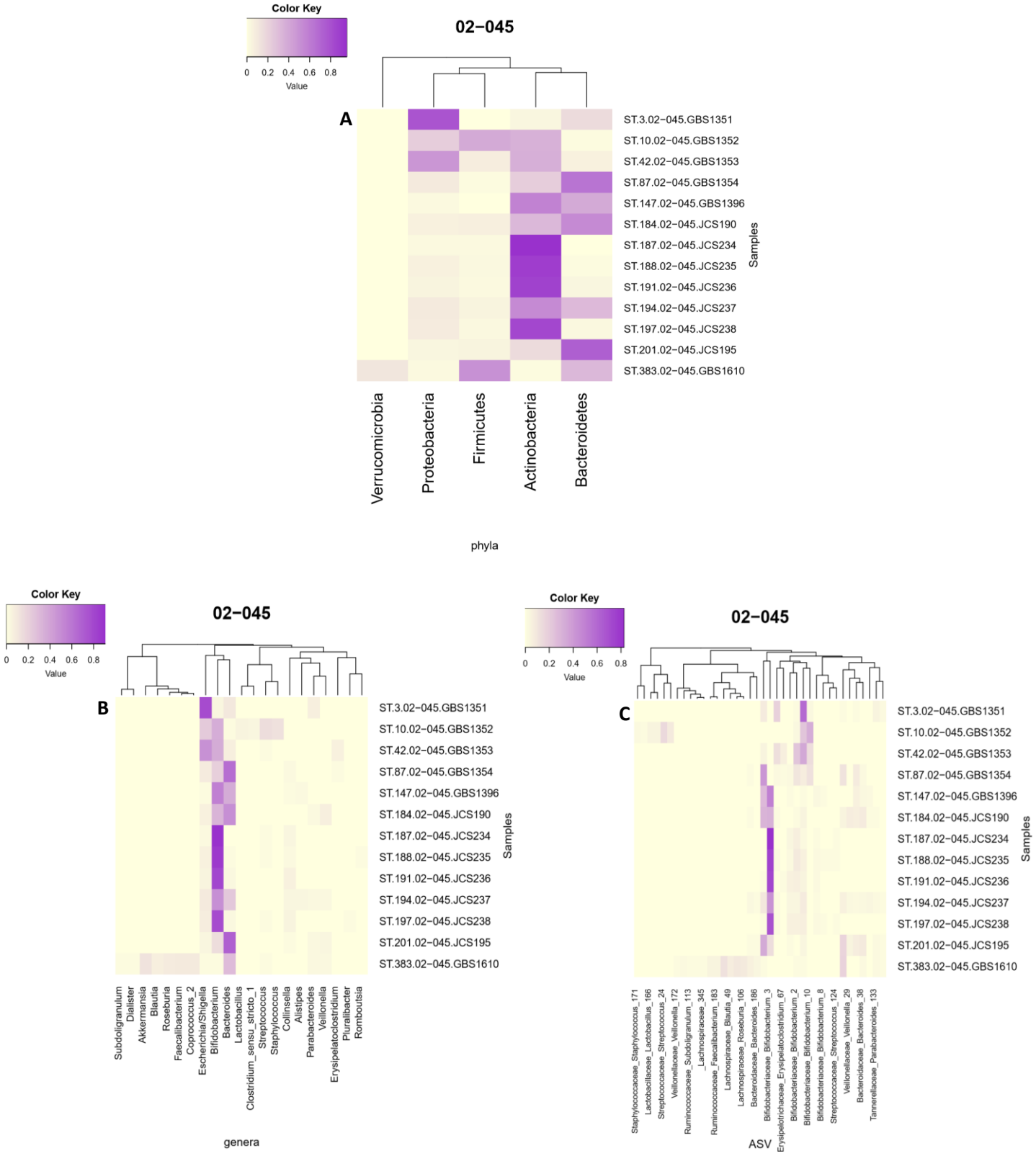


Figure 62: Heat maps of relative abundance for 02-045. A: Phylum level, B: Genus level, C: ASV level

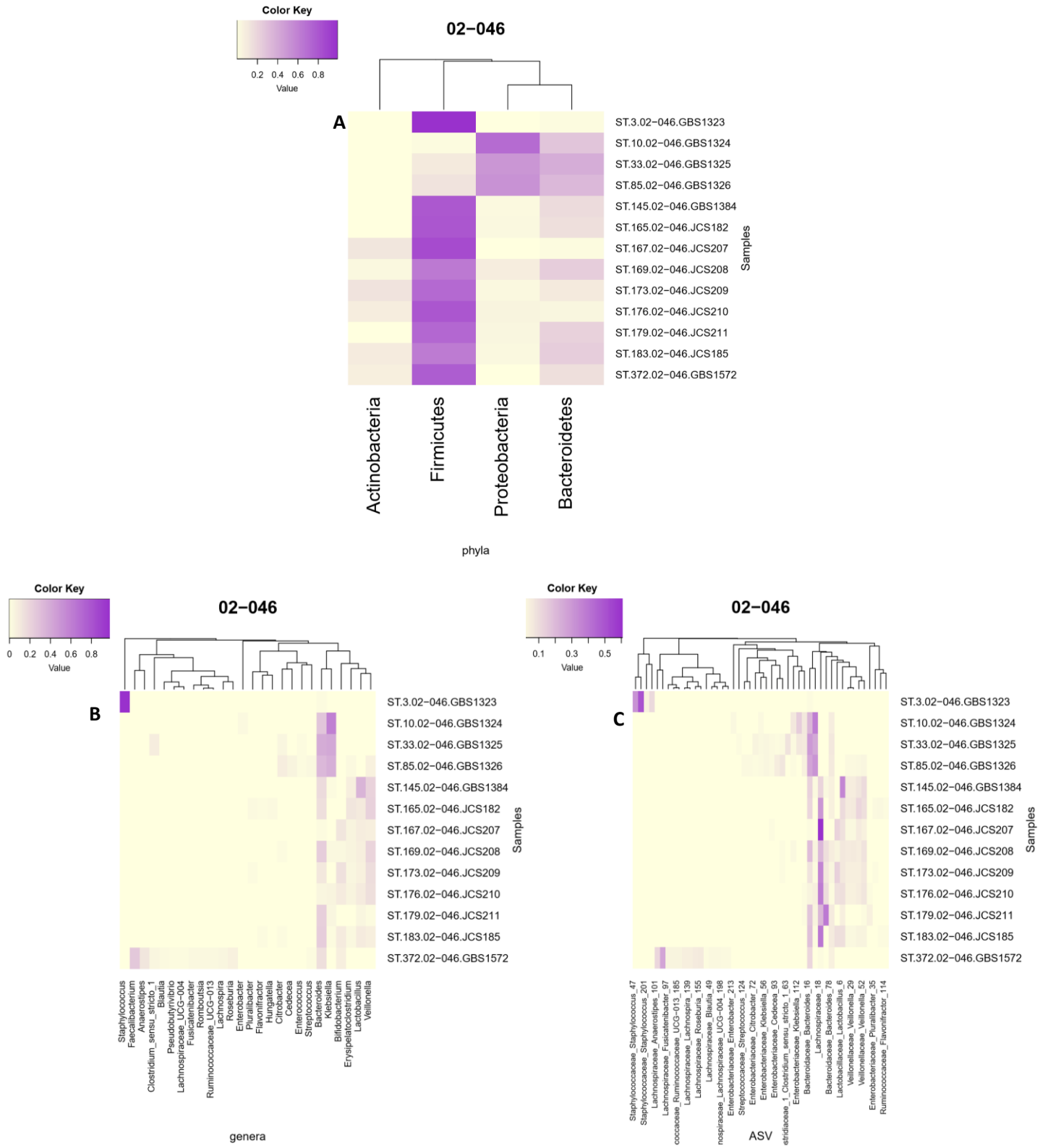


Figure 63:Heat maps of relative abundance for 02-046. A: Phylum level, B: Genus level, C: ASV level

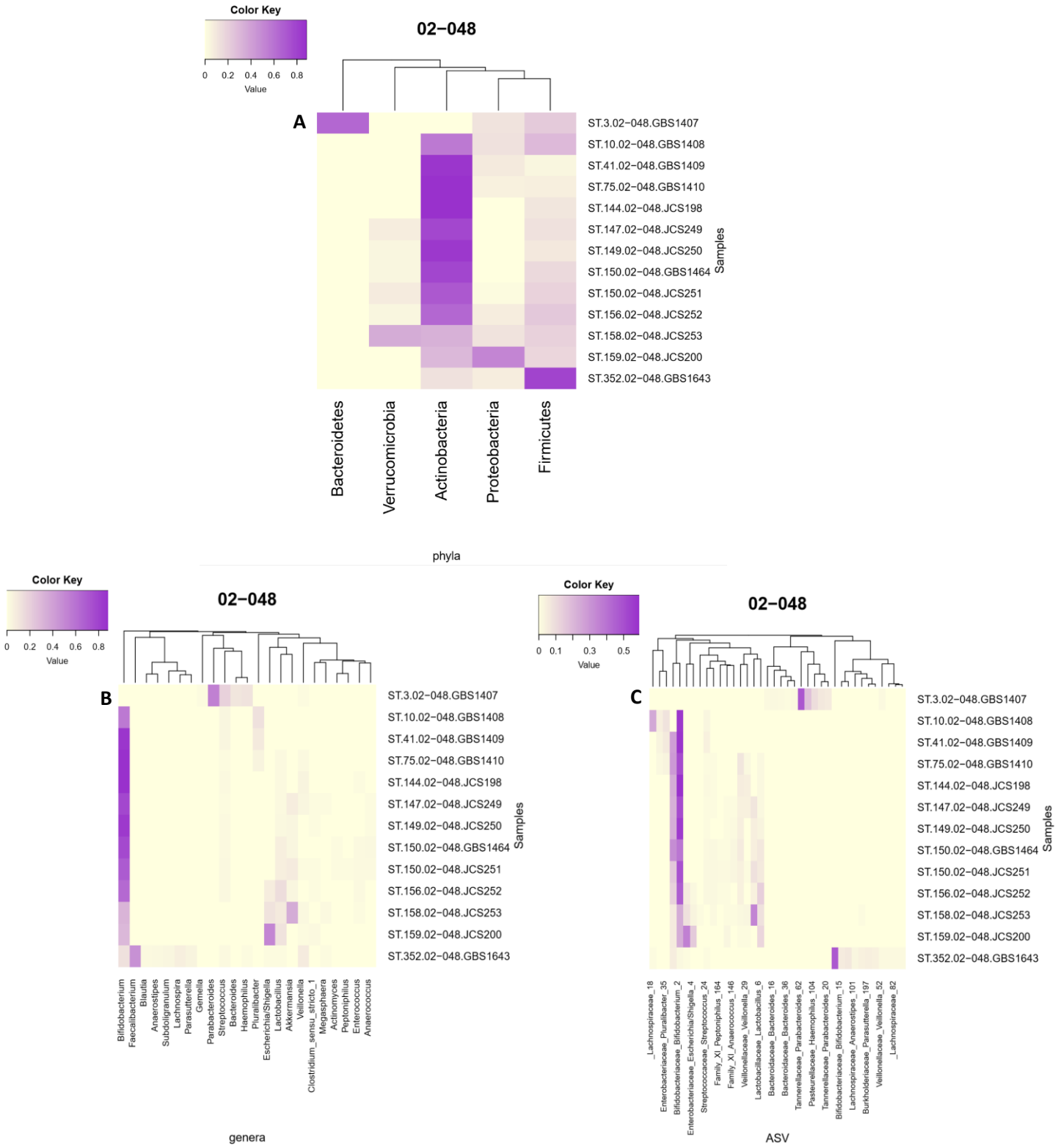


Figure 64: Heat maps of relative abundance for 02-048. A: Phylum level, B: Genus level, C: ASV level

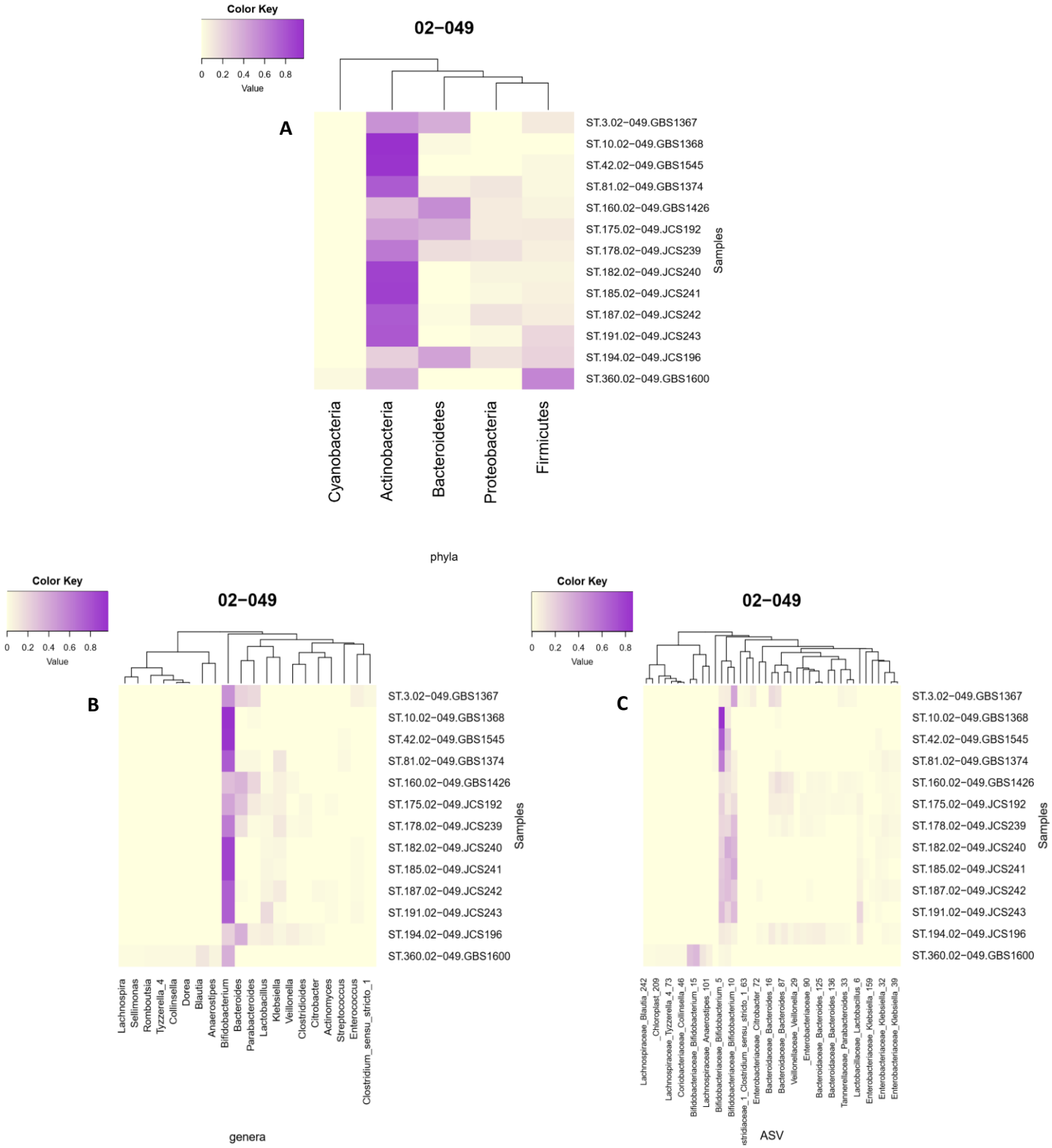


Figure 65: Heat maps of relative abundance for 02-049. A: Phylum level, B: Genus level, C: ASV level

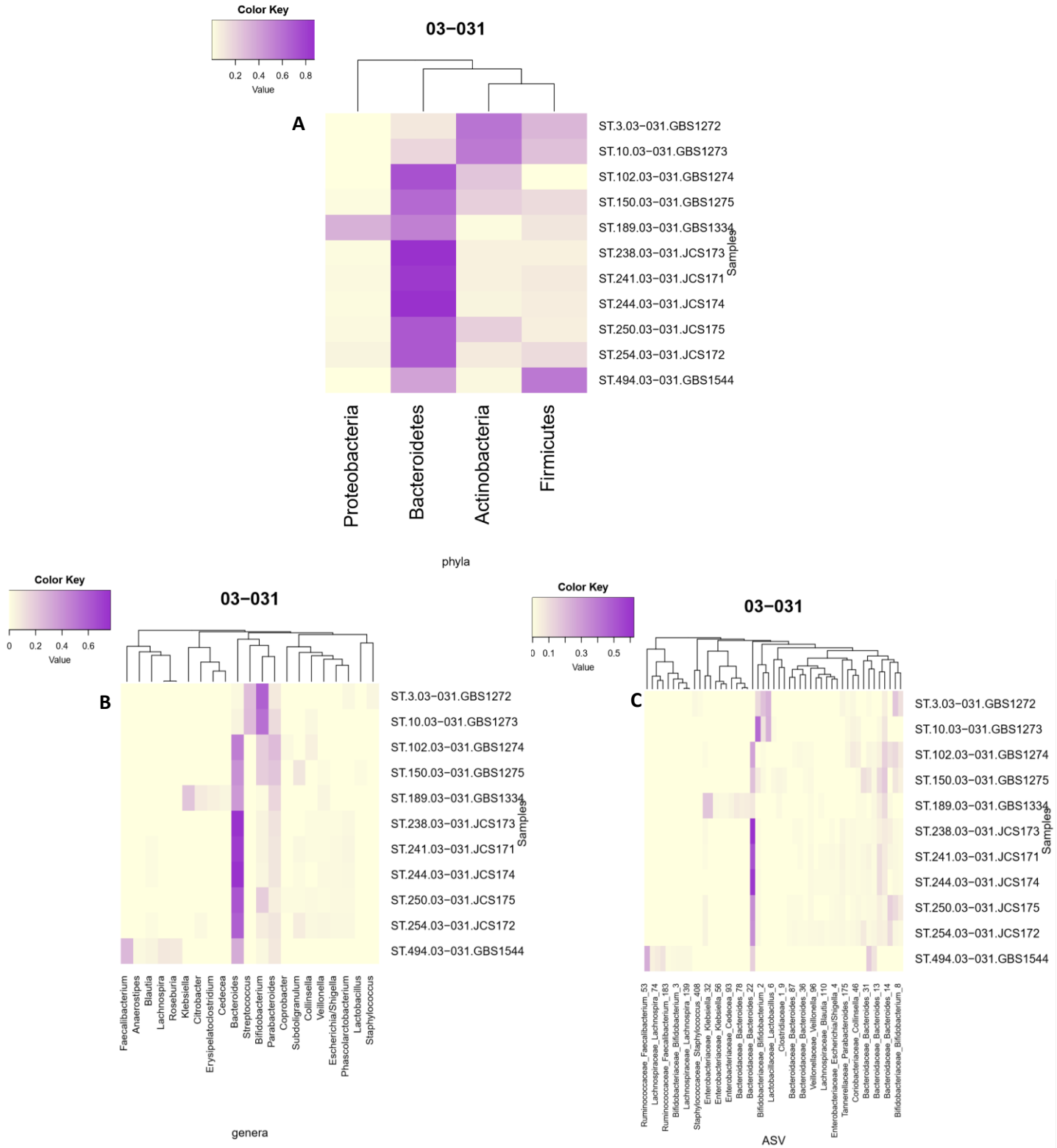


Figure 66: Heat maps of relative abundance for 03-031. A: Phylum level, B: Genus level, C: ASV level

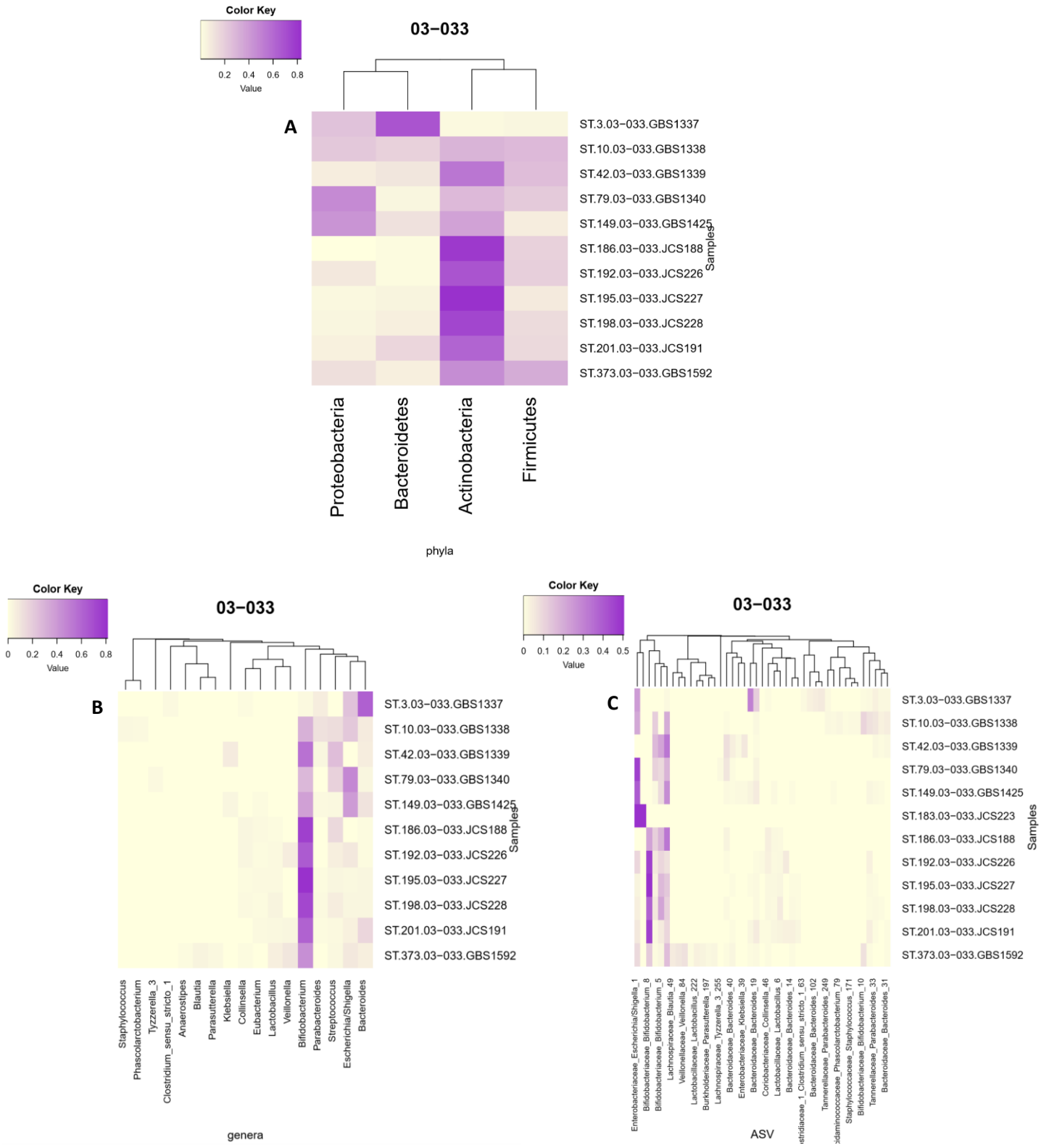


Figure 67: Heat maps of relative abundance for 03-033. A: Phylum level, B: Genus level, C: ASV level

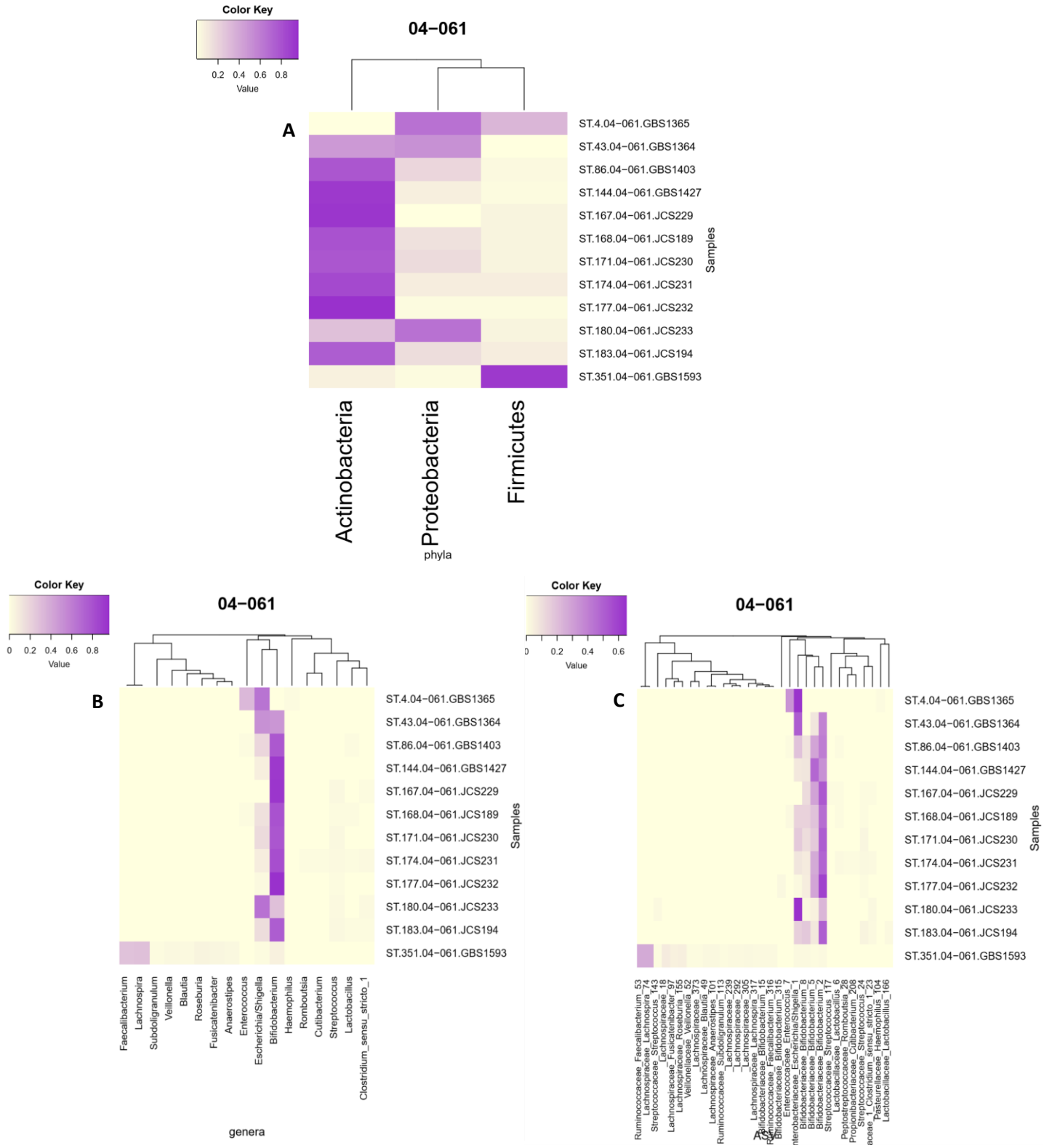


Figure 68: Heat maps of relative abundance for 04-061. A: Phylum level, B: Genus level, C: ASV level

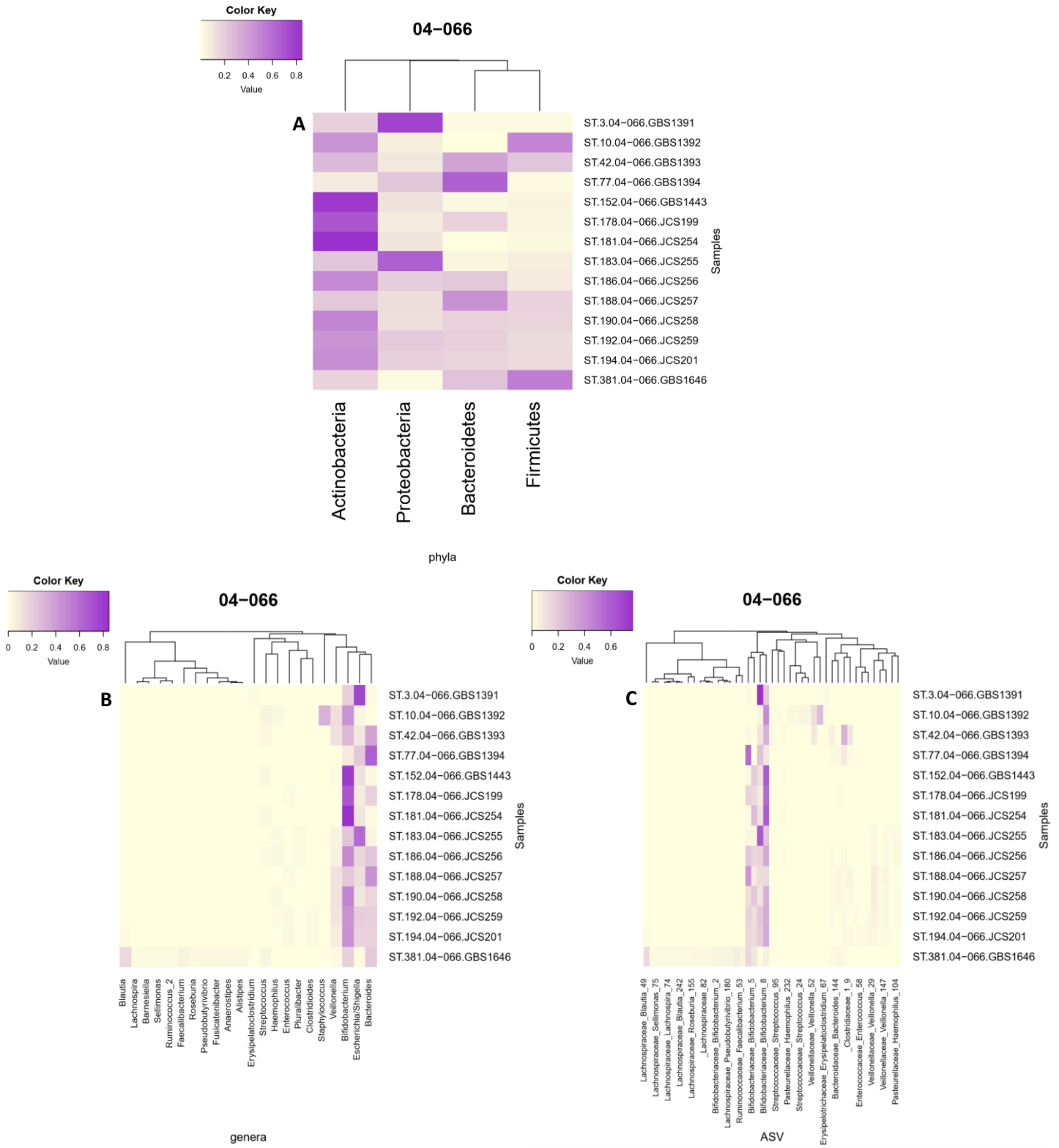


Figure 69: Heat maps of relative abundance for 04-066. A: Phylum level, B: Genus level, C: ASV level

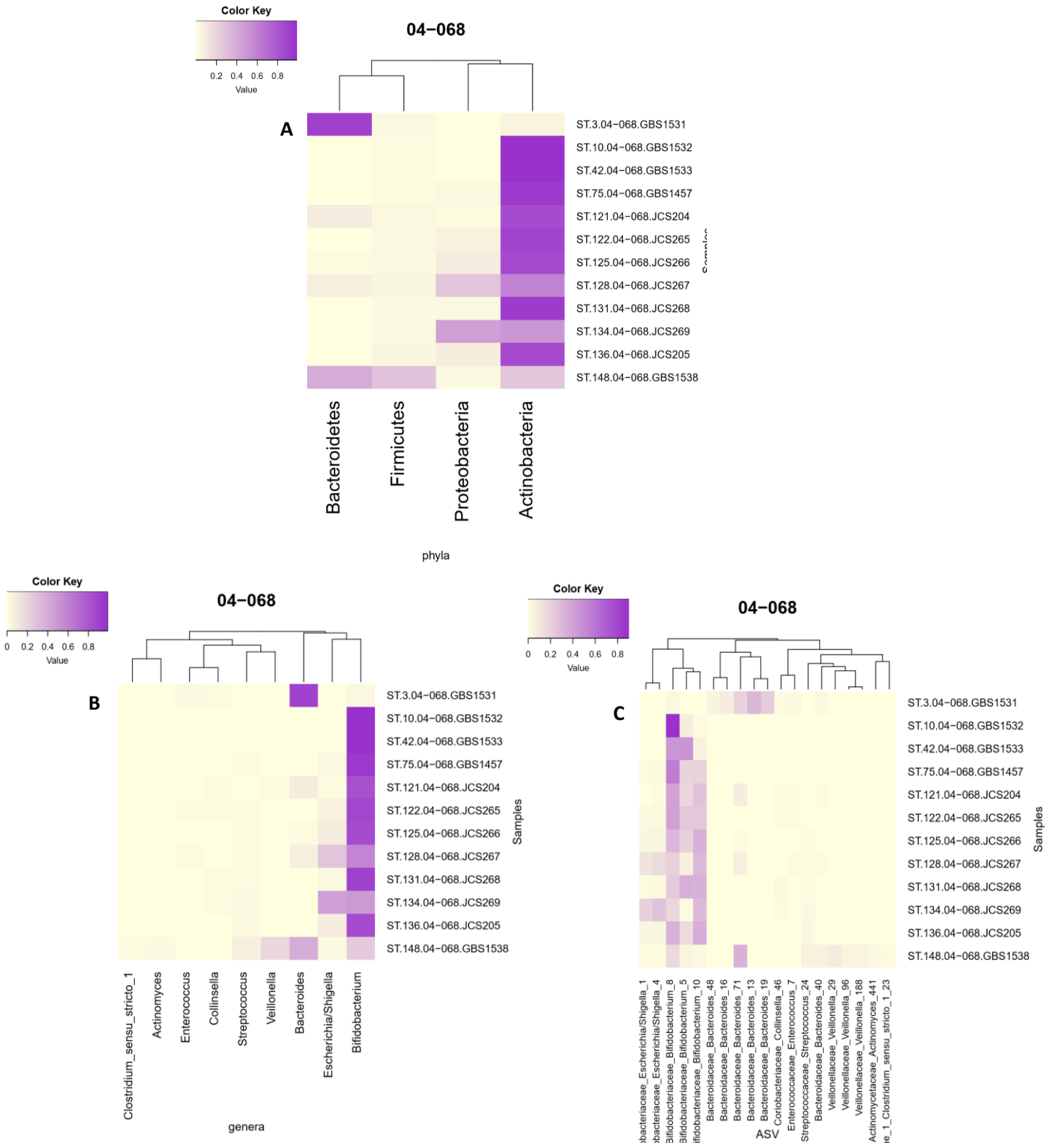


Figure 70: Heat maps of relative abundance for 04-068. A: Phylum level, B: Genus level, C: ASV level

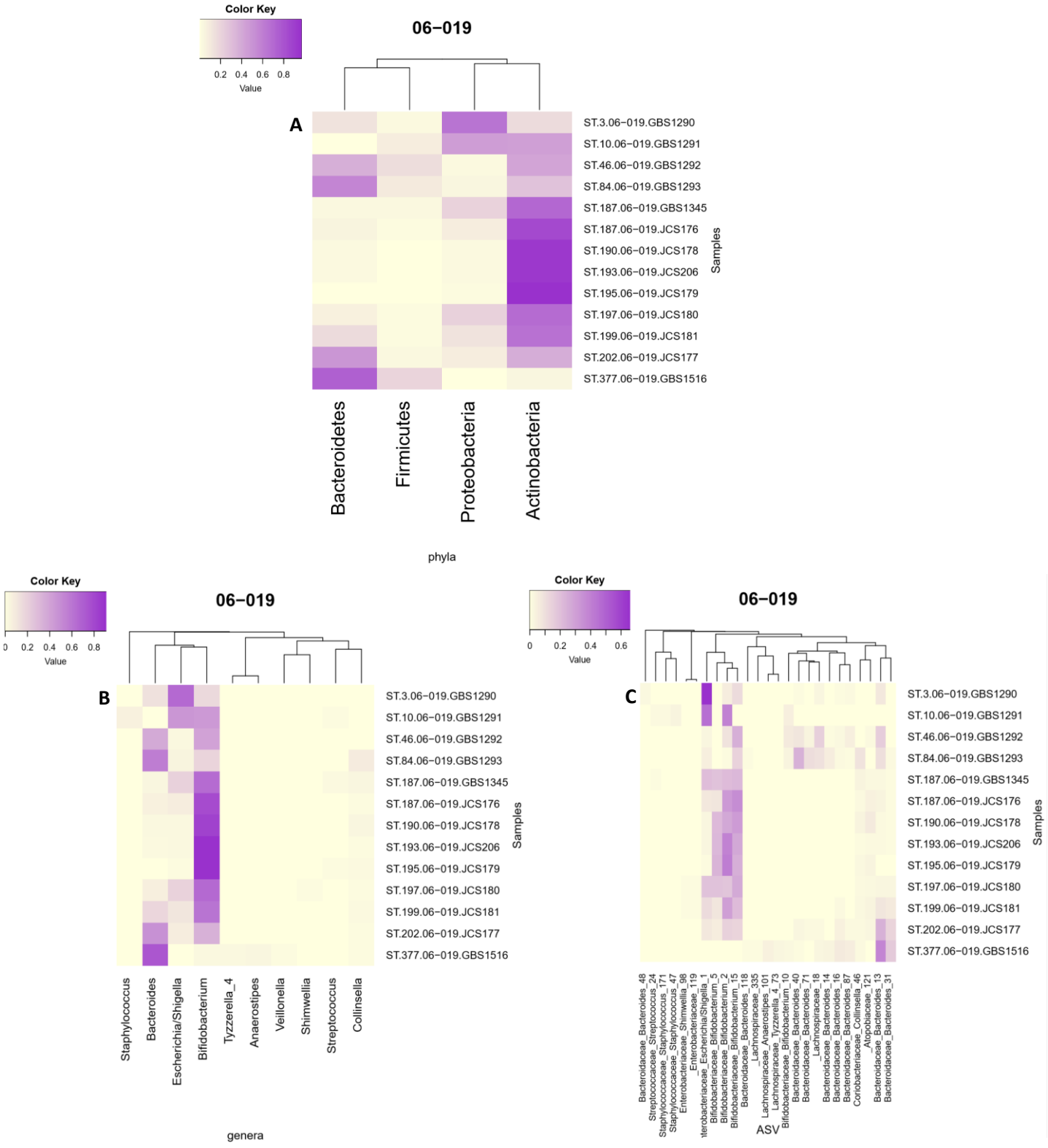


Figure 71: Heat maps of relative abundance for 06-019. A: Phylum level, B: Genus level, C: ASV level

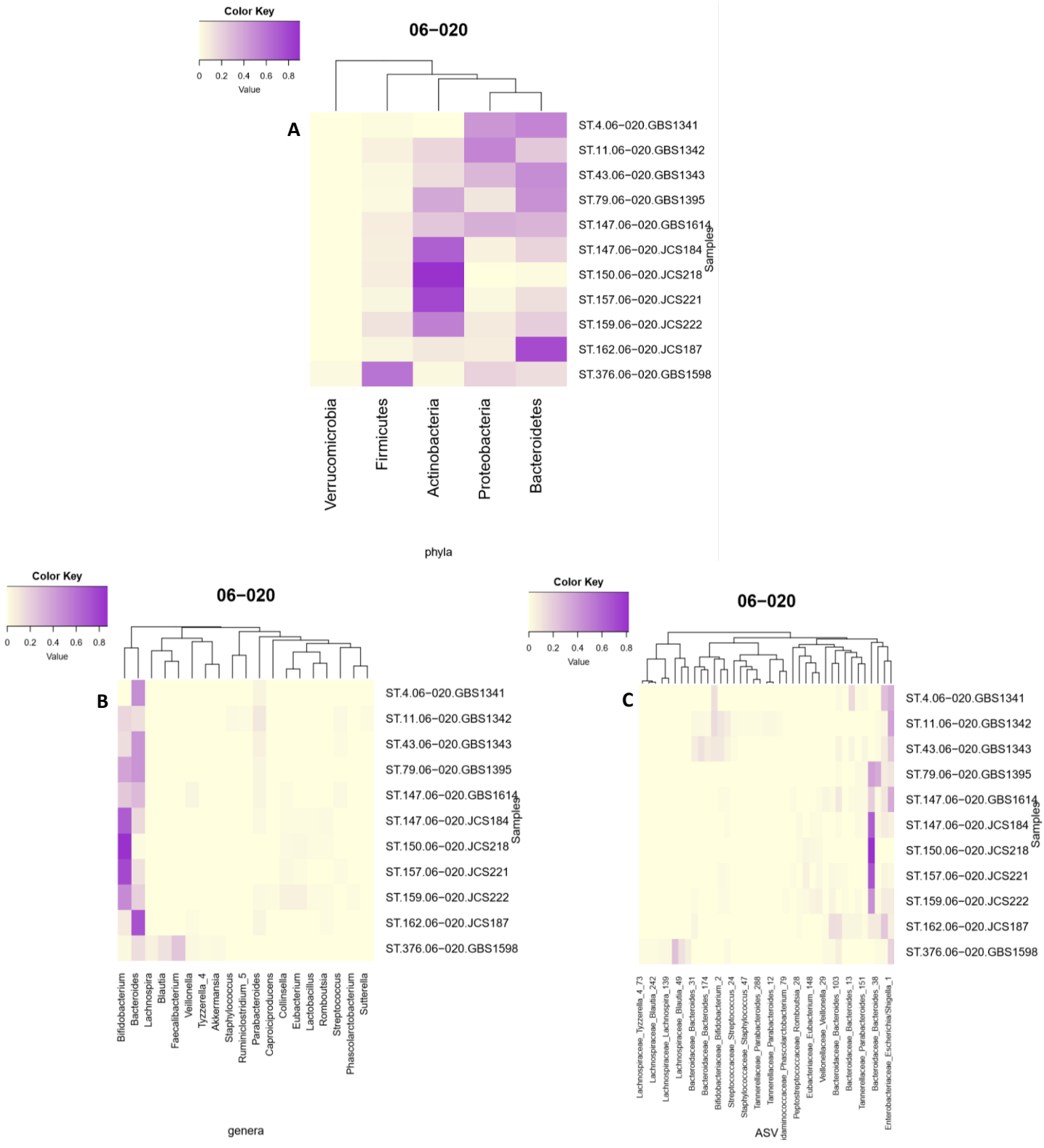


Figure 72: Heat maps of relative abundance for 06-020. A: Phylum level, B: Genus level, C: ASV level

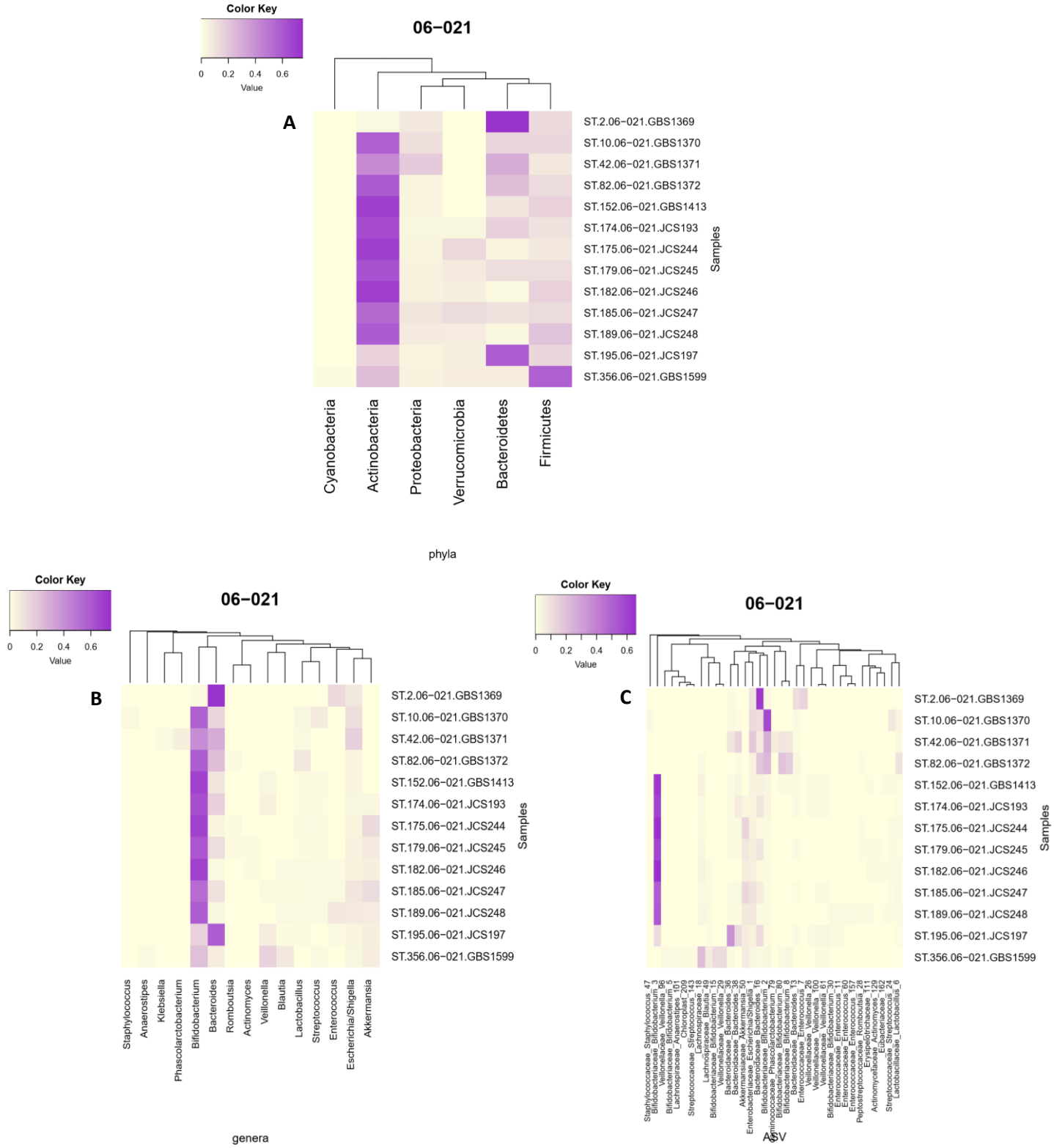


Figure 73: Heat maps of relative abundance for 06-021. A: Phylum level, B: Genus level, C: ASV level

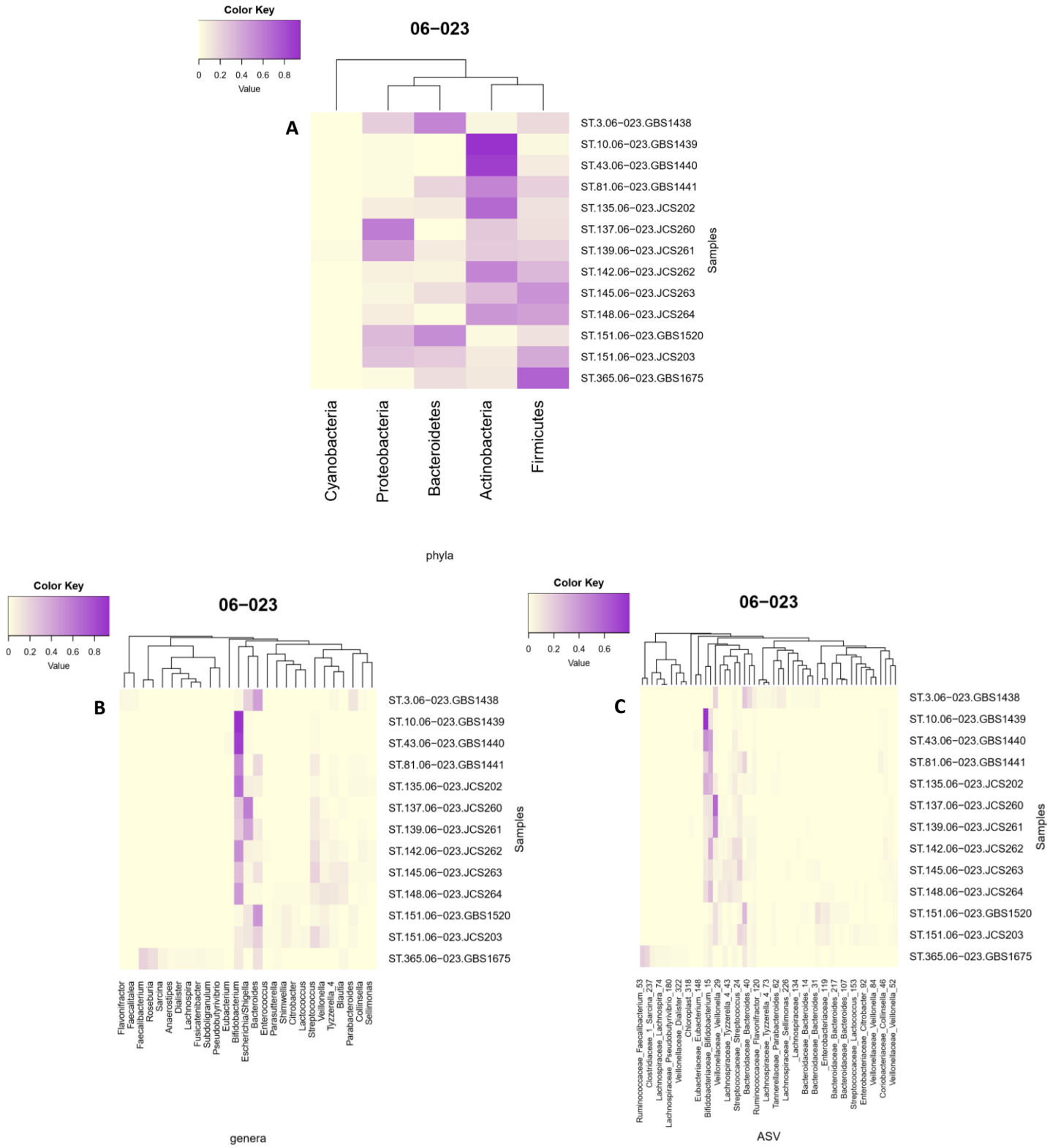


Figure 74: Heat maps of relative abundance for 06-023. A: Phylum level, B: Genus level, C: ASV level

Individual Analysis – Taxa Bar Charts

The following figures show relative abundances of bacterial ASVs in a different form than as heat maps. The top 25 ASVs are named in the legend.

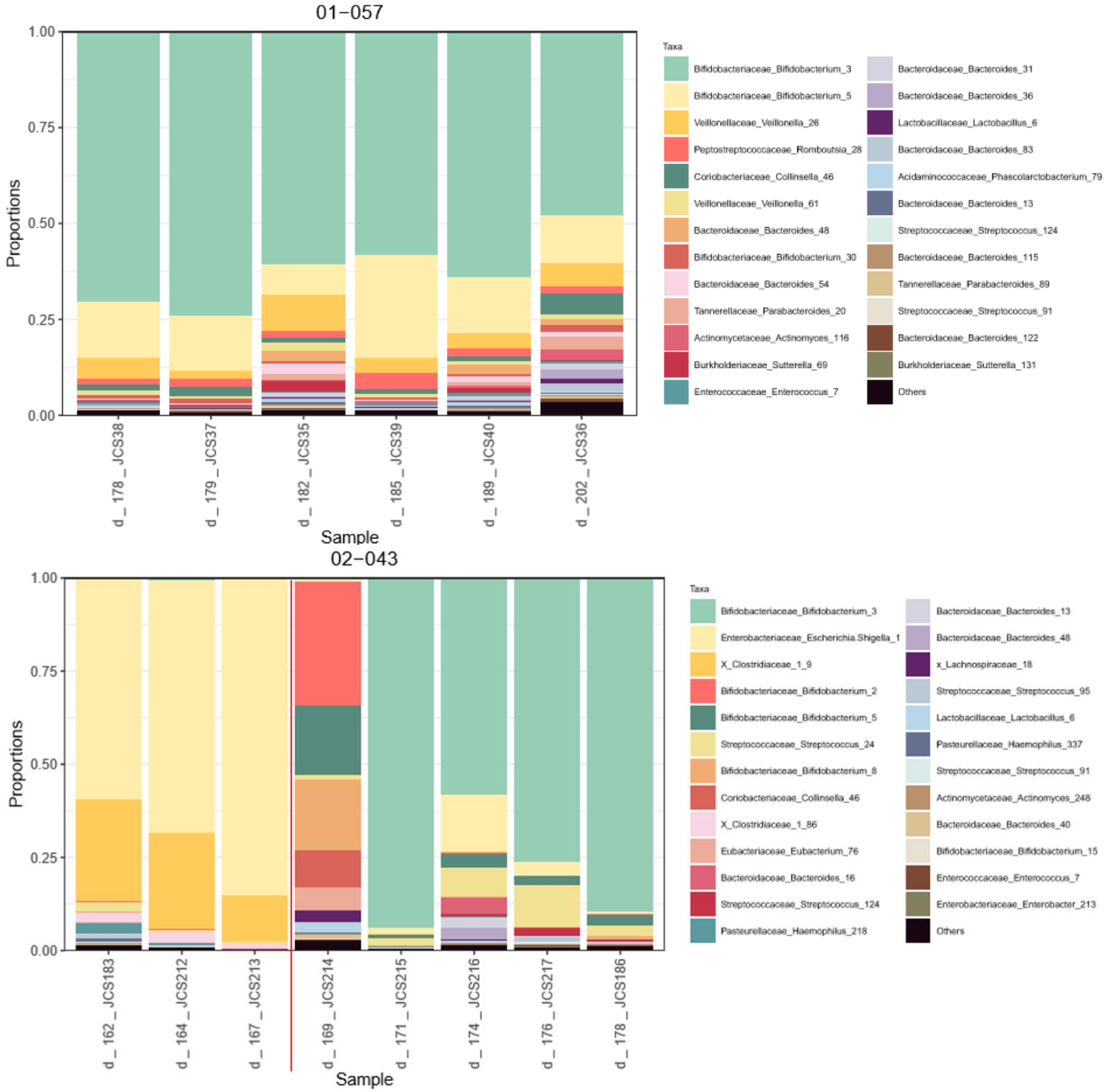


Figure 76: Taxa bar chart of relative abundance for the top 25 bacterial ASVs for 02-043

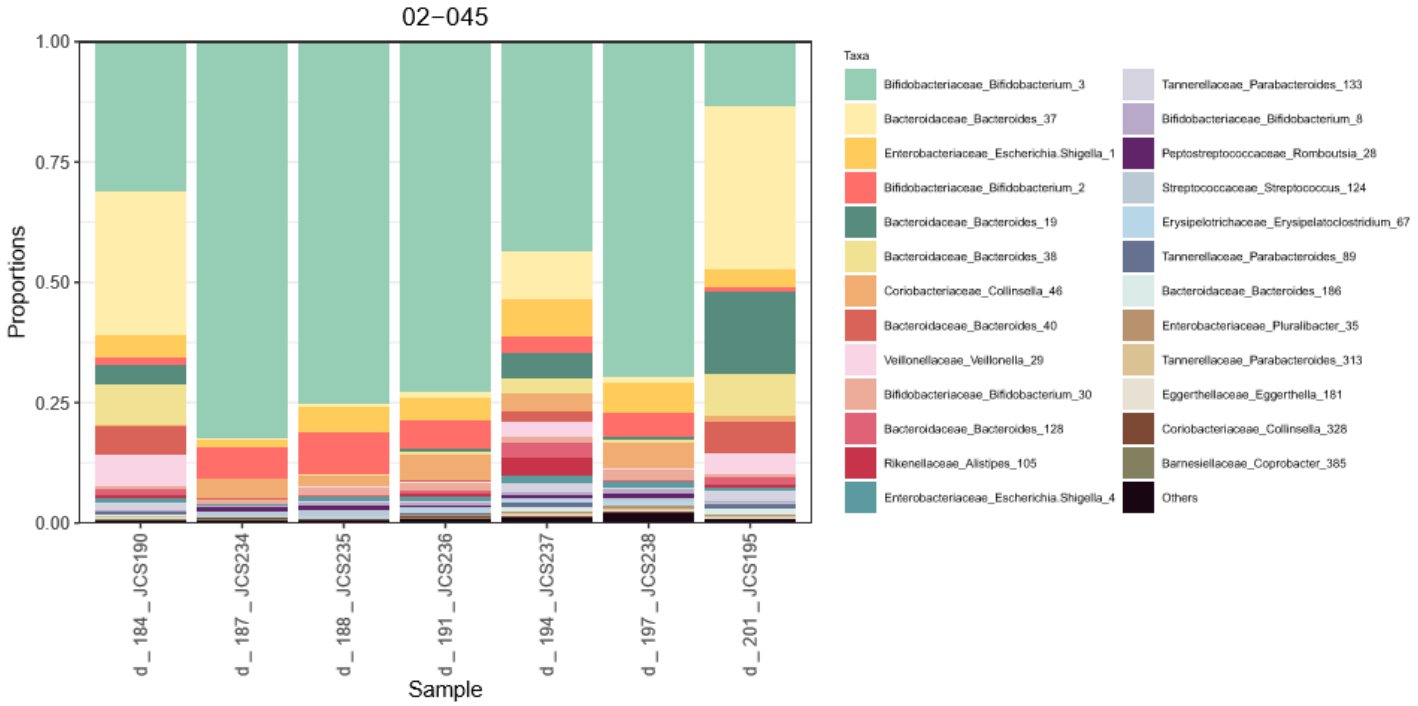


Figure 78: Taxa bar chart of relative abundance for the top 25 bacterial ASVs for 02-045.

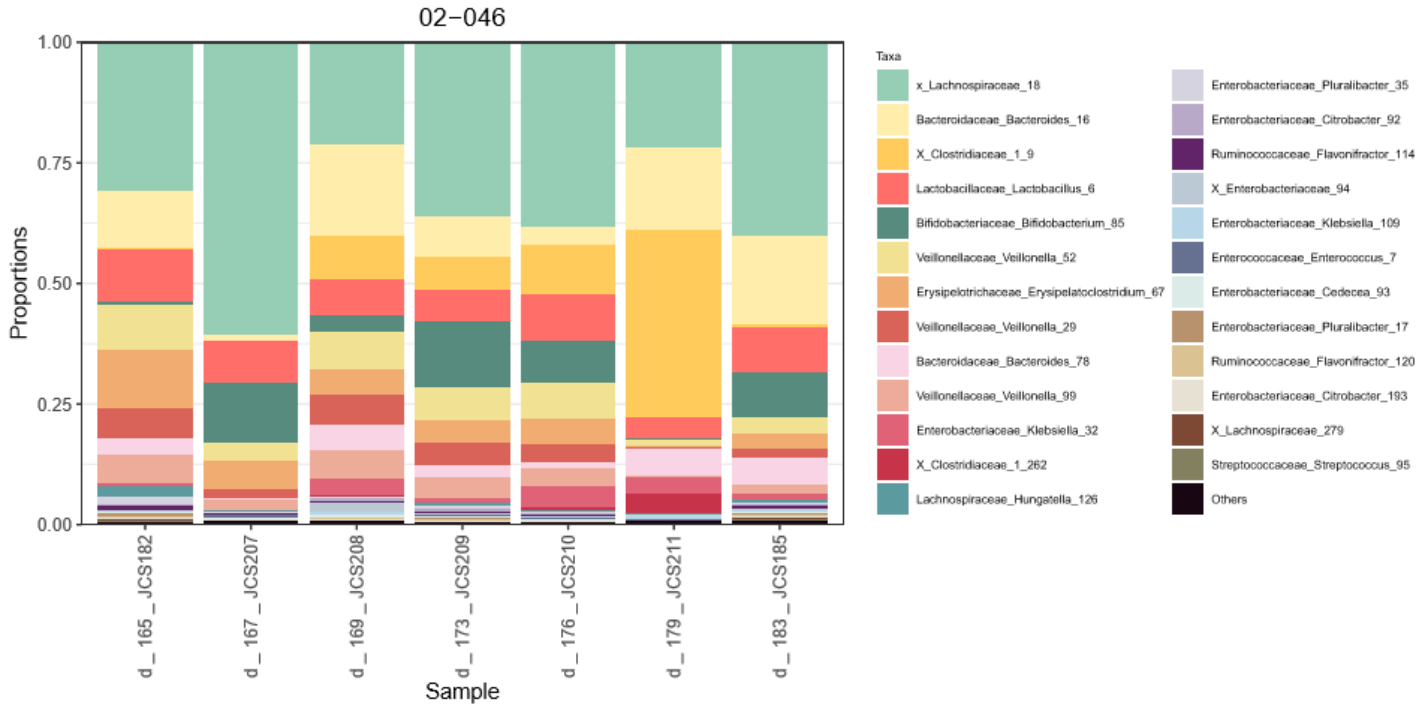


Figure 77: Taxa bar chart of relative abundance for the top 25 bacterial ASVs for 02-046.

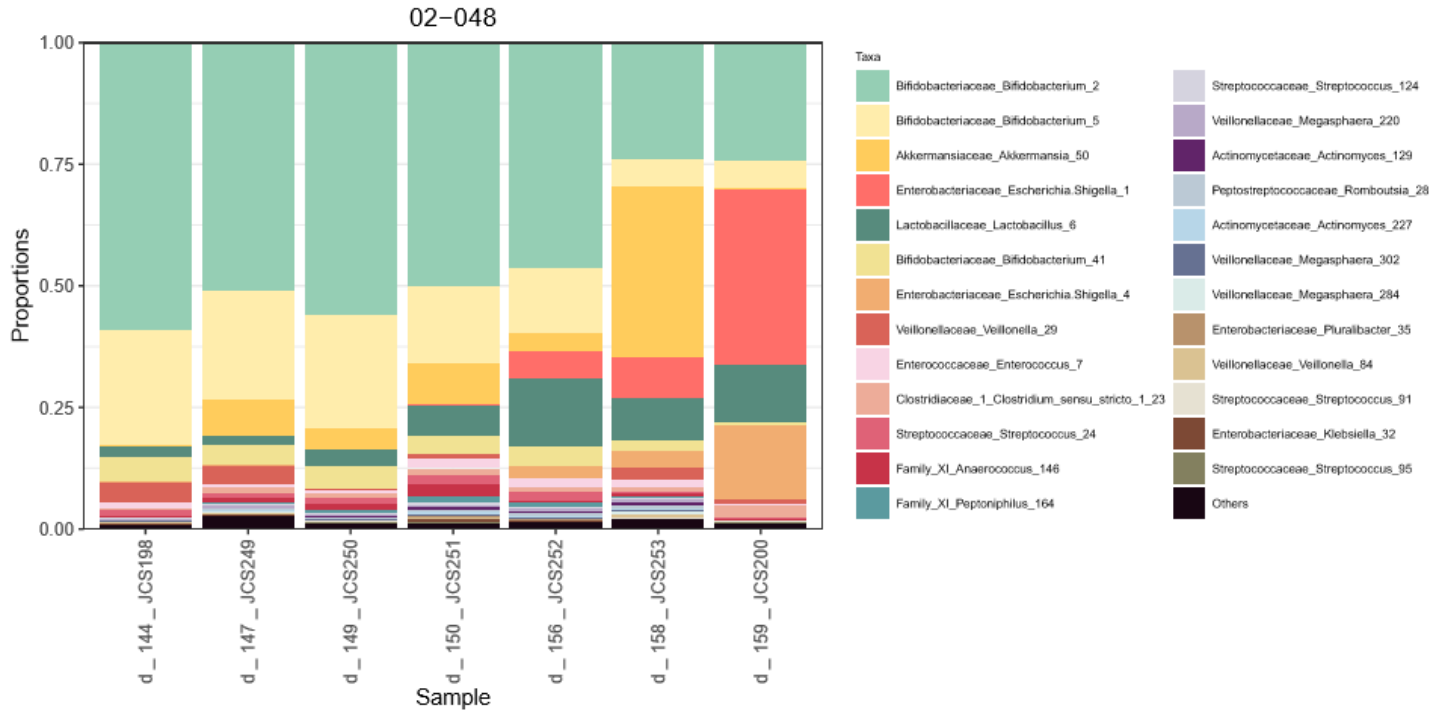


Figure 79: Taxa bar chart of relative abundance for the top 25 bacterial ASVs for 02-048.

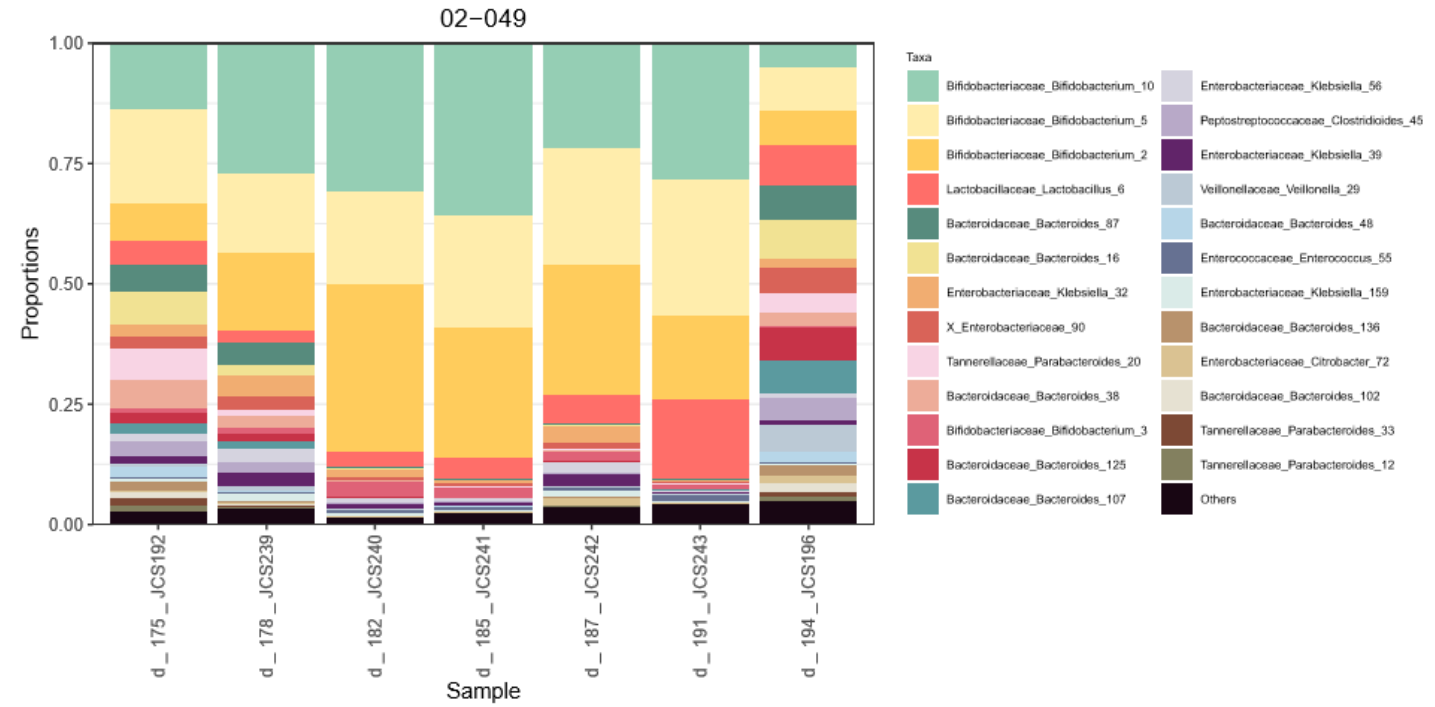


Figure 80: Taxa bar chart of relative abundance for the top 25 bacterial ASVs for 02-049.

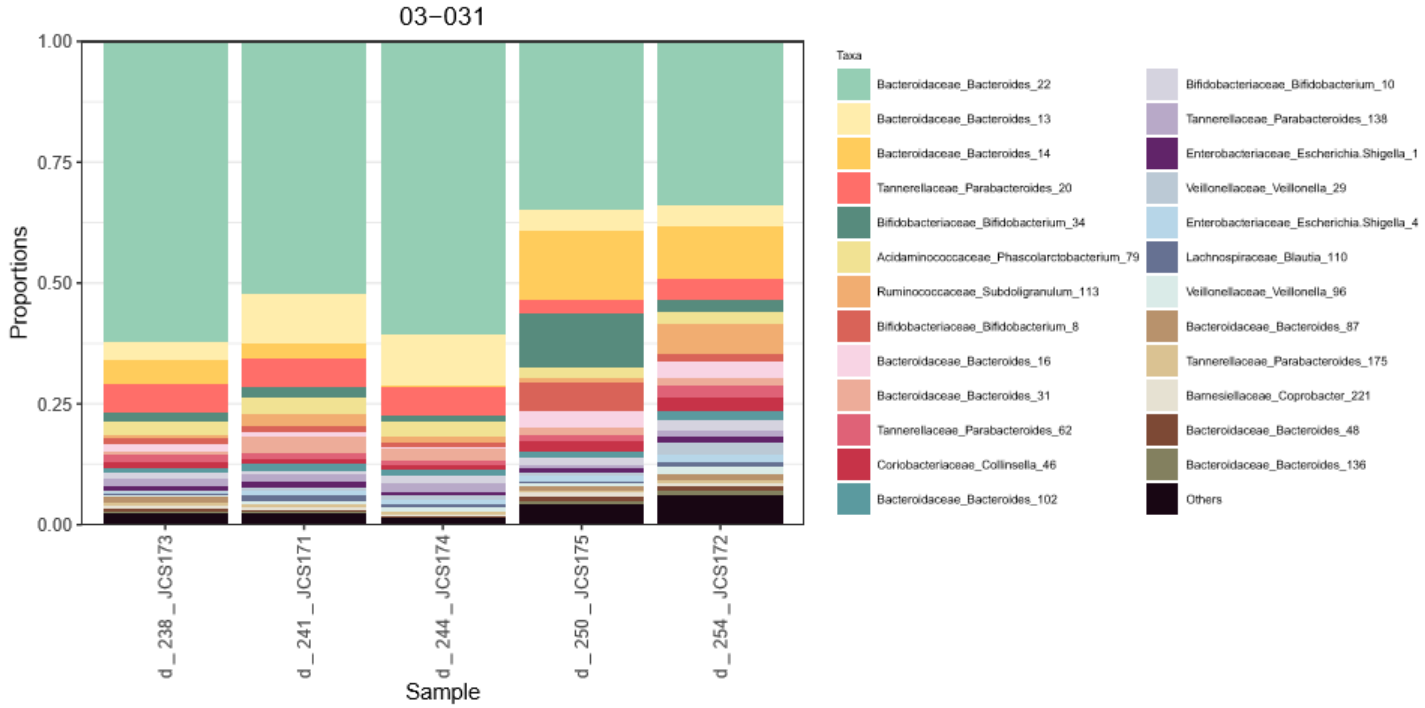


Figure 81: Taxa bar chart of relative abundance for the top 25 bacterial ASVs for 03-031.

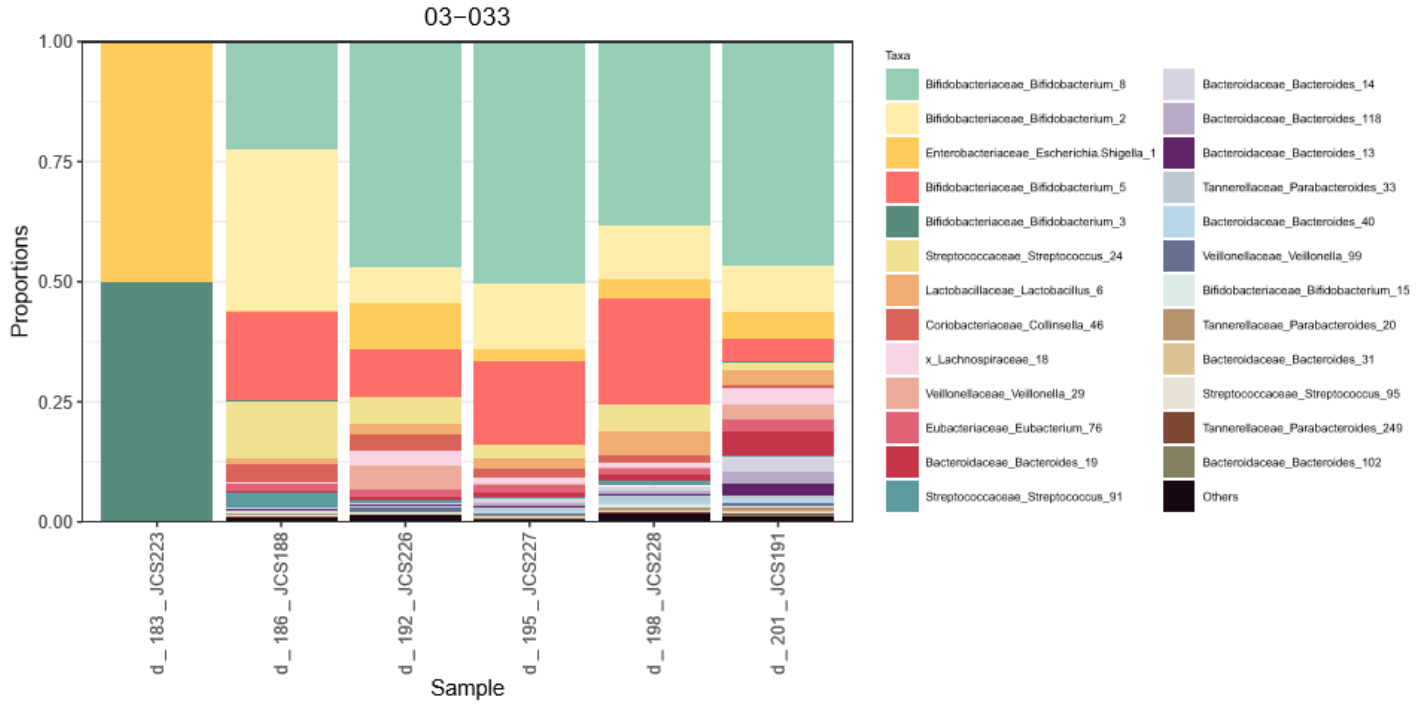


Figure 82: Taxa bar chart of relative abundance for the top 25 bacterial ASVs for 03-033.

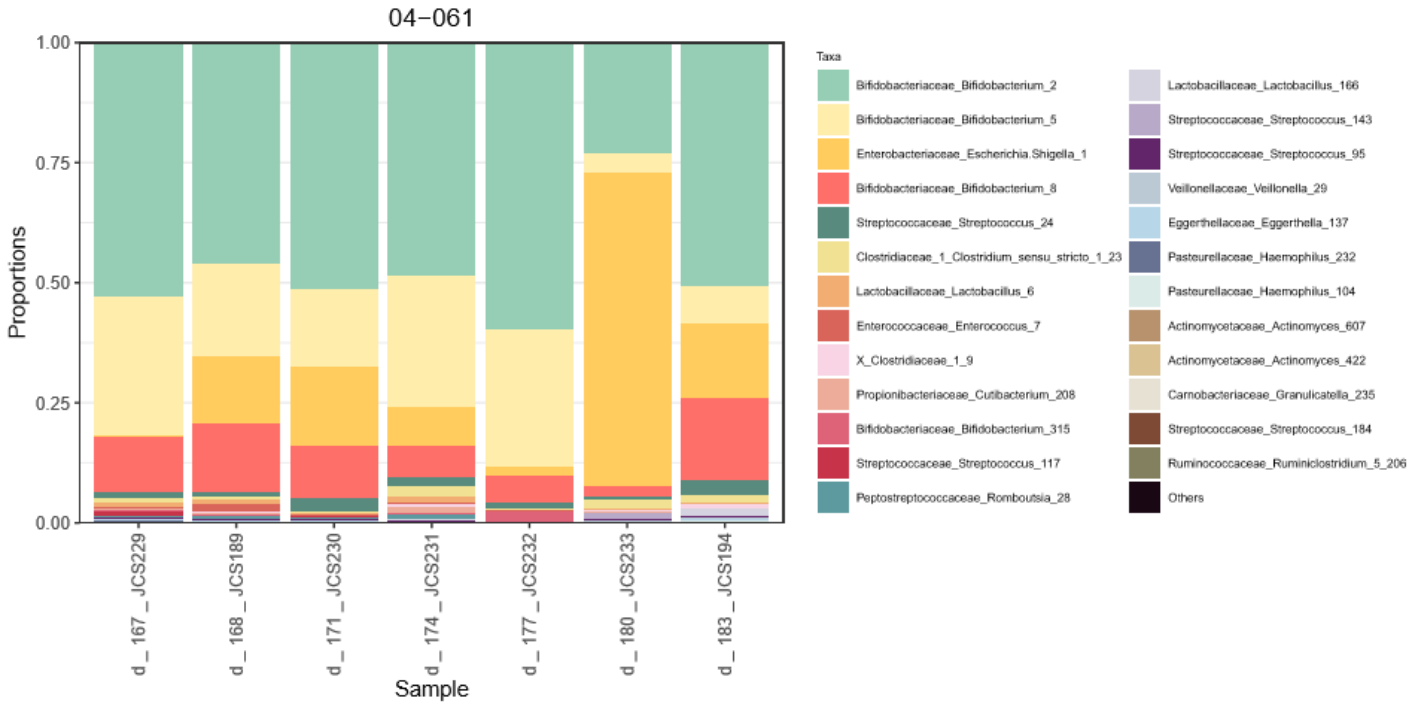


Figure 84: Taxa bar chart of relative abundance for the top 25 bacterial ASVs for 04-061.

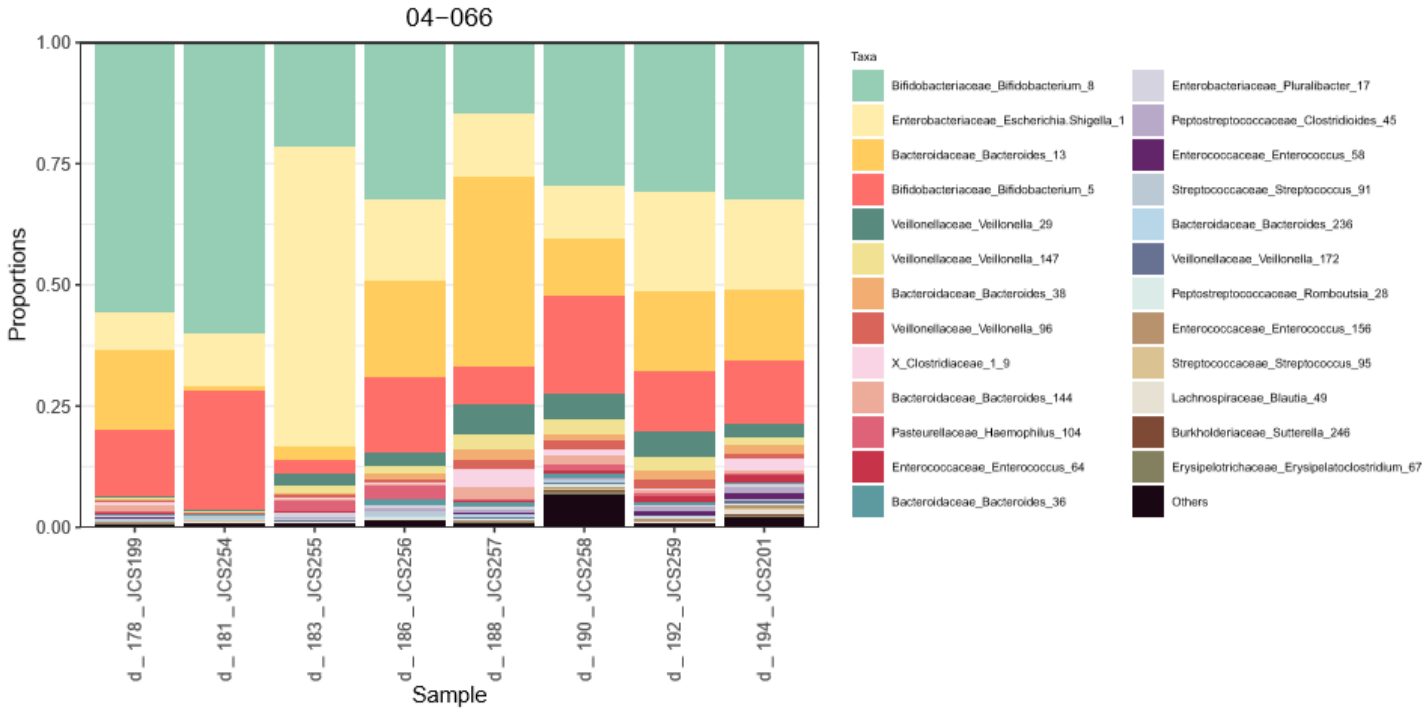


Figure 83: Taxa bar chart of relative abundance for the top 25 bacterial ASVs for 04-066.

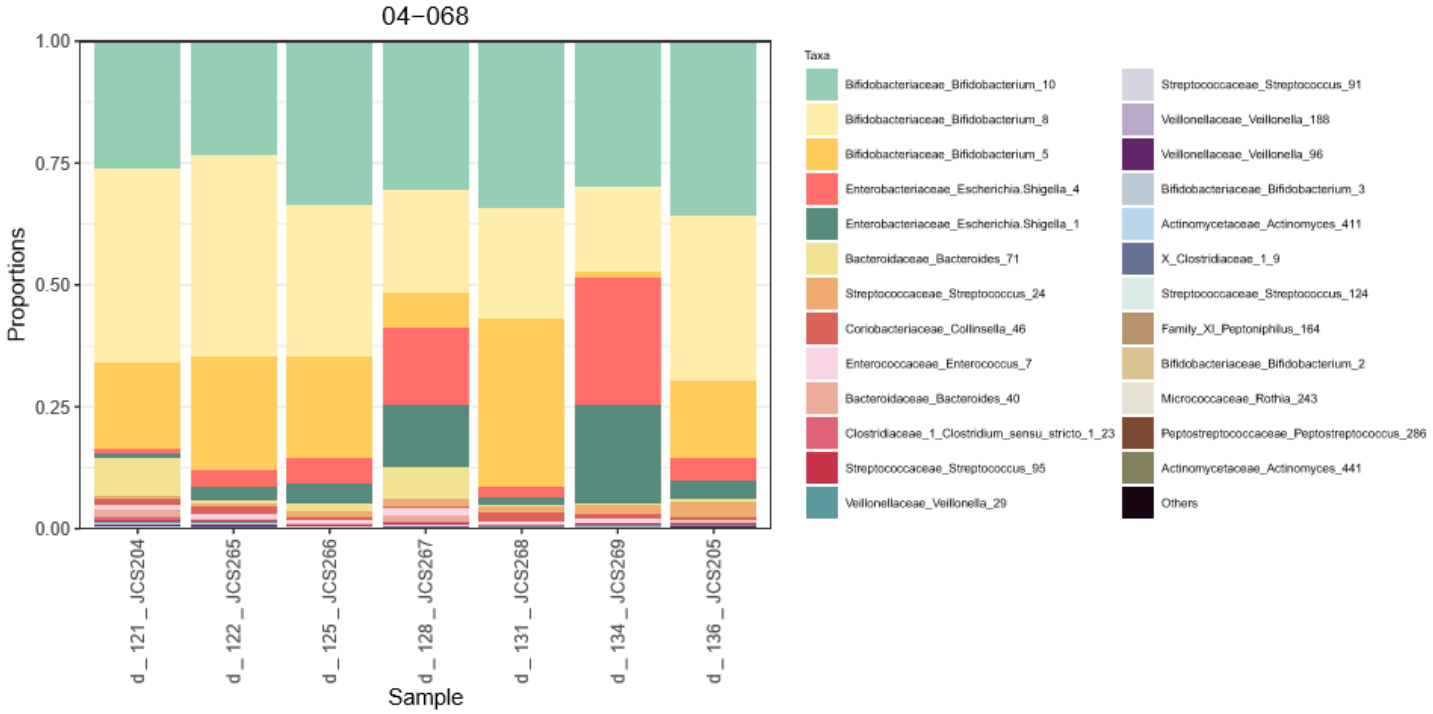


Figure 86:Taxa bar chart of relative abundance for the top 25 bacterial ASVs for 06-019.

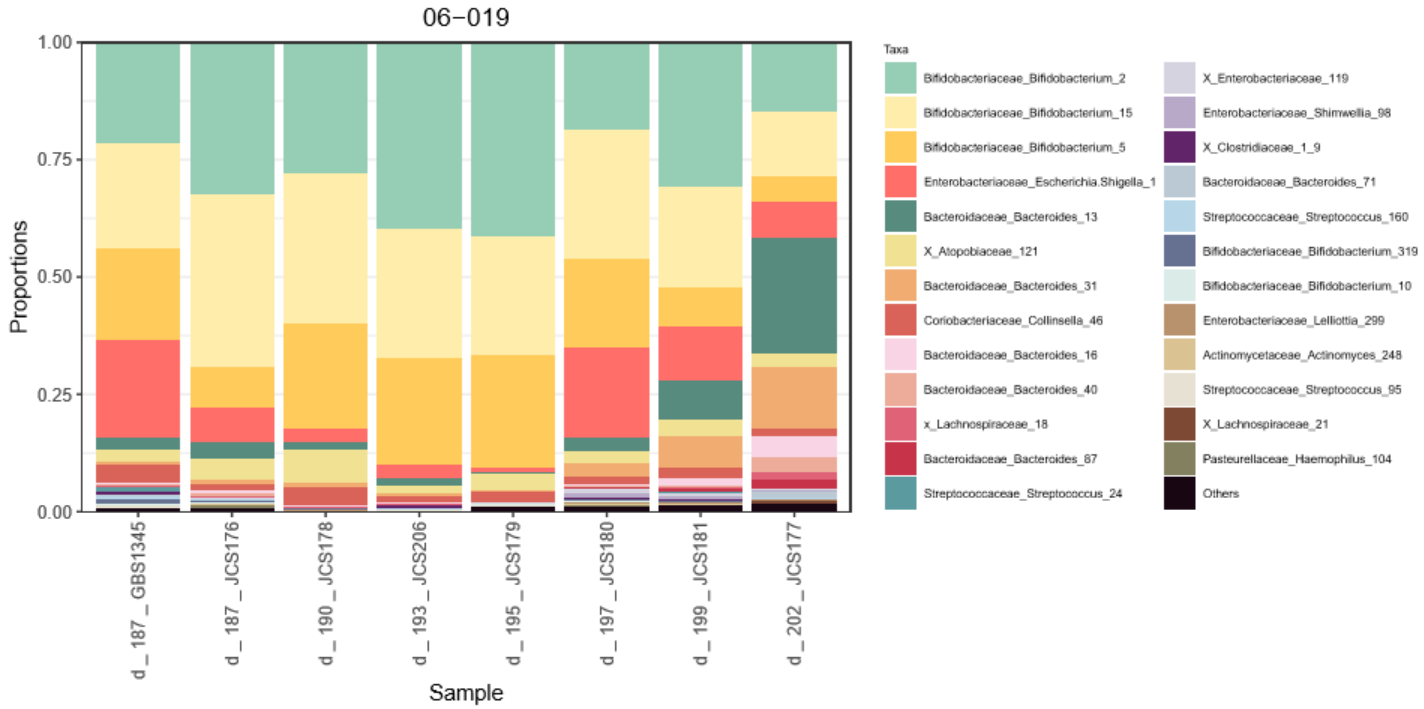


Figure 85: Taxa bar chart of relative abundance for the top 25 bacterial ASVs for 06-019.

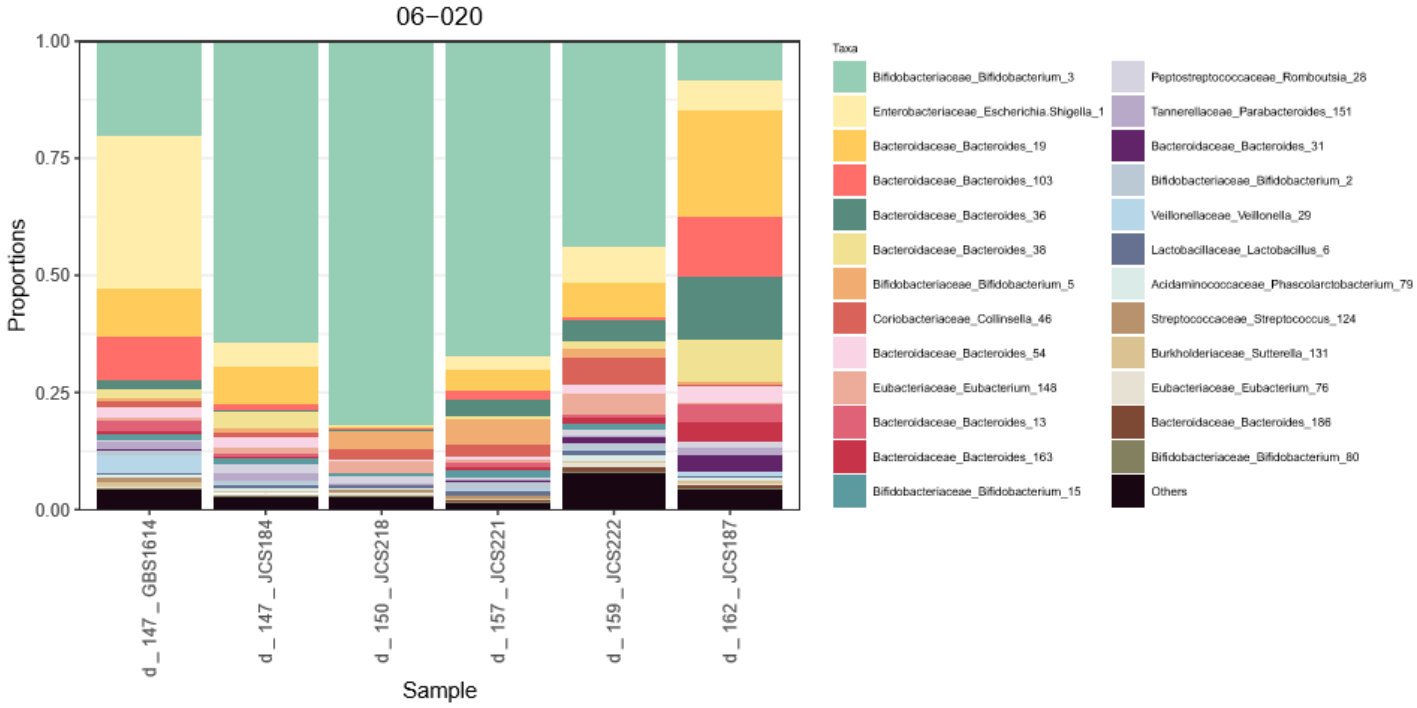


Figure 87: Taxa bar chart of relative abundance for the top 25 bacterial ASVs for 06-020.

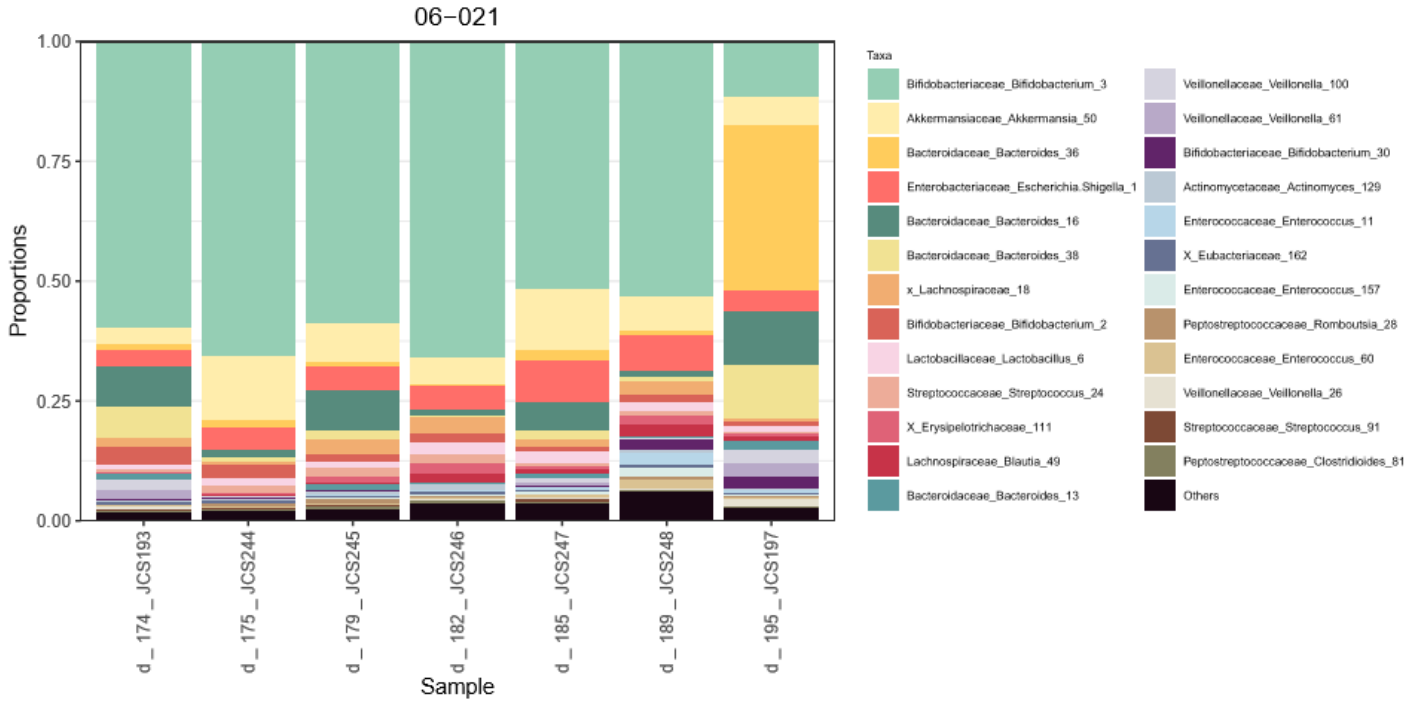


Figure 88: Taxa bar chart of relative abundance for the top 25 bacterial ASVs for 06-021.

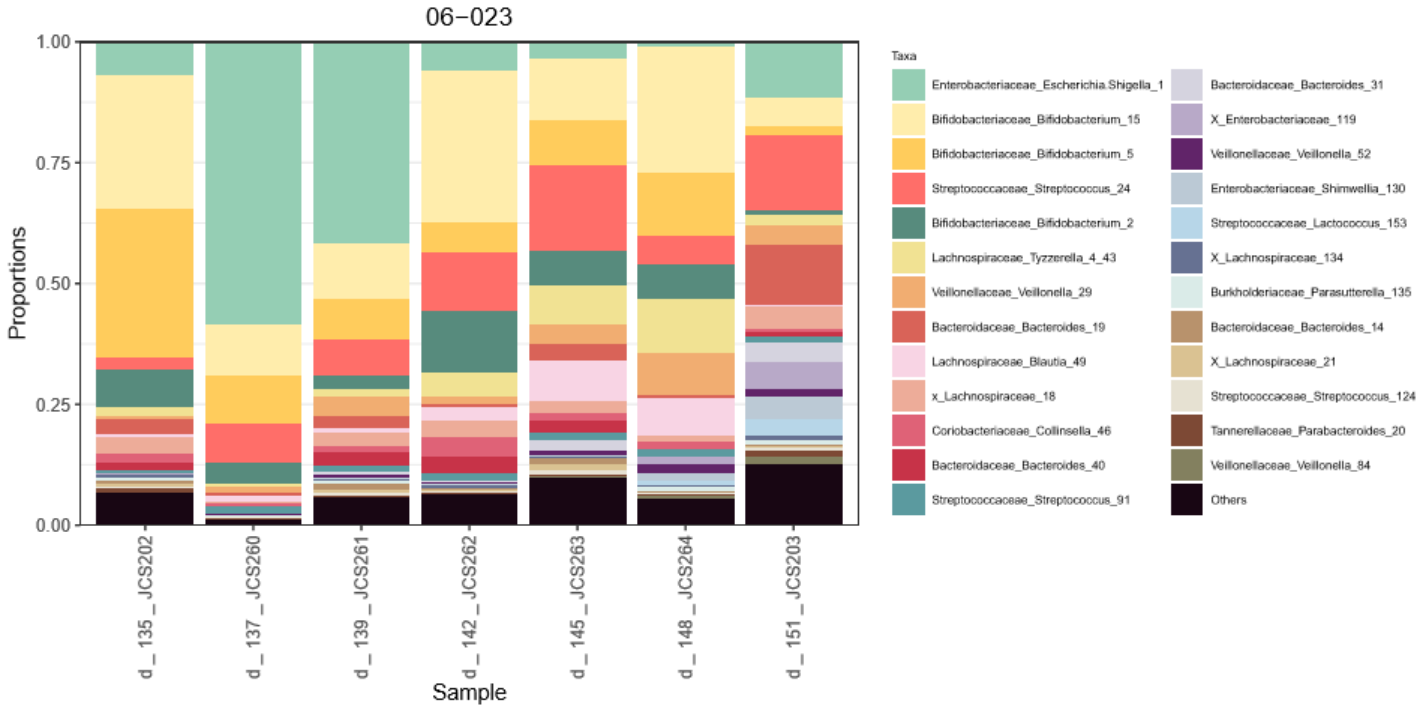


Figure 89: Taxa bar chart of relative abundance for the top 25 bacterial ASVs for 06-023.

Population Level Analysis

Beta Diversity

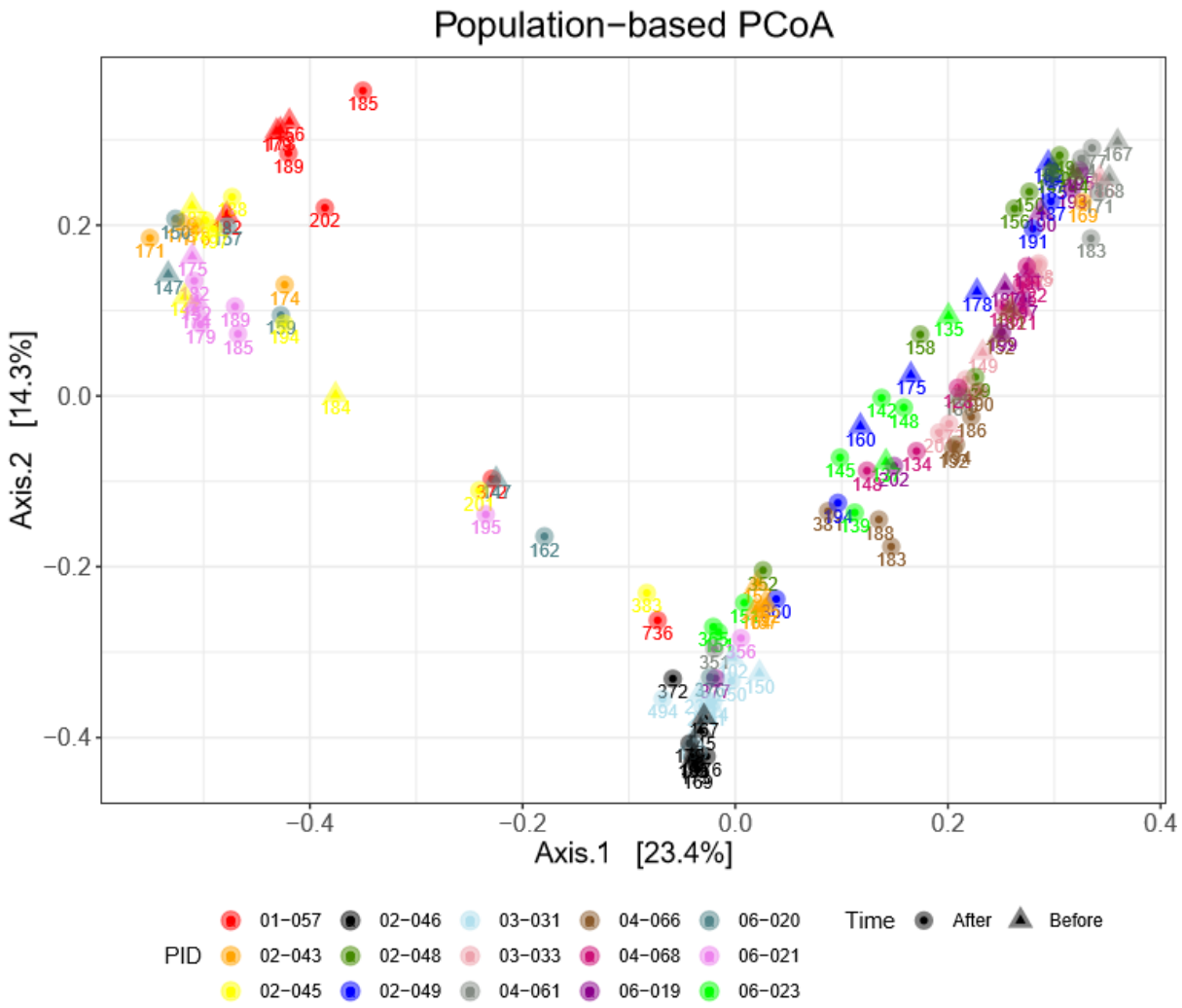


Figure 90: Population level PCoA with samples depicted over 90 days in age. Points are numbered according to age in days.

Covariates

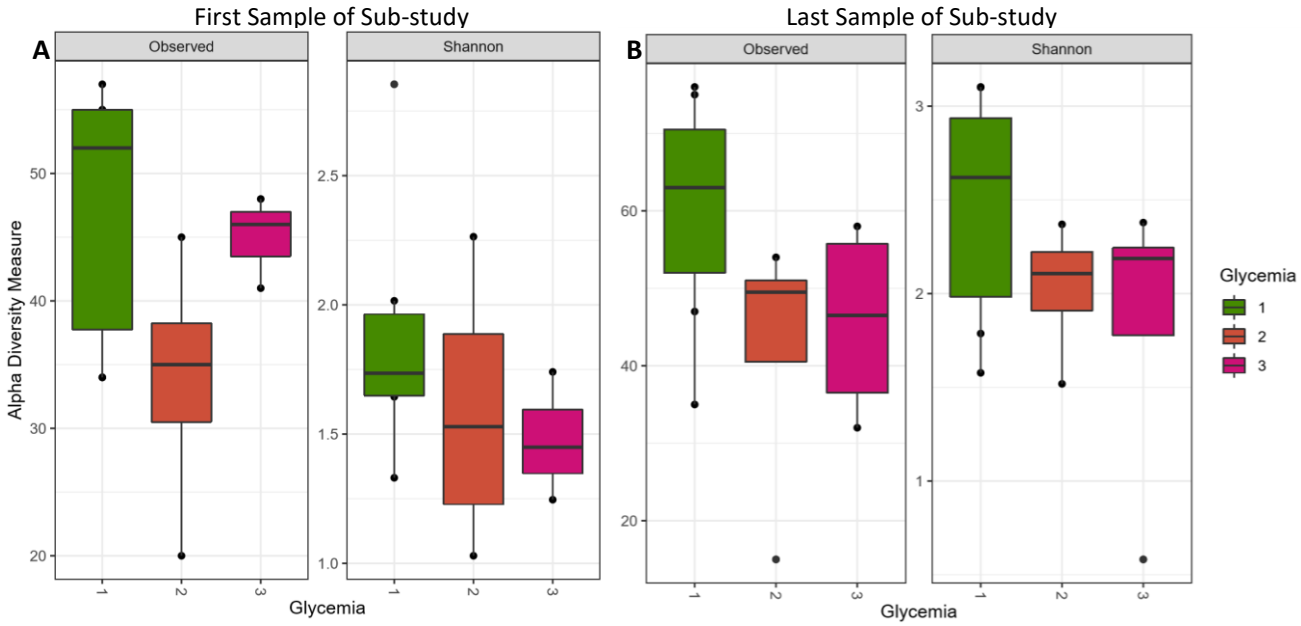


Figure 91: Graph showing the effect hyperglycemia during pregnancy has on alpha diversity (green (1) = non-elevated glucose levels, brown (2) = elevated glucose levels, pink (3) = unknown/not tested). A: first sample available for intensively sampled period. B: last sample available for intensively sampled period.

Appendix C: Nutrition and the Microbiome

Alpha Diversity

Macronutrients

Linear mixed effects analyses were carried out for calories from the macronutrients.

Tables. 15 - 18 shows the results for fat and protein calories.

*Table 15: Results from the linear mixed effects analyses for calories from protein and Shannon alpha diversity. * = p <0.1, ** = p < 0.05, *** = p < 0.01. Adjustments are made for age in days, age at introduction and GBS prophylaxis as these were significant.*

	Unadjusted Model: Estimate (SE)	Standard Multivariate 1: Estimate (SE)	Standard Multivariate 2: Estimate (SE)	Energy Partition Model 1: Estimate (SE)	Energy Partition Model 2: Estimate (SE)
Protein calories	0.011* (0.006)	-0.008 (0.012)	-0.012 (0.012)	-0.005 (0.011)	-0.009 (0.010)
Intercept	1.734*** (0.116)	1.705*** (0.120)	3.236*** (0.888)	1.705*** (0.120)	3.236*** (0.888)
AIC	99.47	109.34	110.96	109.34	110.96

*Table 16: Results from the linear mixed effects analyses for calories from protein and observed species richness. * = p <0.1, ** = p < 0.05, *** = p < 0.01. Adjustments are not made for age in days, age at introduction and GBS prophylaxis as these were not significant in univariate analyses.*

	Unadjusted Model: Estimate (SE)	Standard Multivariate 1: Estimate (SE)	Energy Partition Model 1: Estimate (SE)
Protein calories	0.055 (0.147)	0.052 (0.289)	0.053 (0.256)

Intercept	50.421*** (4.002)	50.406*** (4.029)	50.406*** (4.029)
AIC	495.33	501.96	501.96

*Table 17: Results from the linear mixed effects analyses for calories from fat for Shannon alpha diversity. * = p <0.1, ** = p < 0.05, *** = p < 0.01. Adjustments are made for age in days, age at introduction and GBS prophylaxis as these were significant. Adjustments are made for age in days, age at introduction and GBS prophylaxis as these were significant.*

	Unadjusted Model: Estimate (SE)	Standard Multivariate 1: Estimate (SE)	Standard Multivariate 2: Estimate (SE)	Energy Partition Model 1: Estimate (SE)	Energy Partition Model 2: Estimate (SE)
Fat calories	0.003* (0.001)	-0.003 (0.004)	-0.001 (0.003)	0.001 (0.002)	0.001 (0.002)
Intercept	1.759*** (0.114)	1.668*** (0.123)	3.197*** (0.869)	1.668*** (0.123)	3.197*** (0.869)
AIC	102.30	111.29	114.27	111.29	114.27

*Table 18: Results from the linear mixed effects analyses for calories from fat and observed species richness. * = p <0.1, ** = p < 0.05, *** = p < 0.01. Adjustments are not made for age in days, age at introduction and GBS prophylaxis as these were not significant in univariate analyses.*

	Unadjusted Model: Estimate (SE)	Standard Multivariate 1: Estimate (SE)	Energy Partition Model 1: Estimate (SE)
Fat calories	-0.001 (0.032)	-0.086 (0.087)	-0.028 (0.041)
Intercept	50.83*** (3.969)	49.501*** (4.021)	49.501*** (4.021)
AIC	498.54	503.44	503.44

The tables for Shannon alpha diversity for calories from carbohydrates and fiber are visible in the main text, however *Tables 19 & 20* show the output for linear mixed analyses for observed species richness.

*Table 19: Results from the linear mixed effects analyses for calories from carbohydrates and observed species richness. * = p <0.1, ** = p < 0.05, *** = p < 0.01. Adjustments are not made for age in days, age at introduction and GBS prophylaxis as these were not significant in univariate models.*

	Unadjusted Model: Estimate (SE)	Standard Multivariate 1: Estimate (SE)	Energy Partition Model 1: Estimate (SE)
CHO calories	0.048 (0.055)	0.091 (0.093)	0.072 (0.069)
Intercept	49.577*** (4.050)	49.507*** (4.037)	49.507*** (4.037)
AIC	496.70	503.31	503.31

*Table 20: Results from the linear mixed effects analyses for calories from fiber and observed species richness. * = p <0.1, ** = p < 0.05, *** = p < 0.01. Adjustments are not made for age in days, age at introduction and GBS prophylaxis as these were not significant in univariate models.*

	Unadjusted Model: Estimate (SE)	Standard Multivariate 1: Estimate (SE)	Standard Multivariate 2: Estimate (SE)
Fiber (g/d)	0.380 (0.820)	1.245 (1.492)	1.326 (1.509)
Intercept	50.464*** (3.980)	49.118*** (4.050)	80.022** (36.636)
AIC	485.44	495.42	488.26

Food Categories

"A priori" groups – Alpha Diversity

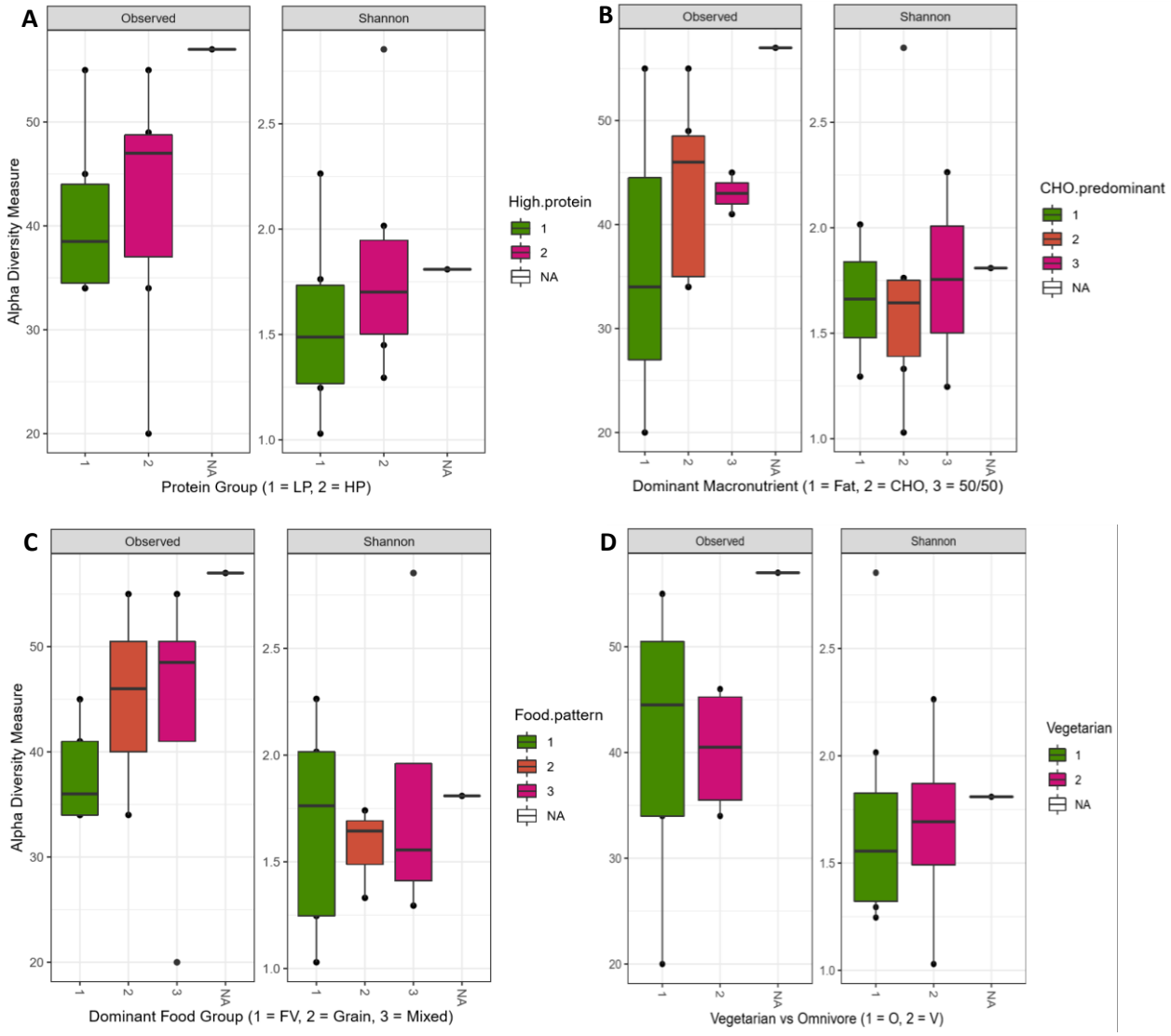


Figure 92: Alpha diversity plots for the different "a priori" groupings with the first sample of the intensively sampled sub-study (before sample). No statistically significant differences can be seen in the before samples between groups. A: Protein groups. Green (1) = low protein, pink (2) = high protein; B: Dominant macronutrient. Green (1) = fat predominant, brown (2) = carbohydrate predominant, pink (3) = mixed fat/carbohydrate; C: Food patterns. Green (1) = fruit and vegetable based, brown (2) = grain-based, pink (3) = mixed diet; D: Vegetarian vs non vegetarian. Green (1) = non-vegetarian, pink (2) = vegetarian.

Linear mixed effects analyses were run for the food categories as well, the output for these analyses are seen below. Observed species richness and dominant macronutrient grouping can be seen in the main text under *4.6.1.2*.

*Table 21: Output from the linear mixed effects analysis of protein grouping and Shannon alpha diversity. * = p <0.1, ** = p < 0.05, *** = p < 0.01. Adjustments are made for age in days, age at introduction and GBS prophylaxis as these were significant.*

	Model 1: Estimate (SE)	Model 2: Estimate (SE)	Model 3: Estimate (SE)	Model 4: Estimate (SE)
High Protein	0.236 (0.219)	0.076 (0.235)	0.119 (0.231)	0.055 (0.200)
Intercept	1.499*** (0.346)	1.591*** (0.360)	2.051** (0.965)	3.095*** (0.936)
AIC	105.04	103.78	107.68	106.20

*Model 1 is unadjusted, Model 2 adjusts for total caloric intake, Model 3 adjusts for total caloric intake, age in days and age at introduction and model 4 adjusts for total caloric intake, age in days, age at introduction and GBS prophylaxis.

*Table 22: Output from the linear mixed effects analysis of protein grouping and observed species richness. * = p <0.1, ** = p < 0.05, *** = p < 0.01. Adjustments are made for age in days, age at introduction and GBS prophylaxis as these were significant.*

	Model 1: Estimate (SE)	Model 2: Estimate (SE)	Model 3: Estimate (SE)	Model 4: Estimate (SE)
High Protein	8.732 (7.708)	8.077 (7.934)	8.350 (8.283)	6.804 (8.251)
Intercept	46.400*** (5.453)	46.612*** (5.521)	52.233 (34.165)	74.228* (38.343)
AIC	513.88	494.32	493.30	487.47

*Model 1 is unadjusted, Model 2 adjusts for total caloric intake, Model 3 adjusts for total caloric intake, age in days and age at introduction and model 4 adjusts for total caloric intake, age in days, age at introduction and GBS prophylaxis.

*Table 23: Output from the linear mixed effects analysis of dominant macronutrient grouping and Shannon alpha diversity. * = p <0.1, ** = p < 0.05, *** = p < 0.01. Adjustments are made for age in days, age at introduction and GBS prophylaxis as these were significant.*

	Model 1: Estimate (SE)	Model 2: Estimate (SE)	Model 3: Estimate (SE)	Model 4: Estimate (SE)
CHO predominant	0.374 (0.274)	0.349 (0.277)	0.333 (0.276)	0.220 (0.247)
50/50 Diet	0.076 (0.330)	0.204 (0.340)	0.169 (0.338)	0.168 (0.296)
Intercept	1.623*** (0.234)	1.458*** (0.245)	1.824* (0.964)	2.864*** (0.985)
AIC	106.18	104.67	108.80	108.07

*Model 1 is unadjusted, Model 2 adjusts for total caloric intake, Model 3 adjusts for total caloric intake, age in days and age at introduction and model 4 adjusts for total caloric intake, age in days, age at introduction and GBS prophylaxis.

*Table 24: Output from the linear mixed effects analysis of dominant food group grouping and Shannon alpha diversity. * = p <0.1, ** = p < 0.05, *** = p < 0.01. Adjustments are made for age in days, age at introduction and GBS prophylaxis as these were significant.*

	Model 1: Estimate (SE)	Model 2: Estimate (SE)	Model 3: Estimate (SE)	Model 4: Estimate (SE)
Grain based	0.100 (0.295)	0.078 (0.305)	-0.047 (0.376)	0.011 (0.315)
Mixed	0.313 (0.254)	0.169 (0.269)	0.137 (0.293)	0.151 (0.248)
Intercept	1.720*** (0.172)	1.629*** (0.180)	2.155* (1.302)	3.012*** (1.158)
AIC	107.14	106.01	109.60	108.23

*Model 1 is unadjusted, Model 2 adjusts for total caloric intake, Model 3 adjusts for total caloric intake, age in days and age at introduction and model 4 adjusts for total caloric intake, age in days, age at introduction and GBS prophylaxis

*Table 25: Output from the linear mixed effects analysis of dominant food group grouping and observed species richness. * = p <0.1, ** = p < 0.05, *** = p < 0.01. Adjustments are made for age in days, age at introduction and GBS prophylaxis as these were significant.*

	Model 1: Estimate (SE)	Model 2: Estimate (SE)	Model 3: Estimate (SE)	Model 4: Estimate (SE)
Grain based	6.387 (10.664)	5.990 (10.751)	9.108 (13.831)	10.177 (13.410)
Mixed	9.321 (9.156)	8.512 (9.331)	10.330 (10.615)	10.261 (10.282)
Intercept	46.063*** (6.179)	46.276*** (6.264)	27.337 (47.894)	49.850 (49.428)
AIC	509.36	489.80	488.14	481.98

*Model 1 is unadjusted, Model 2 adjusts for total caloric intake, Model 3 adjusts for total caloric intake, age in days and age at introduction and model 4 adjusts for total caloric intake, age in days, age at introduction and GBS prophylaxis

*Table 26: Output from the linear mixed effects analysis of a vegetarian diet and Shannon alpha diversity. * = p <0.1, ** = p < 0.05, *** = p < 0.01. Adjustments are made for age in days, age at introduction and GBS prophylaxis as these were significant.*

	Model 1: Estimate (SE)	Model 2: Estimate (SE)	Model 3: Estimate (SE)	Model 4: Estimate (SE)
Vegetarian	0.056 (0.239)	0.102 (0.238)	0.095 (0.235)	-0.121 (0.216)
Intercept	1.833*** (0.142)	1.664*** (0.149)	2.149** (0.940)	3.331*** (0.924)
AIC	106.00	103.67	107.75	105.79

*Model 1 is unadjusted, Model 2 adjusts for total caloric intake, Model 3 adjusts for total caloric intake, age in days and age at introduction and model 4 adjusts for total caloric intake, age in days, age at introduction and GBS prophylaxis

*Table 27: Output from the linear mixed effects analysis of a vegetarian diet and observed species richness. * = p <0.1, ** = p < 0.05, *** = p < 0.01. Adjustments are made for age in days, age at introduction and GBS prophylaxis as these were significant.*

	Model 1: Estimate (SE)	Model 2: Estimate (SE)	Model 3: Estimate (SE)	Model 4: Estimate (SE)
Vegetarian	-5.897 (8.301)	-5.194 (8.363)	-5.222 (8.729)	-11.621 (8.716)
Intercept	52.861*** (4.943)	52.337*** (5.108)	55.345 (35.198)	91.932** (37.851)
AIC	514.52	494.88	493.87	486.28

*Model 1 is unadjusted, Model 2 adjusts for total caloric intake, Model 3 adjusts for total caloric intake, age in days and age at introduction and model 4 adjusts for total caloric intake, age in days, age at introduction and GBS prophylaxis

Dietary Diversity

Linear mixed effects analyses were also performed for the dietary diversity scores. Tables below show the results of the analyses for scores one, two and three. The fourth table is visible in the main text.

*Table 28: Results of the linear mixed effects analyses for the relationship between the first dietary diversity score and Shannon/observed species richness. * = p <0.1, ** = p < 0.05, *** = p < 0.01. Adjustments are made for age in days, age at introduction and GBS prophylaxis as these were significant.*

Independent Variables	Shannon Alpha Diversity Unadjusted: Estimate (SE)	Shannon Alpha Diversity Standard Multivariate: Estimate (SE)	Observed species richness Unadjusted: Estimate (SE)	Observed species richness Standard Multivariate: Estimate (SE)
Diversity Score 1	0.014 (0.023)	0.025 (0.019)	1.105 (0.771)	1.432* (0.755)
Intercept	1.764*** (0.187)	3.314*** (0.838)	43.742*** (6.174)	85.496** (33.856)
AIC	110.33	109.29	517.75	489.62

*Table 29: Results of the linear mixed effects analyses for the relationship between the second dietary diversity score and Shannon/observed species richness. * = p <0.1, ** = p < 0.05, *** = p < 0.01. Adjustments are made for age in days, age at introduction and GBS prophylaxis as these were significant.*

Independent Variables	Shannon Alpha Diversity Unadjusted: Estimate (SE)	Shannon Alpha Diversity Standard Multivariate: Estimate (SE)	Observed species richness Unadjusted: Estimate (SE)	Observed species richness Standard Multivariate: Estimate (SE)

Diversity Score 2	-0.019 (0.082)	0.036 (0.073)	2.308 (2.820)	4.064 (2.898)
Intercept	1.922*** (0.319)	3.194*** (0.889)	42.371*** (10.998)	79.704** (35.893)
AIC	108.15	108.04	516.61	488.31

*Table 30: Results of the linear mixed effects analyses for the relationship between the third dietary diversity score and Shannon/observed species richness. * = p <0.1, ** = p < 0.05, *** = p < 0.01. Adjustments are made for age in days, age at introduction and GBS prophylaxis, as these were significant.*

Independent Variables	Shannon Alpha Diversity Unadjusted: Estimate (SE)	Shannon Alpha Diversity Standard Multivariate: Estimate (SE)	Observed species richness Unadjusted: Estimate (SE)	Observed species richness Standard Multivariate: Estimate (SE)
Diversity Score 3	0.014 (0.014)	0.017 (0.012)	0.806* (0.496)	0.964** (0.458)
Intercept	1.670*** (0.215)	3.270*** (0.812)	40.440*** (7.030)	82.749** (32.658)
AIC	110.67	109.76	517.95	489.98

Beta Diversity

Macronutrients

The PCoA plots for the calories from the macronutrients and fiber (g/d) are shown in *Fig. 93.*

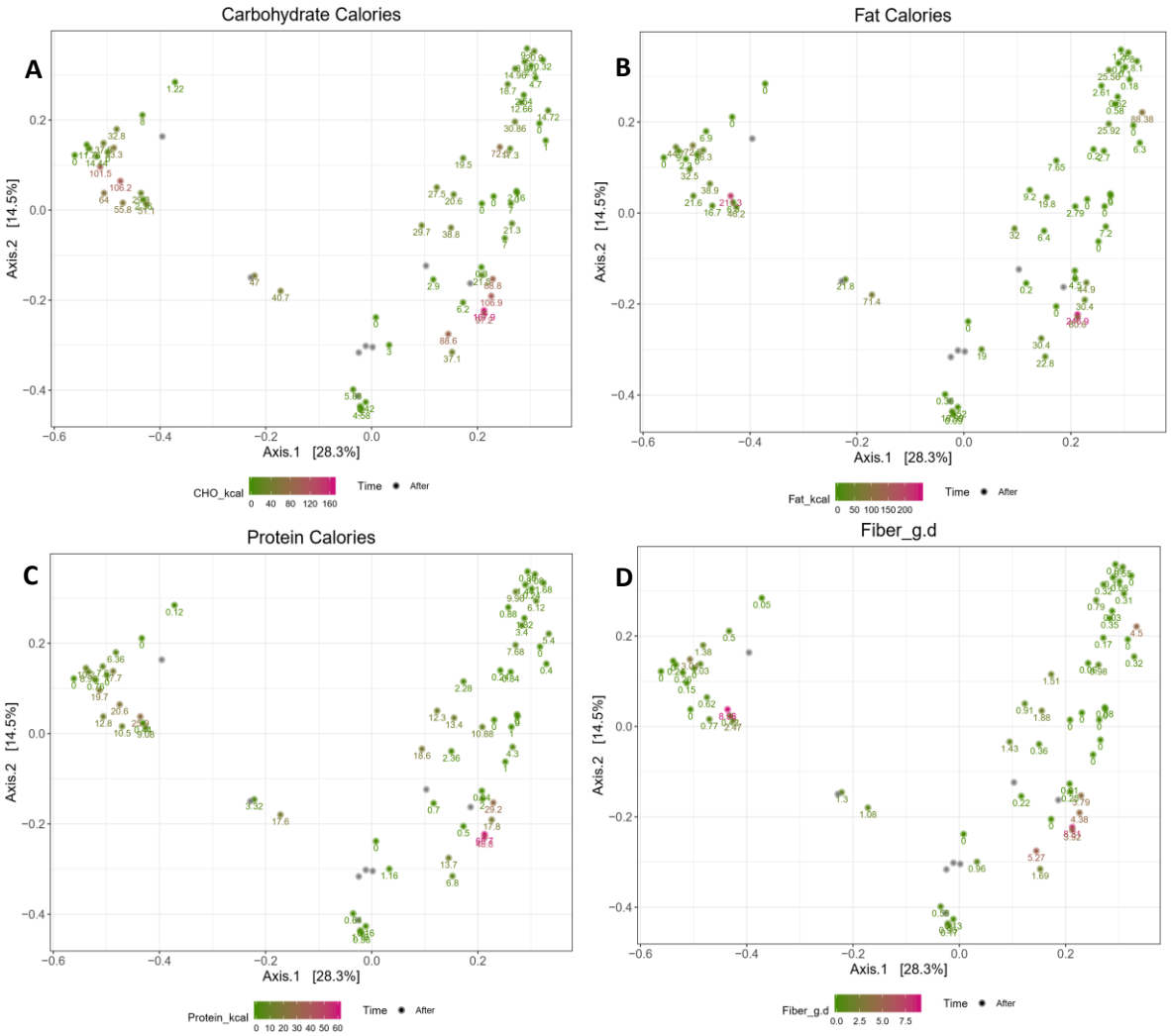


Figure 93: PCoA plots for all samples of the sub-study period for the calories from the macronutrients and fiber (g/d). A: carbohydrate calories, B: fat calories, C: protein calories, D: fiber (g/d).

Food Categories

PCoA plots were created for the food categories for the last sample only, as well as all sub- study samples. *Fig. 94* shows all the PCoA plots with the last sample only.

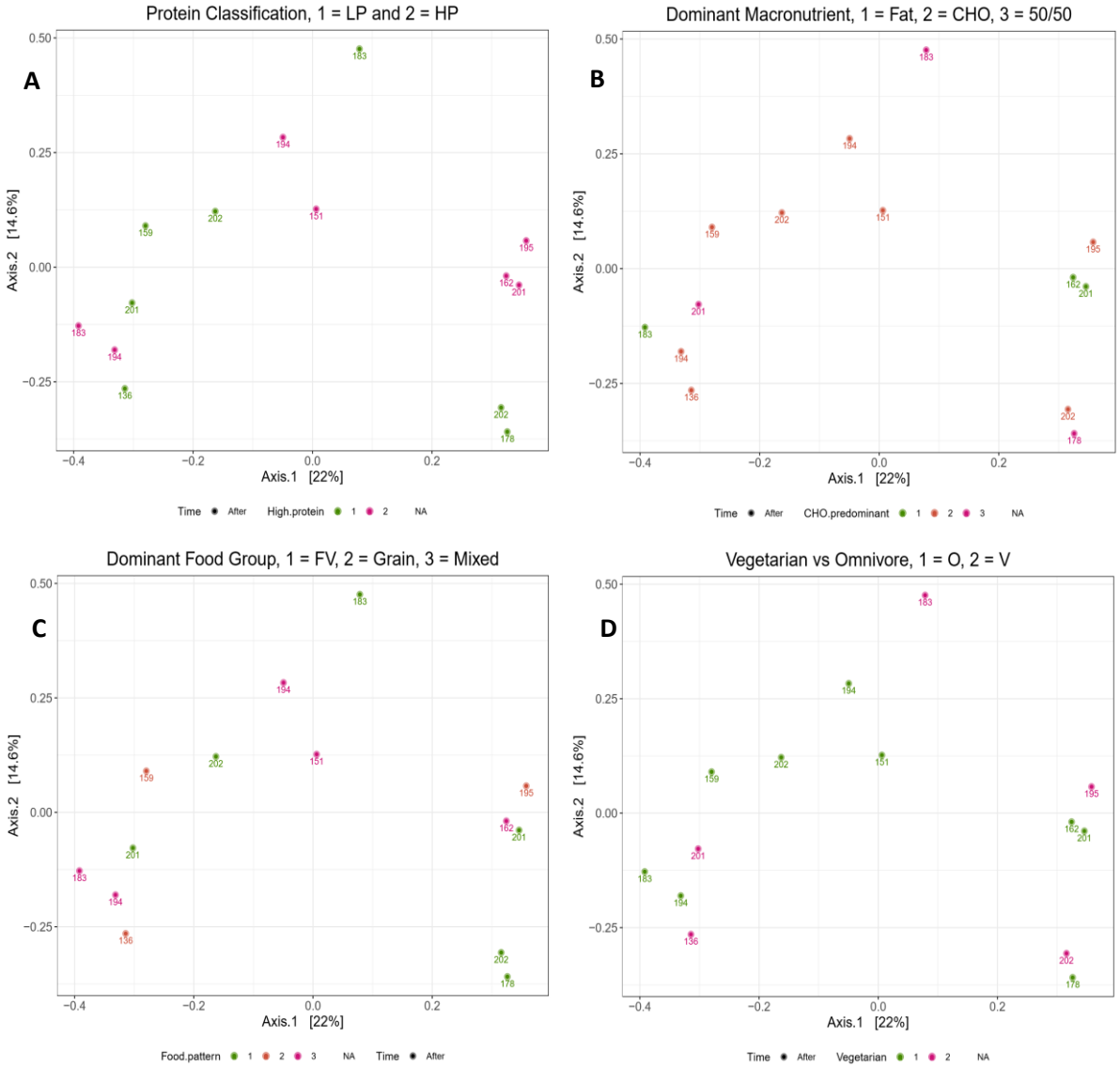


Figure 94: PCoA plots for the last sample of the sub-study period for the food categories. A: Protein group; green (1) = low protein, pink (2) = high protein, B: dominant macronutrient grouping; green (1) = fat predominant, brown (2) = carbohydrate predominant, pink (3) = mixed fat/carbohydrate, C: dominant food group grouping; green (1) = fruit and vegetable based, brown (2) = grain-based, pink (3) = mixed diet, D: vegetarian diet; green (1) = non-vegetarian, pink (2) = vegetarian.

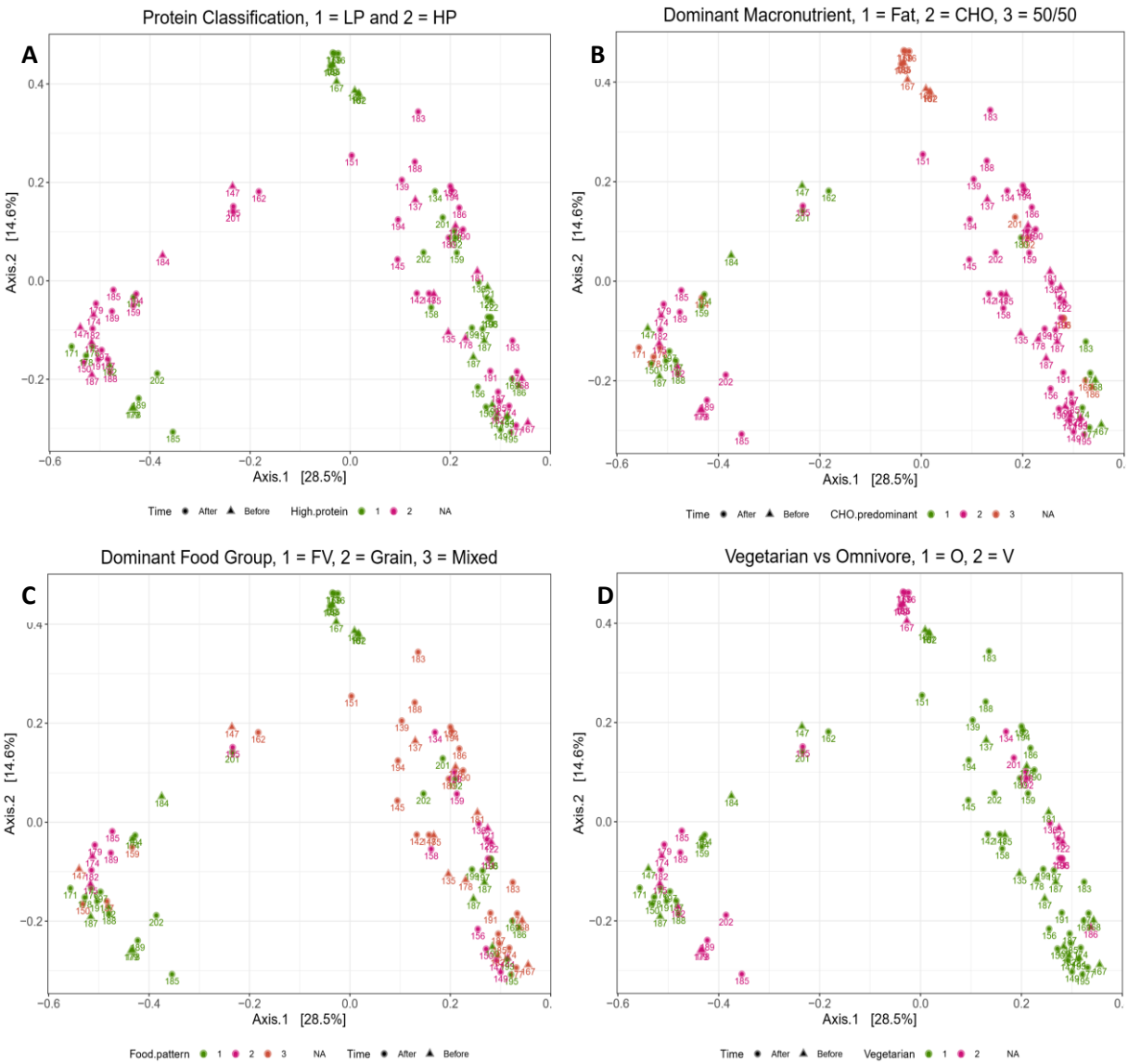


Figure 95: PCoA plots for all samples of the sub-study period for the food categories. A: Protein group; green (1) = low protein, pink (2) = high protein, B: dominant macronutrient grouping; green (1) = fat predominant, brown (2) = carbohydrate predominant, pink (3) = mixed fat/carbohydrate, C: dominant food group grouping; green (1) = fruit and vegetable based, brown (2) = grain-based, pink (3) = mixed diet, D: vegetarian diet; green (1) = non-vegetarian, pink (2) = vegetarian.

Dietary Diversity

Dietary Diversity Scores were linked to degree of change using a simple linear regression.

Table 31: Linear regression output for the relationship between dietary diversity scores and the degree of change during the sub-study period.

Diversity Scores	**Intercept**	**Beta Coefficient**	**p-value**	**R-squared value**
1	-1.855 (2.584)	3.96 (1.638)	**0.0326**	0.27
2	1.374 (1.0505)	1.092 (0.480)	**0.0422**	0.24
3	1.230 (6.018)	5.586 (2.751)	0.065	0.19
4	-1.702 (3.825)	4.578 (1.749)	**0.0225**	0.31

Taxonomic Distribution

Macronutrients

Table 32: Output from the negative binomial regressions for the different models for calories from Carbohydrates and Bacteroides_22

	Bacteroidaceae_Bacteroides_22									
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Energy Partition 1		Energy Partition 2	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-12 (0.824)	<0.001	-11 (0.65)	<0.001	-16 (3.30)	<0.001	-11 (0.65)	<0.001	-16 (3.30)	<0.001
CHO_kcal	0.0069 (0.0175)	0.692	0.078 (0.0440)	0.078	0.052 (0.0404)	0.197	0.040 (0.0276)	0.153	0.023 (0.0258)	0.368
Total_kcal			-0.038 (0.0190)	0.045	-0.029 (0.0174)	0.096				
Age_days					0.030 (0.0196)	0.132			0.029 (0.0196)	0.132
PF_kcal							-0.038 (0.0190)	0.045	-0.029 (0.0174)	0.096
ICC	0.36									
Observations	75		75		75		75		75	
AIC	150.871		149.845		149.934		149.845		149.934	

Table 33: Output from the negative binomial regressions for the different models for calories from Fat and Bacteroides_22

	Bacteroidaceae_Bacteroides_22									
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Energy Partition 1		Energy Partition 2	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-11 (0.842)	<0.001	-11 (0.653)	<0.001	-16 (3.20)	<0.001	-11 (0.653)	<0.001	-16 (3.20)	<0.001
Fat_kcal	-0.023 (0.0203)	0.251	-0.12 (0.0631)	0.055	-0.094 (0.0591)	0.110	-0.080 (0.0398)	0.045	-0.064 (0.0371)	0.083
Total_kcal			0.041 (0.0261)	0.113	0.030 (0.0248)	0.223				
Age_days					0.029 (0.0186)	0.119			0.029 (0.0186)	0.119

PC_kcal							0.041 (0.0261)	0.113	0.030 (0.0248)	0.224
ICC	0.26									
Observations	75		75		75		75		75	
AIC	149.443		148.077		148.077		148.077		148.077	

Table 34: Output from the negative binomial regressions for the different models for calories from fat and Enterococcus_7

	Enterococcaceae_Enterococcus_7									
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Energy Partition 1		Energy Partition 2	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-6.4 (0.695)	<0.001	-7.0 (0.763)	<0.001	0.42 (5.07)	0.933	-7.0 (0.763)	<0.001	0.42 (5.07)	0.933
Fat_kcal	-0.000088 (0.00606)	0.988	-0.040 (0.0182)	0.029	-0.042 (0.0187)	0.026	-0.015 (0.00821)	0.073	-0.015 (0.00857)	0.081
Total_kcal			0.025 (0.0112)	0.025	0.027 (0.0113)	0.019				
Age_days					-0.042 (0.029)	0.142			-0.042 (0.0289)	0.142
PC_kcal							0.025 (0.0112)	0.025	0.027 (0.0113)	0.019
ICC	0.86		0.88		0.82		0.88		0.82	
Observations	75		75		75		75		75	
AIC	977.712		975.016		975.318		975.016		975.318	

Table 35: Output from the negative binomial regressions for the different models for Fiber (g/d) and Bacteroides_13

	Bacteroidaceae_Bacteroides_13							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-7.1 (1.03)	<0.001	-7.2 (1.05)	<0.001	-20 (2.98)	<0.001	-23 (6.40)	<0.001
Fiber_g.d	0.091 (0.271)	0.738	0.013 (0.341)	0.970	0.52 (0.256)	0.042	0.028 (0.339)	0.933
Total_kcal			0.0035 (0.0102)	0.734			0.0037 (0.0114)	0.749
Age_days					0.083 (0.0176)	<0.001	0.091 (0.0364)	0.012
ICC	0.85		0.84				0.76	
Observations	75		75		75		75	

AIC	709.718	711.588	708.770	707.873

Table 36: Output from the negative binomial regressions for the different models for Fiber (g/d) and Enterococcus_7

	Enterococcaceae_Enterococcus_7							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-6.2 (0.705)	<0.001	-6.4 (0.681)	<0.001	-1.1 (5.30)	0.837	1.04 (4.53)	0.818
Fiber_g.d	-0.21 (0.164)	0.199	-0.52 (0.201)	0.009	-0.19 (0.169)	0.254	-0.56 (0.206)	0.006
Total_kcal			0.0092 (0.00477)	0.053			0.011 (0.00503)	0.029
Age_days					-0.029 (0.0300)	0.332	-0.043 (0.0257)	0.097
ICC	0.86		0.85		0.82		0.77	
Observations	75		75		75		75	
AIC	976.276		974.159		977.501		974.143	

Food Categories

Table 37: Output from the negative binomial regressions for the different models for the dominant macronutrient and Bifidobacterium_5

	Bifidobacteriaceae_Bifidobacterium_5							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-5.0 (1.13)	<0.001	-4.8 (1.28)	<0.001	-3.5 (1.51)	0.002	5.9 (5.31)	0.266
CHO.predominant [2]	2.6 (1.33)	0.053	2.4 (1.50)	0.111	-0.29 (1.54)	0.026	2.2 (1.67)	0.193
CHO.predominant [3]	0.61 (1.61)	0.705	0.62 (1.82)	0.734	-0.76 (1.85)	0.650	1.1 (2.03)	0.595
Total_kcal			-0.0010 (0.00241)	0.665			0.00073 (0.00257)	0.776
Age_days					0.0017 (0.00439)	0.002	-0.062 (0.0295)	0.036
ICC	0.78		0.84		0.76		0.88	
Observations	95		75		95		75	
AIC	1623.039		1301.501		1617.693		1298.382	

Table 38: Output from the negative binomial regressions for the different models for the dominant macronutrient and Bifidobacterium_15

	Bifidobacteriaceae_Bifidobacterium_15							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-5.7 (0.702)	<0.001	-11 (3.46)	0.002	-16 (3.31)	<0.001	-6.1 (12)	0.605
CHO.predominant [2]	2.8 (0.885)	0.002	0.12 (3.99)	0.976	0.37 (2.75)	0.893	0.082 (4.11)	0.984
CHO.predominant [3]	0.43 (1.15)	0.710	4.1 (4.62)	0.370	1.8 (3.23)	0.587	4.5 (4.83)	0.352
Total_kcal			0.0060 (0.00774)	0.437			0.0073 (0.00868)	0.397
Age_days					0.039 (0.0110)	<0.001	-0.028 (0.0678)	0.676
ICC			0.94		0.86		0.94	
Observations	95		75		95		75	
AIC	856.319		602.572		845.181		604.393	

Table 39: Output from the negative binomial regressions for the different models for the dominant macronutrient and _Lachnospiraceae_18

	x_Lachnospiraceae_18							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-5.5 (0.934)	<0.001	-8.9 (1.39)	<0.001	-13 (1.43)	<0.001	-8.5 (5.09)	0.095
CHO.predominant [2]	0.92 (1.13)	0.416	1.8 (1.55)	0.256	1.8 (1.32)	0.171	1.7 (1.56)	0.263
CHO.predominant [3]	2.6 (1.38)	0.064	5.7 (1.90)	0.003	5.3 (1.61)	0.001	5.7 (1.91)	0.003
Total_kcal			-0.0013 (0.00621)	0.835			-0.0012 (0.00623)	0.842
Age_days					0.026 (0.00386)	<0.001	-0.0026 (0.0281)	0.927
ICC	0.47		0.78		0.71		0.78	
Observations	95		75		95		75	

AIC	1140.027	755.352	1088.924	757.344

Table 40: Output from the negative binomial regressions for the different models for the dominant macronutrient and Lactobacillus_6

	Lactobacillaceae_Lactobacillus_6							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-6.9 (1.36)	<0.001	-6.8 (1.41)	<0.001	-4.1 (1.46)	0.005	-5.0 (5.27)	0.340
CHO.predominant [2]	0.68 (1.60)	0.671	0.70 (1.63)	0.670	0.84 (1.58)	0.595	0.68 (1.66)	0.684
CHO.predominant [3]	3.2 (1.93)	0.102	3.2 (1.99)	0.110	3.2 (1.91)	0.094	3.3 (2.03)	0.109
Total_kcal			0.00050 (0.00312)	0.872			0.00080 (0.00324)	0.805
Age_days					-0.016 (0.00300)	<0.001	-0.011 (0.0294)	0.720
ICC	0.81		0.85		0.83		0.85	
Observations	95		75		95		75	
AIC	1144.719		927.993		1129.316		929.862	

Table 41: Output from the negative binomial regressions for the different models for the dominant Food Pattern and Bifidobacterium_10

	Bifidobacteriaceae_Bifidobacterium_10							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-8.0 (1.39)	<0.001	-10 (1.75)	<0.001	-13 (2.06)	<0.001	-19 (9.96)	0.060
Food.pattern [2]	2.8 (2.26)	0.217	4.3 (2.79)	0.123	3.6 (2.44)	0.135	5.7 (3.29)	0.083
Food.pattern [3]	0.73 (1.95)	0.706	1.8 (2.44)	0.460	1.1 (2.09)	0.605	2.4 (2.58)	0.347
Total_kcal			-0.0011 (0.00413)	0.785			-0.0023 (0.00427)	0.595
Age_days					0.018 (0.00625)	0.003	0.045 (0.0525)	0.394

ICC	0.80	0.90	0.85	0.90
Observations	95	75	95	75
AIC	751.904	563.215	742.198	564.474

Table 42: Output from the negative binomial regressions for the different models for the dominant Food Pattern and Bifidobacterium_15

	Bifidobacteriaceae_Bifidobacterium_15							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-5.9 (1.57)	<0.001	-7.2 (2.21)	0.001	-14 (2.72)	<0.001	4.6 (13)	0.725
Food.pattern [2]	-0.073 (2.49)	0.977	-7.7 (4.62)	0.095	-3.1 (2.81)	0.268	-9.0 (4.80)	0.061
Food.pattern [3]	1.5 (2.10)	0.476	-2.6 (3.23)	0.419	0.34 (2.31)	0.884	-4.0 (3.61)	0.273
Total_kcal			0.0062 (0.00775)	0.426			0.0097 (0.00948)	0.307
Age_days					0.038 (0.0106)	<0.001	-0.064 (0.0701)	0.359
ICC	0.75		0.92		0.83		0.92	
Observations	95		75		95		75	
AIC	859.072		600.540		843.916		601.687	

Table 43: Output from the negative binomial regressions for the different models for the dominant Food Pattern and Bacteroides_22

	Bacteroidaceae_Bacteroides_22							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-10 (0.663)	<0.001	-9.9 (0.763)	<0.001	-7.7 (2.67)	0.004	-16 (5.33)	0.003
Food.pattern [2]	-1.7 (1.11)	0.120	-2.2 (1.15)	0.058	-1.2 (2.16)	0.580	-0.75 (1.76)	0.671
Food.pattern [3]	-1.7 (0.923)	0.062	-1.7 (0.973)	0.083	-1.2 (1.84)	0.529	-0.87 (1.27)	0.494
Total_kcal			-0.0029 (0.00750)	0.702			-0.0061 (0.00839)	0.464

Age_days					-0.018 (0.0132)	0.173	0.032 (0.0280)	0.256
ICC					0.30			
Observations	95		75		95		75	
AIC	167.586		153.034		167.972		153.756	

Table 44: Output from the negative binomial regressions for the different models for the dominant Food Pattern and Enterococcus_7

	Enterococcaceae_Enterococcus_7							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-8.1 (0.891)	<0.001	-8.1 (0.876)	<0.001	-2.5 (1.23)	0.047	-6.1 (6.26)	0.329
Food.pattern [2]	3.1 (1.49)	0.035	3.2 (1.47)	0.032	1.8 (1.46)	0.208	2.8 (1.82)	0.123
Food.pattern [3]	2.6 (1.28)	0.042	2.6 (1.29)	0.042	1.8 (1.26)	0.157	2.4 (1.40)	0.082
Total_kcal			0.00093 (0.00392)	0.812			0.0012 (0.00410)	0.762
Age_days					-0.028 (0.00401)	<0.001	-0.010 (0.0329)	0.755
ICC	0.73		0.80		0.77		0.78	
Observations	95		75		95		75	
AIC	1122.312		976.024		1091.293		977.931	

Table 45: Output from the negative binomial regressions for the different models for a vegetarian diet and Bifidobacterium_2

	Bifidobacteriaceae_Bifidobacterium_2							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-2.7 (0.648)	<0.001	-2.8 (0.958)	0.003	-3.1 (1.03)	0.003	-0.83 (6.62)	0.900
Vegetarian [2]	-2.5 (1.11)	0.023	-3.4 (1.61)	0.034	-2.6 (1.12)	0.020	-3.5 (1.67)	0.038

Total_kcal			-0.00017 (0.00448)	0.969			0.00016 (0.00457)	0.972
Age_days					0.0022 (0.00411)	0.598	-0.012 (0.0382)	0.763
ICC	0.69		0.82		0.69		0.83	
Observations	95		75		95		75	
AIC	1486.840		1137.442		1488.547		1139.347	

Table 46: Output from the negative binomial regressions for the different models for a vegetarian diet and Bifidobacterium_5

	Bifidobacteriaceae_Bifidobacterium_5							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-3.1 (0.746)	<0.001	-2.9 (0.817)	<0.001	-3.1 (1.03)	0.053	8.9 (5.17)	0.086
Vegetarian [2]	-0.80 (1.25)	0.526	-1.1 (1.36)	0.423	-2.6 (1.12)	0.639	-1.0 (1.48)	0.489
Total_kcal			-0.00095 (0.00243)	0.696			0.00089 (0.00256)	0.729
Age_days					0.0022 (0.00411)	0.004	-0.068 (0.0295)	0.020
ICC	0.82		0.86		0.83		0.89	
Observations	95		75		95		75	
AIC	1624.696		1301.677		1620.568		1297.522	

Table 47: Output from the negative binomial regressions for the different models for a vegetarian diet and Escherichia/Shigella_1

	Enterobacteriaceae_Escherichia/Shigella_1							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-2.4 (0.551)	<0.001	-2.3 (0.937)	0.012	-1.9 (0.816)	0.017	-0.47 (5.33)	0.929
Vegetarian [2]	-1.5 (0.926)	0.102	-3.0 (1.60)	0.059	-1.4 (0.887)	0.116	-3.0 (1.55)	0.053

Total_kcal			-0.0019 (0.00257)	0.468			-0.0016 (0.00269)	0.552
Age_days					-0.0023 (0.00318)	0.466	-0.011 (0.0302)	0.721
Observations	95		75		95		75	
AIC	1793.0		1458.2		1794.5		1460.0	

Dietary Diversity

Table 48: Output from the negative binomial regressions for the different models for dietary diversity score 1 and Bifidobacterium_2

	Bifidobacteriaceae_Bifidobacterium_2							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-5.6 (0.733)	<0.001	-6.7 (1.22)	<0.001	-6.1 (1.14)	<0.001	-3.4 (5.65)	0.552
Diversity	0.34 (0.0923)	<0.001	0.44 (0.138)	0.002	0.34 (0.0932)	<0.001	0.46 (0.149)	0.002
Total_kcal			0.00098 (0.00451)	0.829			0.0015 (0.00449)	0.735
Age_days					0.0020 (0.00391)	0.614	-0.021 (0.0341)	0.539
ICC	0.57		0.75		0.58		0.77	
Observations	95		75		95		75	
AIC	1481.480		1133.088		1483.212		1134.686	

Table 49: Output from the negative binomial regressions for the different models for dietary diversity score 1 and Bifidobacterium_5

	Bifidobacteriaceae_Bifidobacterium_5							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-4.6 (0.908)	<0.001	-4.7 (0.995)	<0.001	-3.1 (1.04)	0.003	7.0 (4.67)	0.134
Diversity	0.19 (0.114)	0.088	0.22 (0.123)	0.073	0.19 (0.112)	0.097	0.28 (0.130)	0.030
Total_kcal			-0.00074 (0.00241)	0.760			0.0011 (0.00252)	0.660
Age_days					-0.0074 (0.00259)	0.004	-0.070 (0.0272)	0.010
ICC	0.80		0.84		0.80		0.86	
Observations	95		75		95		75	
AIC	1622.451		1299.392		1618.279		1293.836	

Table 50: Output from the negative binomial regressions for the different models for dietary diversity score 1 and Bifidobacterium_15

	Bifidobacteriaceae_Bifidobacterium_15							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-8.2 (1.52)	<0.001	-11 (2.71)	<0.001	-8.1 (1.32)	<0.001	-7.8 (11)	0.478
Diversity	0.43 (0.173)	0.013	0.24 (0.320)	0.461	0.34 (0.0933)	<0.001	0.25 (0.330)	0.443

Total_kcal			0.0051 (0.00744)	0.494			0.0060 (0.00816)	0.466
Age_days					0.0080 (0.00479)	0.096	-0.020 (0.0655)	0.760
ICC	0.66		0.93		0.00		0.93	
Observations	95		75		95		75	
AIC	851.017		601.321		NA		603.227	

Table 51: Output from the negative binomial regressions for the different models for dietary diversity score 1 and Bacteroides_22

	Bacteroidaceae_Bacteroides_22							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-13 (0.632)	<0.001	-14 (0.896)	<0.001	-11 (2.14)	<0.001	-14 (3.37)	<0.001
Diversity	0.24 (0.0696)	<0.001	0.30 (0.0744)	<0.001	0.28 (0.095))	0.004	0.31 (0.0969)	0.001
Total_kcal			0.0022 (0.00672)	0.739			0.0025 (0.00741)	0.731
Age_days					-0.017 (0.0129)	0.196	-0.0023 (0.0223)	0.919
ICC					0.18			
Observations	95		75		95		75	
AIC	158.476		141.155		158.307		143.144	

Table 52: Output from the negative binomial regressions for the different models for dietary diversity score 2 and Bifidobacterium_2

	Bifidobacteriaceae_Bifidobacterium_2							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-7.5 (1.31)	<0.001	-9.2 (2.01)	<0.001	-8.1 (1.64)	<0.001	-6.2 (5.79)	0.288
Diversity2	1.1 (0.337)	0.001	1.4 (0.487)	0.003	1.1 (0.340)	0.001	1.5 (0.523)	0.004
Total_kcal			0.0014 (0.00456)	0.753			0.0019 (0.00455)	0.669
Age_days					0.0024 (0.00402)	0.554	-0.020 (0.0349)	0.570
ICC	0.61		0.76		0.61		0.79	
Observations	95		75		95		75	
AIC	1482.970		1134.047		1484.599		1135.705	

Table 53: Output from the negative binomial regressions for the different models for dietary diversity score 2 and Bifidobacterium_5

	Bifidobacteriaceae_Bifidobacterium_5							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*

(Intercept)	-5.6 (0.607)	<0.001	-5.8 (1.72)	0.001	-4.0 (1.66)	0.017	5.6 (4.82)	0.246
Diversity2	0.61 (0.404)	0.133	0.70 (0.437)	0.109	0.56 (0.400)	0.161	0.92 (0.464)	0.048
Total_kcal			-0.00064 (0.00242)	0.793			0.0012 (0.00254)	0.624
Age_days					-0.0074 (0.00260)	0.005	-0.071 (0.0276)	0.010
ICC	0.80		0.85		0.81		0.87	
Observations	95		75		95		75	
AIC	1622.996		1299.927		1618.945		1294.447	

Table 54: Output from the negative binomial regressions for the different models for dietary diversity score 2 and Bifidobacterium_15

	Bifidobacteriaceae_Bifidobacterium_15							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-10 (2.54)	<0.001	-12 (4.47)	0.009	-17 (3.71)	<0.001	-8.8 (11.4)	0.444
Diversity2	1.3 (0.620)	0.039	0.56 (1.12)	0.616	1.0 (0.726)	0.152	0.62 (1.16)	0.593
Total_kcal			0.0050 (0.00745)	0.501			0.0058 (0.00817)	0.476
Age_days					0.032 (0.0124)	0.009	-0.018 (0.0661)	0.780
ICC	0.70		0.93		0.82		0.94	
Observations	95		75		95		75	
AIC	853.250		601.597		841.677		603.518	

Table 55: Output from the negative binomial regressions for the different models for dietary diversity score 2 and Bacteroides_22

	Bacteroidaceae_Bacteroides_22							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-15 (1.01)	<0.001	-16 (1.37)	<0.001	-12 (2.33)	<0.001	-16 (2.95)	<0.001
Diversity2	0.82 (0.242)	0.001	1.1 (0.268)	<0.001	0.90 (0.339)	0.008	1.0 (0.353)	0.003
Total_kcal			0.0042 (0.00724)	0.558			0.0041 (0.00808)	0.613
Age_days					-0.017 (0.0130)	0.198	0.00091 (0.0215)	0.966
ICC					0.24			
Observations	95		75		95		75	
AIC	159.773		142.823		159.714		144.821	

Table 56: Output from the negative binomial regressions for the different models for dietary diversity score 3 and Bifidobacterium_2

	Bifidobacteriaceae_Bifidobacterium_2			
	Unadjusted	Standard Multivariate 1	Standard Multivariate 2	Standard Multivariate 3

Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-6.2 (0.887)	<0.001	-7.3 (1.46)	<0.001	-6.4 (1.23)	<0.001	-3.3 (5.84)	0.576
Diversity3	0.21 (0.0602)	0.001	0.26 (0.0917)	0.004	0.21 (0.0606)	0.001	0.29 (0.0990)	0.004
Total_kcal			0.00078 (0.00453)	0.863			0.0014 (0.00449)	0.757
Age_days					0.0012 (0.00390)	0.749	-0.026 (0.0353)	0.469
ICC	0.59		0.77		0.59		0.79	
Observations	95		75		95		75	
AIC	1482.294		1134.242		1484.189		1135.676	

Table 57: Output from the negative binomial regressions for the different models for dietary diversity score 3 and Bifidobacterium_5

	Bifidobacteriaceae_Bifidobacterium_5							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-5.0 (1.06)	<0.001	-5.2 (1.15)	<0.001	-3.5 (1.16)	0.003	6.5 (4.58)	0.158
Diversity3	0.13 (0.0707)	0.074	0.15 (0.0762)	0.056	0.12 (0.0694)	0.074	0.19 (0.0800)	0.019
Total_kcal			-0.00085 (0.00241)	0.725			0.0010 (0.00251)	0.690
Age_days					-0.0075 (0.00259)	0.004	-0.071 (0.0268)	0.008
ICC	0.79		0.84		0.79		0.86	
Observations	95		75		95		75	
AIC	1622.220		1299.049		1617.918		1293.169	

Table 58: Output from the negative binomial regressions for the different models for dietary diversity score 3 and Bifidobacterium_15

	Bifidobacteriaceae_Bifidobacterium_15							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-8.9 (1.82)	<0.001	-11 (3.26)	0.001	-16 (3.34)	<0.001	-7.7 (11.2)	0.492
Diversity3	0.27 (0.113)	0.017	0.078 (0.214)	0.716	0.21 (0.138)	0.133	0.089 (0.221)	0.689
Total_kcal			0.0048 (0.00739)	0.515			0.0056 (0.00809)	0.490
Age_days					0.031 (0.0131)	0.019	-0.018 (0.0666)	0.788
ICC	0.67		0.94		0.81		0.94	
Observations	95		75		95		75	
AIC	851.599		601.712		841.514		603.639	

Table 59: Output from the negative binomial regressions for the different models for dietary diversity score 3 and Bacteroides_22

	Bacteroidaceae_Bacteroides_22							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-14 (1.04)	<0.001	-14 (1.11)	<0.001	-11 (2.37)	<0.001	-16 (3.41)	<0.001
Diversity3	0.16 (0.0660)	0.015	0.21 (0.0590)	<0.001	0.17 (0.0749)	0.020	0.19 (0.0783)	0.017
Total_kcal			-0.0012 (0.00764)	0.878			-0.0023 (0.00785)	0.769
Age_days					-0.019 (0.0136)	0.154	0.0099 (0.0235)	0.673
ICC	0.15				0.33			
Observations	95		75		95		75	
AIC	161.242		144.627		160.465		146.458	

Table 60: Output from the negative binomial regressions for the different models for dietary diversity score 4 and Bifidobacterium_3

	Bifidobacteriaceae_Bifidobacterium_3							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-2.8 (2.79)	0.322	-2.6 (2.80)	0.354	5.3 (4.14)	0.197	2.6 (11.1)	0.814
Diversity4	-0.63 (0.330)	0.057	-0.62 (0.330)	0.059	-0.72 (0.393)	0.066	-0.61 (0.344)	0.075
Total_kcal			0.00020 (0.00379)	0.957			0.0022 (0.00424)	0.793
Age_days					-0.041 (0.0122)	0.001	-0.031 (0.0633)	0.625
ICC	0.95		0.96		0.97		0.96	
Observations	95		75		95		75	
AIC	1055.901		928.733		1042.092		930.489	

Table 61: Output from the negative binomial regressions for the different models for dietary diversity score 4 and Bifidobacterium_2

	Bifidobacteriaceae_Bifidobacterium_2							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-5.8 (0.768)	<0.001	-6.9 (1.33)	<0.001	-6.2 (1.17)	<0.001	-6.1 (5.87)	0.298
Diversity4	0.30 (0.0804)	<0.001	0.37 (0.128)	0.003	0.30 (0.0814)	<0.001	0.38 (0.131)	0.004
Total_kcal			0.00092 (0.00451)	0.839			0.0011 (0.00457)	0.818
Age_days					0.0018 (0.00394)	0.646	-0.0050 (0.0347)	0.886

ICC	0.56	0.77	0.57	0.77
Observations	95	75	95	75
AIC	1481.583	1134.137	1483.362	1136.116

Table 62: Output from the negative binomial regressions for the different models for dietary diversity score 4 and Bifidobacterium_5

	Bifidobacteriaceae_Bifidobacterium_5							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-5.0 (0.934)	<0.001	-5.1 (1.03)	<0.001	-3.5 (1.07)	0.001	5.6 (4.86)	0.251
Diversity4	0.21 (0.0988)	0.036	0.23 (0.107)	0.034	0.20 (0.0978)	0.043	0.24 (0.116)	0.041
Total_kcal			-0.00081 (0.00240)	0.736			0.00085 (0.00252)	0.737
Age_days					-0.0074 (0.00261)	0.004	-0.062 (0.0275)	0.024
ICC	0.78		0.83		0.78		0.86	
Observations	95		75		95		75	
AIC	1621.275		1298.403		1617.196		1294.425	

Table 63: Output from the negative binomial regressions for the different models for dietary diversity score 4 and Bifidobacterium_15

	Bifidobacteriaceae_Bifidobacterium_15							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	0.00	<0.001	0.00	<0.001	0.00	<0.001	0.00	0.442
Diversity4	1.40	0.032	1.21	0.506	1.33	0.115	1.21	0.503
Total_kcal			1.00	0.501			1.01	0.485
Age_days					1.03	0.012	0.99	0.819
ICC	0.65		0.93		0.82		0.94	
Observations	95		75		95		75	
AIC	852.527		601.413		841.309		603.360	

Table 64: Output from the negative binomial regressions for the different models for dietary diversity score 4 and Bacteroides_22

	Bacteroidaceae_Bacteroides_22							
	Unadjusted		Standard Multivariate 1		Standard Multivariate 2		Standard Multivariate 3	
Predictors	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*	*Estimate (SD)*	*p*
(Intercept)	-8.0 (1.83)	<0.001	-11 (2.86)	<0.001	-16 (3.63)	<0.001	-8.6 (11.2)	<0.001
Diversity4	0.34 (0.157)	0.001	0.19 (0.283)	<0.001	0.29 (0.183)	0.010	0.19 (0.289)	0.006
Total_kcal			0.0050 (0.00739)	0.964			0.0056 (0.00803)	0.843

Age_days					0.032 (0.0128)	0.227	-0.015 (0.0651)	0.428
ICC					0.21			
Observations	95		75		95		75	
AIC	160.108		143.520		160.518		144.947	