GUT MICROBIOTA, INFANT FEEDING, AND NEURODEVELOPMENT: AN ANALYSIS IN EARLY LIFE

By

Sihan Bu

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Human Nutrition – Doctor of Philosophy

2023

ABSTRACT

The human gut microbiota is a complex community of microorganisms Infant diet influences the composition and diversity of the gut microbiota, which may impact neurodevelopmental outcomes. Herein, the mediating role of the infant gut microbiota in the associations between infant diet and infant neurodevelopment and an analysis of the influence of breastfeeding patterns on infant gut microbiota are presented.

Participants in the Michigan Archive for Research on Child Health (MARCH), a cohort study in Michigan, provided infant fecal samples at 3 months of age and neurodevelopment information using the Ages and Stages Questionnaire at 9 months of age. 16S rRNA sequencing data was processed through mothur. Microbiota and statistical analyses were conducted using R.

In Chapter 2, associations between gut microbiota and neurodevelopmental outcomes are described. Gut microbiota richness (Chao 1) was negatively associated with gross motor scores. However, gut microbial diversity (Shannon index) was positively associated with problem-solving scores. Beta diversity (Bray-Curtis) was associated with fine motor and communication scores. Thus, the gut microbiota was associated with cognitive development.

Chapter 3 examined the potential mediating role of early-life gut microbiota in the associations between infant diet and neurodevelopmental outcomes. The gut microbiota was impacted by diet. Breastfeeding and vitamin D supplementation was positively associated with fine motor scores. Infant gut microbial composition, measured by the Bray-Curtis dissimilarity index, mediated the association between infant feeding and fine motor scores. These results suggest the importance of promoting optimal gut health through nutrition to support healthy cognitive development.

In Chapter 4 relationships between breastfeeding patterns (breastfed, bottle-fed, and mix-

fed), the proportions of breastmilk intake and infant gut microbiota among exclusively breastmilk-fed infants at 3 months of age are described. Infants fed at the breast had a lower abundance of *Bifidobacterium* but a higher abundance of *Enterobacteriaceae* compared to bottle- and mixed-fed infants. These microbiotas were then compared to those of infants fed some formula. Though bottle-fed infants were 100% breastmilk fed, they had similar microbiota composition as infants fed with >50% and <50% breastmilk. Thus, breastfeeding patterns influence the gut microbiota of infants.

In summary, this work describes relationships among infant diet, breastfeeding patterns, gut microbiota, and neurodevelopment. The work underscores the importance of promoting optimal gut health through infant feeding practices and nutritional interventions, such as vitamin D supplementation, to support neurodevelopment. Notably, this work advances prior work by using infant dietary intake data collected in the week, as well as in the 24 hours, immediately prior to stool collection. Overall, these results contribute to our understanding of the role of gut microbiota in infant development and may inform the development of interventions aimed at promoting healthy gut microbiota and neurodevelopmental outcomes in early life.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my PhD advisor, Dr. Sarah Comstock, for her unwavering support and guidance throughout my graduate studies. As an international student and non-native English speaker, I was initially unsure of my ability to succeed in this program. However, her patience, encouragement, and willingness to work with me despite our linguistic and cultural differences have been invaluable to my success. I have learned so much from her not only about my field of study but also about academic research, critical thinking, and professionalism. Her mentorship has been a highlight of my academic journey, and I am grateful for the opportunity to have worked under her supervision.

I would like to extend my sincere appreciation to my dissertation committee members Drs. Jean Kerver, Lixin Zhang, and Rita Strakovsky for their invaluable support, insightful feedback, and invaluable contributions to my research. Their expertise and guidance were instrumental in shaping my research ideas, improving the quality of my work, and helping me to achieve my academic goals. I am deeply grateful for their time, effort, and dedication to my success. Thank you for believing in me and for your invaluable contributions to my academic journey.

Finally, I greatly appreciate the funding agency National Institutes of Health for supporting this research. I would also like to thank the financial assistance: John Harvey Kellogg endowment, Glenn R. Dean and Anita C. Dean Endowed Fellowship Within the College of Human Ecology, P. Vincent Hegarty Food Science and Human Nutrition Quality in Education, International Peace Scholarship by P.E.O for supporting my PhD studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA: Analysis of variance

ASQ-3: Ages and Stages Questionnaire, third edition

BMI: Body mass index

CH: Calinski-Harabasz

CI: Confidence interval

DNA: Deoxyribonucleic acid

FDR: False discovery rate

GBA: Gut-microbiota-axis

GF: Germ-free

GI: Gastrointestinal tract

HMO: Human milk oligosaccharides

HSD: Honest Significant Differences

IPL: Inferior parietal lobule

JSD: Jensen-Shannon distance

MaAsLin: Microbiome multivariate association with linear models

MARCH: Michigan Archive for Research on Child Health

MRI: Magnetic Resonance Imaging

MSEL ELC: Mullen Scales of Early Learning Composite

NB: Negative binomial generalized linear model

OUT: Operational Taxonomic Unit

PAM: Partitioning around medoids

PCoA: Principal coordinate analysis

PERMANOVA: Permutational multivariate analysis of variance

rRNA: Ribosomal ribonucleic acid

SD: Standard deviation

SMA: Supplementary motor area

CHAPTER 1: INTRODUCTION

The human gut microbiota is a complex and diverse community of microorganisms that live in the human gastrointestinal tract (Thursby & Juge, 2017). This community consists of trillions of microorganisms, including bacteria, viruses, fungi, and archaea (Matijašić et al., 2020). The gut microbiota plays a vital role in human health, with research showing that it is involved in a wide range of processes such as digestion (Oliphant & Allen-Vercoe, 2019), nutrient absorption (Krajmalnik-Brown et al., 2012), and immune system regulation (Belkaid & Hand, 2014). Furthermore, studies have linked alterations in the gut microbiota to various health conditions such as obesity (Liu et al., 2021), diabetes (Li et al., 2020), and inflammatory bowel disease (Qiu et al., 2022). Other research has also explored the potential links between the gut microbiota and mental health disorders such as depression and anxiety (Clapp et al., 2017). Understanding the relationship between the gut microbiota and human health is an area of active research, and it has the potential to lead to new treatments and interventions to improve human health and well-being.

Development of gut microbiota in early life is tightly related to health later in life (Kundu et al., 2017). Healthy breastfed infants are primarily colonized with *Bifidobacterium* strains (Saturio et al., 2021). However, infants with later atopic disease displayed a reduced ratio of bifidobacteria to clostrida, caused by reduced bifodobacteria and increased clostridia colonization (Björkstén et al., 2001; Kalliomäki et al., 2001). A higher risk of obesity later in life was also attributed to the decreased fecal bifodobacteria early infancy as compared by healthy children (Kalliomäki et al., 2008). Colicky infants have been shown to have increased colonization of *Clostridium difficile* compared with non-colicky infants (Savino et al., 2004). Finally, the gut microbial developmental trajectory during infancy were found to be associated with later development type 1 diabetes (Kostic et al., 2015).

Infant feeding practices include breastfeeding, formula feeding and mix feeding (both breastfeeding and formula feeding). Breastmilk feeding refers to the practice of feeding infants with breast milk produced by their mothers. Human milk is the ideal source of nutrition for newborns, providing necessary nutrients and immune factors for optimal growth and development (C. R. Martin et al., 2016). The modes of breastmilk feeding are comprised of direct breastfeeding, expressed breastfeeding, and mixed feeding (Pang et al., 2017; Pérez-Escamilla et al., 2023). Direct breastfeeding is when the infant feeds directly from the mother's breast, while expressed breast milk feeding is when the infant is fed with human milk that has been extracted from the breast using a pump and provided through a bottle, cup, or spoon. Mixed feeding is a combination of both, where the infant is fed directly at the breast and also given expressed breast milk (Pang et al., 2017). Although breastmilk is the ideal source of nutrition for infants, it may not provide enough vitamin D for optimal growth and development (Balasubramanian, 2011). Vitamin D deficiency in breastfed infants can result in nutritional rickets which is a bone-related condition (Shore & Chesney, 2013). Formula feeding, on the other hand, is a viable alternative for infants who cannot be breastfed (Stevens et al., 2009). Formula provides a complete source of nutrition for infants, and it is designed to mimic the composition of breast milk.

Infant feeding practices profoundly influence the colonization and maturation of the infant gut microbiome (Li et al., 2021; O'Sullivan et al., 2015). The human milk oligosaccharides (HMOs) are one of the main components of breast milk and they are utilized by *Bifidobacterium* in infant’s gut, which can inhibit the growth of pathogenic bacteria and modulate the mucosal barrier function and immune response (Le Huërou-Luron et al., 2010; Marcobal et al., 2010; Sudo et al., 1997). Formula-fed infants have a distinct gut microbial composition from breastfed

infants (Ma et al., 2020; O'Sullivan et al., 2015; Yatsunenko et al., 2012). Exclusively breastfed infants had lower bacterial diversity, increased abundance of *Bifidobacterium,* and decreased abundance of *Lachnospiraceae* compared to partially or non-breastfed infants (Baumann-Dudenhoeffer et al., 2018; Forbes et al., 2018). Exclusively formula-fed infants displayed a more diverse gut microbiota with a lower abundance of *Bifidobacterium* species and an increased abundance of *Clostridium* species and *Enterobacteriaceae* species due to lacking HMOs and higher protein contents in infant formula compared to breastfed infants (Bäckhed et al., 2015; Benno et al., 1984; Penders et al., 2007). *Lactobacilli*/*Enterococci* counts were also higher in breastfed infants compared to formula-fed infants (Rinne et al., 2005). Recently, infant formula has been improved by adding oligosaccharides, making it possible to establish Bifidobacterium-rich gut microbiota in infants (Veereman-Wauters et al., 2011). Furthermore, formula-fed infants form an adult-like gut microbiota composition at an early age (Bäckhed et al., 2015). In conclusion, breastfeeding plays a critical role in the development of the infant gut microbiota, promoting the growth of beneficial bacteria and providing long-term health benefits for the infant.

Pumping breastmilk into a bottle is one of the common breastmilk feeding modes; however, it can impact the bacterial composition of breast milk (Differding & Mueller, 2020; Moossavi & Azad, 2020; Weiss, 2005). Human milk bacteria is a potential source of bacteria that colonize the infant gut (Urbaniak et al., 2016). Therefore, the changes of human milk microbial composition could possibly influence the infant gut microbiota. However, the consequences of pumping and breastfeeding on infant gut microbiota have not been well studied. *Streptococcus spp*. and *Veillonella dispar* co-occurred in breast milk and infant's stool but this co-occurrence was depleted when infants were fed with pumped breastmilk (Fehr et al., 2020).

They also reported that infant gut microbiota composition was not associated with breastmilk feeding patterns (breastfeeding versus pumping) (Fehr et al., 2020). While the impact of pumping breast milk on the infant gut microbiota is not yet fully understood, studies suggest that it may alter the microbial composition of breast milk, potentially affecting the bacterial colonization of the infant gut.

The introduction of complementary foods during weaning is a critical period for the development of the infant gut microbiota. During weaning, the introduction of complementary food causes an increase in alpha diversity of gut microbiota, resulting in the replacement of *Proteobacteria* and *Actinobacteria* by *Firmicutes* and *Bacteroidetes* phyla as the dominant species (Fallani et al., 2011; Koenig et al., 2011). The infant gut microbial diversity increased significantly with the consumption of solid foods at 9 months of age compared to milk-based diet at 4 months of age (McKeen et al., 2022). The timing of the introduction to solid in infancy was associated with altered gut microbial composition, which differed by duration of breastfeeding (Differding et al., 2020).

Delivery mode is recognized as an essential driver of early gut microbiota composition in full-term born infants (Mitchell et al., 2020; Munyaka et al., 2014). The maternal vaginal microbiome is considered the first natural microbial exposure to newborn babies, which results in neonatal gut colonization by the mother's vaginal microbiota, such as *Lactobacillus* and *Prevotella* (Biasucci et al., 2010; Dominguez-Bello et al., 2010). In contrast, cesarean section (C-section) born infants are not directly exposed to vaginal microbiota; however, they are more likely to be colonized by some environmental microorganisms from maternal skin, the hospital staff, or the hospital environment (Bäckhed et al., 2015; Biasucci et al., 2010; Bokulich et al., 2016; Fouhy et al., 2012; Rodríguez et al., 2015), such as *Staphylococcus*, *Corynebacterium*, and

Propionibacterium spp. Additionally, C-section delivered infants also show a reduced diversity of gut microbiota, and they are less likely to be colonized by *Bifidobacterium* and *Bacteroides* but are more frequently colonized by *Clostridium sensu stricto* (cluster I) and *Clostridium difficile* (Adlerberth et al., 2007; Akagawa et al., 2019; Biasucci et al., 2010; Del Chierico et al., 2015; Dominguez-Bello et al., 2010; Hill et al., 2017; Jakobsson et al., 2014; Neu & Rushing, 2011; Penders et al., 2006). Therefore, c-section might be leading to the dysbiosis of infant gut microbiota since it reduced the gut microbial diversity compared to vaginal delivery (Hoang et al., 2021). However, breast milk might help reverse this adverse outcome induced by c-section (Zhang et al., 2021). These gut microbial differences between vaginally and C-section-born babies decrease at 4 months and 12 months, but the gut microbiota of C-section-born infants remain more heterogeneous (Bäckhed et al., 2015; R. Martin et al., 2016).

In the early stages, antibiotics exposure is a significant factor disrupting the normal gut microbiota colonization and development. Intrapartum antibiotic prophylaxis (IAP) is commonly used to prevent severe the bacterial infections, sepsis, and meningitis caused by *Streptococcus agalactiae*, group B *Streptococcus* (GBS), in newborn and young infants (Le Doare & Heath, 2013; Moore et al., 2003; Schrag et al., 2000; Thigpen et al., 2011). The impact of maternal IAP on infant gut microbiota colonization is present at the age of two days (Nogacka et al., 2017). IAP infants have been shown to have a higher abundance of *Enterobacteriaceae* (Mazzola et al., 2016), and a lower abundance of *Bifidobacterium spp*. at the age of one week (Corvaglia et al., 2016). At three months of age, a decreased infant gut microbiota richness, a depletion of *Bacteroidetes*, and increased *Firmicutes* were observed, which persisted to 1 year among IAP-exposed infants delivered by emergency C-section born babies and were not breastfed exclusively at 3 months (Azad et al., 2016). In addition to prenatal exposure, postnatal antibiotic

use also has a potential impact on gut microbiota development. Early empiric antibiotic use in preterm infants in the first week of life was associated with lower gut microbial diversity in the second and third weeks (Greenwood et al., 2014). Antibiotics administration caused a lower abundance of *Bacteroides spp* during the first three months of life (Eck et al., 2020). Broad-spectrum antibiotics are used to treat suspected early-onset neonatal sepsis (sEONS). The gut microbial composition changed significantly after the antibiotics treatment (Reyman et al., 2022). In addition, antibiotics treated infants showed a decreased abundance of *Bifidobacterium spp*. and increased abundance of *Klebsiella* and *Enterococcus spp*. compared to non-antibiotics treated infants (Greenwood et al., 2014; Korpela et al., 2020; Reyman et al., 2022). Epidemiological studies have been conducted that early exposure to antibiotics is associated with asthma, allergic diseases, overweight, inflammatory bowel disease, and celiac disease in childhood (Chelimo et al., 2020; Dydensborg Sander et al., 2019; Kronman et al., 2012; Murk et al., 2011; Saari et al., 2015; Zven et al., 2020).

Associations between maternal pre-pregnancy BMI and infant gut microbiota are modified by delivery mode (Mueller et al., 2016; Singh et al., 2020). In one study, mothers who were overweight or obese before becoming pregnant had a significantly different gut microbial community structure, such as the enrichment in the *Bacteroides* and depletion in the *Enterococcus*, *Acinetobacter*, *Pseudomonas*, and *Hydrogenophilus* in vaginally born infants (Mueller et al., 2016). On the contrary, maternal pre-pregnancy BMI was not associated with infant gut microbial community structure (Mueller et al., 2016). Another study observed that maternal overweight or obesity was associated with increased infant gut microbial diversity in vaginally born infants, while there was no association in C-section born infants (Singh et al., 2020). In addition, Sugino et al. found that infant gut microbiota membership tended to differ by

maternal pre-pregnancy BMI category ($18.5 \leq BMI < 25$, $25 \leq BMI < 30$, $BMI \geq 30$) (Sugino et al., 2019).

Gut microbiota secretions, such as peptides, gut hormones and neuroactive substances and microbiota-derived products, and microbiota-derived metabolites, which will modulate the brain through immune system, neuroendocrine system, enteric nervous system, circulatory system, and vagus nerve by altering receptor activity and neurotransmission due to microbial metabolites entry (Bonaz et al., 2018; Braniste et al., 2014; Brown et al., 2003; Carabotti et al., 2015; Farzi et al., 2018; Onyszkiewicz et al., 2020). This process might lead to negative results, such as neurodegenerative diseases and neurodevelopmental and neuropsychiatric diseases (Luczynski et al., 2016; Zhang et al., 2022).

Early life is crucial for brain development and the establishment of cognitive abilities (Gilmore et al., 2018), which might impact the future life of the child (Longo et al., 2021; Nelson et al., 2007). The gut microbiota that colonizes the gastrointestinal tract also develops rapidly after birth in response to the environmental factors mentioned above. Therefore, due to the GBA, the microbiota's colonization of the gastrointestinal tract appears to happen in parallel and interactively with brain development (Carlson et al., 2018; Gao et al., 2019). Loughman et al. observed a clear relationship between the decreased abundance of *Prevotella* collected when the infants were 12 months of age and increased behavioral problems at 2 years of age (Loughman et al., 2020). Carlson et al. showed that infants with a high abundance of beneficial gut microbiota such as *Lactobacillus* and *Bacteroides* might improve overall cognitive performance (Carlson et al., 2018). Lower alpha diversity was associated with lower cognitive performance as a result of adverse health outcomes, including type 1 diabetes and asthma in the future (Abrahamsson et al., 2014; Carlson et al., 2018; Kostic et al., 2015). Animal studies have also provided insights into

the gut microbiota and brain development in the early postnatal period. In adulthood, at 8-9 weeks of age, GF mice exhibited an anxiety-related behavior compared to SPF mice. Besides, the colonization of adolescent (5-6 weeks) GF mice by gut microbiota could not reverse the monoamine neurotransmitter-related gene expressions (Pan et al., 2019). Therefore, the early identification of abnormal neurodevelopment is essential to lead to earlier treatment and positively alter the long-term outcomes (Bian et al., 2012; Chaudhari & Kadam, 2012; Cioni et al., 2016; Hadders-Algra, 2021; Siller et al., 2013).

The Ages and Stages Questionnaire (ASQ) is a parent-completed screening tool that pinpoints developmental progress in children. The ASQ was developed by D. Bricker and J. Squires from the University of Oregon, US. The Ages & Stages Questionnaires, Third Edition (ASQ-3) can take 10-15 minutes for parents to complete at home, in a waiting room, during a home visit, or in an interview, as well as 2-3 minutes for professionals to score. In addition, the ASQ-3 is available in different languages, such as Arabic, Chinese, English, French, Spanish, and Vietnamese. The ASQ-3 has been shown to effectively differentiate between children with developmental delays and those with typical development. The overall sensitivity of ASQ-3 or the ability of ASQ-3 to correctly identify children with developmental delay is 86%. The overall specificity of ASQ-3, or the ability of ASQ-3 to correctly identify typically developing children, is 85%. The ASQ-3 comprises 5 areas: communication, gross motor, fine motor, problem-solving, and personal-social for children from 1-66 months. Scores for each area fall between 0 and 60. Parents indicate for each item “yes” if child performs the item and scores 10 points, “sometimes” indicating an occasional or emerging skills and child scores 5 points, or “not yet” if child doesn’t perform the behavior and scores 0 points. The cutoff points for ASQ-3 9 months are 26.26, 32.27, 42.82, 39.11, 30.69 for communication, gross motor, fine motor, problem-

solving and personal social, respectively. If the total score of each area is below cutoff, then further assessment with a professional maybe needed (Questionnaires, 2022).

A more recent study, enrolling 309 full-term healthy infants, evaluated the relationships between fecal microbiota composition, also estimated through 16S sequencing, at 3–6 months of age and score of the Age and Stage Questionnaire (ASQ) at 3 years of age (Sordillo et al., 2019). The authors used a co-abundance factor approach, which allowed assigning four scores to each individual based on the co-abundance of the 25 most abundant bacterial taxa. They then mathematically correlated these microbiota scores to the ASQ scores. Interestingly, scores in communication and personal social skills were negatively associated with the microbiota factor comprising relative high abundance of *Lachnospiraceae* and *Clostridiales* and low abundance of *Bacteroidetes*, while fine motor skills scores were negatively correlated with the factor comprising relative high abundance of *Bacteroidetes* and low abundance of *E. coli* and *Bifidobacterium*, two early colonizers. A tendency for increased Shannon diversity index with lower personal and social skills was also noticed. In another study, *Staphylococcus caprae* was negativealy corrated with ASQ scores, but *Escherichia coli* were positively correlated with ASQ scores (Rozé et al., 2020).

Breastfeeding is a nutrient delivery system to continuously transfer all essential nutrients in appropriate amounts from mother to infant (Hinde & German, 2012). In addition to being a meal for infants, it also has a profound long-term impact on their cognitive and behavioral development and mental health (Lockyer et al., 2021; Raju, 2011). Guxens et al. and Leventakou et al. found that a higher duration of exclusive breastfeeding was positively associated with memory performance, early language development, and motor skills at 14 months (Guxens et al., 2011) and 18 months (Leventakou et al., 2015) of age as measured by

Bayley Scales of Infant Development. These cognitive benefits from breastfeeding seem to be extended to childhood and adolescence. Similarly, another study showed communication, and global motor had more delays in preschoolers who were breastfed for only 3 months compared to those with 6- and 12-months breastfeeding duration when using the Ages and Stages Questionnaire-3 (Saliaj, 2015). A large population-based cohort study reported that 4-year-old children with a duration of exclusively breastfeeding for over 6 months after birth have better executive function (cognitive control) than those with less than a 6-month breastfeeding period (Julvez et al., 2014). Bernard et al. observed that the breastfeeding experience was related to improved cognitive development among 2 and 3 years old children with Communicative Development Inventory and Ages and Stages Questionnaire (Bernard et al., 2013). In addition, Mandy et al. reported that predominant breast milk feeding in the first 28 days of life was positively associated with IQ, academic achievement, working memory, and motor function at 7 years of age among preterm infants (Belfort et al., 2016). There is conflicting evidence on whether breastfeeding can improve cognitive development. Breastfeeding was found to have little or no effect on intelligence among children aged 5-14 years as measured by Peabody individual achievement test (Der et al., 2006). A long duration of breastfeeding was not associated with later cognitive development in 9- to 10-year-old children in South India using Kaufman Assessment Battery for Children (Veena et al., 2010).

Formula-fed infants gain more weight during infancy than breastfed infants because of the higher protein content in formula (Farrow et al., 2013; Kramer et al., 2004; Ren et al., 2022). Though there is some evidence to suggest a positive association between protein intake and neurodevelopment in infancy, the evidence is mixed. In a cohort study, increased protein intake in the first month of life was not associated with better cognitive, language, and motor scores or

decreased sensory impairments at 2 years of age (Cester et al., 2015). However, other studies reported the opposite results. Increased protein intake in the first week after birth was associated with higher Mental Development Index scores at 18 months in extremely low birth weight infants (Stephens et al., 2009). A positive association was demonstrated between protein intake during the first 28 days and cognitive and motor scores at 2 years in infants born at a gestational age < 31 weeks (Coviello et al., 2018).

Based on the literature reviewed above, the main objective of this dissertation was to investigate the associations between infant feeding practices, infant gut microbiota and infant neurodevelopmental outcomes. The first aim of this body of work was to examine the associations between infant gut microbiota at 3 months of age and infant neurodevelopmental outcomes at 9 months of age. The second aim was to determine whether infant feeding practices during early infancy influence the infant neurodevelopmental outcomes, and also investigated the mediating role of the early life gut microbiota in the association between infant feeding methods and neurodevelopment. The third aim was to examine the effects of breastfeeding patterns (breastfeeding at the breast, breastfeeding from the bottle and breastfeeding from both breast and bottle) on infant gut microbiota.

CHAPTER 2: THE RELATIONSHIPS BETWEEN INFANT GUT MICROBIOTA AND INFANT NEURODEVELOPMENT, AS MEASURED BY THE AGES AND STAGES QUESTIONNAIRE

2.1 Abstract

The gut-microbiota-axis (GBA) refers to the bidirectional communication between gut microbiota and the central nervous system. Infancy is a critical period for colonizing gut microbiota and brain development. The abnormal compositional gut microbiota development during early life can lead to worse cognitive performance later in life. However, the association between early-life gut microbiota and later neurodevelopment outcomes is unclear. Therefore, this study aimed to identify the relationship between infant gut microbiota at 3 months of age and neurodevelopment at 9 months of age. Deoxyribonucleic acid (DNA) was extracted from 64 samples, 16S ribosomal ribonucleic acid (rRNA) libraries were made, and libraries were sequenced by Illumina MiSeq. Sequences were processed using mothur, and data were analyzed in R. Infant diet information was reported at three months of age. Neurodevelopment was assessed by the Ages and Stages Questionnaire, third edition (ASQ-3) when the infants were 9 months old. A higher Chao 1 index was associated with lower gross motor skills. Shannon index was positively related to problem-solving. The Bray-Curtis dissimilarity matrix was associated with fine motor and communication. Three clusters of gut microbiota were identified: Cluster 1 (*Lachnospiraceae* unclassified-dominated), Cluster 2 (*Bifidobacterium*-dominated), and Cluster 3 (*Bacteroides*-dominant cluster). Infants whose gut microbiota were in Cluster 3 had lower problem-solving scores than those in Cluster 1. These findings suggest an association between characteristics of the infant gut microbiota at age 3 months and gross motor, fine motor, communication, and problem-solving skills at age 9 months.

2.2 Keywords

Bifidobacterium, *Lachnospiraceae* unclassified, *Bacteroides*, Ages and Stages Questionnaire, gross motor, fine motor, communication, problem-solving, infants

2.3 Introduction

Gut microbiota plays an important role in maintaining human health (Thursby & Juge, 2017). Mounting evidence from animal studies shows the bidirectional communication between the gut and brain, referred to as the GBA (Carabotti et al., 2015). For example, germ-free (GF) mice displayed decreased anxiety-like behavior compared to specific pathogen-free mice with normal gut microbiota in the elevated plus maze and the light-dark box text (Heijtz et al., 2011; Neufeld et al., 2011), which can be reversed by moving GF mice to conventional mice cages covered with feces from conventional mice (Clarke et al., 2013). GF mice were found to have impaired short-term recognition and working memory (Gareau et al., 2011), but increased locomotor and rearing behaviors (Heijtz et al., 2011).

Early life is crucial for brain development and the establishment of cognitive abilities (Gilmore et al., 2018), which might impact a child's future life (Nelson et al., 2007). The gut microbiota that colonizes the gastrointestinal tract (GI) also develops rapidly after birth in response to environmental factors (Sugino, Ma, Paneth, et al., 2021), such as delivery mode (Munyaka et al., 2014; Sugino et al., 2019), antibiotic exposure (Eck et al., 2020) (Reyman et al., 2022), feeding practice (Haddad et al., 2021; O'Sullivan et al., 2015; Sugino, Ma, Kerver, et al., 2021), etc. Breastfed infants with vitamin D supplementation were shown to have different gut microbial diversity compared to non-supplemented, breastfed infants (Ma et al., 2022). Therefore, the microbiota's colonization of the GI appears to happen in parallel and interactively with brain development (Ratsika et al., 2023). An enhanced understanding of the development of the GI reflects how the brain develops in early life and vice versa, allowing gut microbiota to be a regulator of early-life neurodevelopment (Jena et al., 2020). Loughman et al. observed a clear relationship between the decreased abundance of *Prevotella* collected when the infants were 12 months of age and increased behavioral problems at two years of age (Loughman et al., 2020).

Carlson et al. showed that infants with a high abundance of beneficial gut microbiota, such as *Lactobacillus* and *Bacteroides,* demonstrated better overall cognitive performance (Carlson et al., 2018). Lower alpha diversity, indicating a less mature microbiota of infants, was associated with lower cognitive performance and led to adverse health outcomes, including type 1 diabetes and asthma, in the future (Abrahamsson et al., 2014; Carlson et al., 2018; Kostic et al., 2015).

Few studies have investigated the association between early infant gut microbiota and neurodevelopment later in life, accounting for feeding practices, antibiotic use since birth, delivery mode, infant race, maternal education level, gestational age at birth, pre-pregnancy body mass index (BMI), and maternal age. Thus, this cohort study assessed whether infant gut microbiota at 3 months was associated with infant neurodevelopment at 9 months of age measured by the ASQ-3 (Squires J, 2009).

2.4 Materials and methods

2.4.1 Population characteristic

A total of 64 participants were enrolled as part of the Michigan Archive for Research on Child Health (MARCH), an ongoing population-based pregnancy and birth cohort in Michigan's lower peninsula. The participants provided informed consent to obtain the questionnaire and provide the infant stool samples at three months. The covariates used were from the MARCH Prenatal 1 Survey questionnaire that asks about mothers' education level, mother's height, pre-pregnancy weight, and maternal age. The birth certificate information includes infant sex, mode of delivery, and estimated weeks of gestation. MARCH 3-month survey dictionary includes the infant race. Infants with gestational age less than 37 weeks were excluded from the analyses Fecal was collected when the infants were 3 months old. At the same time, the sample collection form was completed, which asks whether the infants received breast milk or formula in the past

24 hours and the past week before collecting the sample, whether the infant was received antibiotics since birth, and other dietary history information Infants with missing data were also excluded. The Michigan State University Human Research Protection Program approved the study (IRB# 16-1429).

2.4.2 Classification of infant dietary intake

Infants were split into four groups based on their dietary intake in the past day: breastfeeding, breastfeeding with vitamin D supplementation, partial breastfeeding, and formula feeding. Seven feeding groups were classified according to the infant's dietary intake in the past week: 100% breastmilk feeding, 80% breastmilk feeding, 50-80% breastmilk feeding, 50% breastmilk feeding, 20-50% breastmilk feeding, 20% breastmilk feeding, and 100% formula feeding.

2.4.3 Ages and Stages Questionnaire

At approximately 9 months old, parents completed the ASQ-3 (Squires J, 2009) during a phone interview as part of the MARCH 9-Month Survey. The ASQ-3 is a parent-completed screening tool that pinpoints developmental progress in children. The ASQ-3 comprises 5 areas: communication, gross motor, fine motor, problem-solving, and personal-social for children from 1-66 months. Scores for each area fall between 0 and 60. Parents indicate for each item “yes” if the child performs the item and the child scores 10 points, “sometimes” indicating an occasional or emerging skill and the child scores 5 points, or “not yet” if the child doesn’t perform the behavior and scores 0 points. The cutoff scores for ASQ-3 at 9 months are 26.26, 32.27, 42.82, 39.11, 30.69 for communication, gross motor, fine motor, problem-solving, and personal social, respectively. If the total score of each area is below the cutoff, then further assessment with a

professional may be needed.

2.4.4 Sample collection

Collection kits were assembled at the dry research lab at MSU and sent to the participants by mail. The collection kits include an instruction for collecting a fecal sample at home, diapers for infant fecal samples, an OMNIgene•GUT tube (DNA genotek, Ontario, CA) for sample collection, and a box with postage to return the sample. Fecal samples were collected by parents from the infant's diaper when the infant was approximately three months of age. Stool samples were returned to the lab in the pre-paid mailer through the United States postal system. Fecal samples were aliquoted into sterile Eppendorf tubes (Thermo Fisher Scientific, Waltham, MA) and stored at -80°C once reaching the lab.

2.4.5 Laboratory procedures

2.4.5.1 DNA extraction and 16S rRNA gene amplification

DNA extractions were performed using the DNeasy Powersoil Pro kit (Qiagen MoBio, Carlsbad, CA). The V4 region of the 16S rRNA gene was amplified using the Schloss lab primers (500B-700A). Primers SB501-SB508 and SA701-SA712 were ordered from Integrated DNA Technologies (Coralville, IA). PCR amplification procedure followed the mothur wet lab documentation (Kozich et al., 2013). A final reaction volume of 20 µL with at most 10 ng of template DNA, primer pairs, and Accumprime Pfx Supermix (Thermo Fisher Scientific, Waltham, MA) was used. The PCR reactions were performed in triplicate and amplified using a thermocycler. A negative control without template DNA was included to control for non-specific amplification. Thermocycler conditions were set as follows: 1x (95 ℃ for 2 min); 30x (95 ℃ for 20 s, 55℃ for 15 s, 72℃ for 5 min); 10 min for 72℃. The PCR amplicons were checked by

agarose gel electrophoresis on a 1% agarose gel using 1X TBE buffer at 200 V for 30 min. Successful PCR triplicate amplicons were pooled and cleaned with Agencourt AMPure XP (Beckman Coulter, Brea, CA) with a few changes to the protocol: PCR products were purified by 0.7X AMPure XP, and DNA was eluted using 20 μL of low EDTA TE buffer (IDT, Coralville, IA). After purification, the 16S rRNA PCR amplicons concentrations were determined by Quant-IT dsDNA assay kit (Invitrogen, Carlsbad, CA). An equal amount (ng) of DNA in each sample was pooled for sequencing. The Michigan State University Research Technology Support Facility Genomics Core conducted paired-end 250 base-pair sequencing on the Illumina MiSeq platform using V2 chemistry. The average number of reads per sample was 21605, with at least 82% of reads per sample having a read quality greater than or equal to 30.

2.4.5.2 Processing and analysis of sequencing data

16S rRNA sequences were processed using mothur, following the mothur Miseq standard operating procedure (Schloss et al., 2009). Taxonomy was assigned to operational taxonomic units (OTU) by phylotype using the RDP reference database (version 18). Samples were rarefied to 2,624 reads per sample before further analysis. Rarefaction curves were generated to confirm adequate community coverage.

2.5 Statistical analysis

All data were analyzed using R (version 4.2.2). Data normality was tested using Shapiro–Wilk test (stats package). For categorical variables of descriptive analysis, the Wilcoxon Rank-Sum test (stats package) was used to determine the relationship of sex, race, delivery mode with ASQ scales. The Kruskal-Wallis test (stats package) examined the associations of maternal education levels, feeding methods with ASQ scales. Data is presented as n (%) and median (min,

max). Univariate linear regression models (stats package) were used for continuous variables to analyze the relationship of pre-pregnancy BMI, maternal age, gestational age at birth with ASQ scales. Data is present as mean ± standard deviation (SD) and β (95% confidence interval, CI). Alpha diversity (Chao1, Shannon, and inverse Simpson indices) was calculated using the vegan package (Jari Oksanen et al., 2020). Multivariate linear regression models (stats package) were used to assess the associations between alpha diversity indices and ASQ scales, adjusted for feeding practice, infant sex, antibiotics use since birth, delivery mode, infant race, maternal education level, gestational age at birth, pre-pregnancy BMI, and maternal age. The Spearman correlation test (stats package) was used to test the association between alpha diversity indices and ASQ scales in each feeding group. For beta diversity, Sorensen and Bray-Curtis dissimilarities were calculated using the vegan package and ordinated using principal coordinate analysis (PCoA). Permutational multivariate analysis of variance (PERMANOVA) was performed using the vegan package to test for significant differences in beta diversity. Three clusters were determined by the partitioning around medoids (PAM) clustering algorithm using cluster package based on the Jensen-Shannon distance (JSD) of beta diversity and were assessed for the optimal number of clusters using the Calinski-Harabasz (CH) Index (Caliński & Harabasz, 1974; Kaufman, 1990). Analysis of variance (ANOVA) from stats package with Tukey's honest Significant Differences (HSD) (stats package) and Kruskal-Wallis with Dunn's test (dunn.test package) were used to examine the relationship between alpha diversity and clusters. Univariate and multivariate linear regression models adjusted by feeding practice, infant sex, antibiotics use since birth, delivery mode, infant race, maternal education level, gestational age at birth, pre-pregnancy BMI, and maternal age were performed to examine the associations between the three microbiota clusters and ASQ scales. Spearman correlation was used to test the

relationship between ASQ and the relative abundance of specific taxa. Chi-square (stats package) was used to assess the association between the three microbiota clusters and feeding methods. P-value<0.05 is significant.

2.6 Results

2.6.1 Population characteristics

A total of 64 participants were included in the final univariate analyses (Table 1). Of these, more than half of the infants were female (51.6%) and White (68.8%). Scores for each of the five ASQ scales were similar between male and female. Non-White infants had a significantly higher communication score compared to White infants (p-value=0.01). Maternal education level was associated with fine motor (p-value=0.04) and problem-solving (p-value=0.03) scores. There was a trend that maternal education level was negatively related to communication scores (p-value=0.06). Breastfed infants had significantly lower fine motor scores compared to the other three feeding groups (p-value < 0.01). Mode of delivery, pre-pregnancy BMI, and maternal age were not associated with ASQ scales. However, higher gestational age at birth tended to be associated with higher problem-solving scores (p-value=0.05).

Table 1. Population characteristics and scores on the five ASQ scales

	N=64	**Gross motor**		**Fine motor**		**Communication**		**Personal-social**		**Problem-solving**	
Categorical variable[1]	N (%) or Mean±SD	Median(min, max) or β(95% CI)	*p*-value	Median(min, max) or β(95% CI)	*p*-value	Median(min, max) or β(95% CI)	*p*-value	Median(min, max) or β(95% CI)	*p*-value	Median(min, max) or β(95% CI)	*p*-value
Infant sex											
Male	31(48.4%)	45(10, 60)	0.77	55(35, 60)	0.70	45(25, 60)	0.49	40(15, 60)	0.89	50(20, 60)	0.46
Female	33(51.6%)	45(10, 60)		60(35, 60)		50(15, 60)		40(20, 60)		55(5, 60)	
Infant race											
White	44(68.8%)	45(10, 60)	0.14	55(35, 60)	0.08	42.5(15, 60)	0.01*	40(20, 60)	0.99	52.5(5, 60)	0.98
Non-White	20(31.2%)	47.5(10, 60)		60(35, 60)		50(25, 60)		40(15, 60)		50(25, 60)	
Maternal education level											
Did not finish high school	3(4.7%)	60(15,60)	0.84	60(60, 60)	0.04*	60(40, 60)	0.06	50(15, 55)	0.41	60(60, 60)	0.03*
High school graduate or GED	11(17.2%)	45(10, 60)		55(45, 60)		50(15, 60)		45(30, 60)		55(5, 60)	
Some college	13(20.3%)	45(30, 60)		60(50, 60)		50(15, 60)		45(20, 60)		60(40, 60)	
College graduate or more	37(57.8%)	45(10, 60)		55(35, 60)		40(20, 60)		40(20, 55)		60(20, 55)	
Delivery mode											
Vaginal	39(60.9%)	45(10, 60)	0.81	60(35, 60)	0.11	50(15, 60)	0.63	40(15, 60)	0.16	50(5, 60)	0.94
C-section	25(39.1%)	45(10, 60)		55(35, 60)		45(20, 60)		45(20, 60)		55(20, 60)	
Feeding method											
Breastfeeding	9(14.06%)	45(10, 60)	0.53	45(35, 55)[a]	<0.01	35(15, 55)	0.08	45(20, 55)	0.53	50(5, 60)	0.42
Breastfeeding with Vitamin D	17(26.56%)	40(10, 60)		60(35, 60)[b]		50(25, 60)		40(20, 55)		55(30, 60)	
Partial breastfeeding	16(25%)	47.5(20, 60)		57.5(50, 60)[b]		50(15, 60)		35(20, 60)		50(20, 60)	
Formula feeding	22(34.38%)	45(15, 60)		60(40, 60)[b]		47.5(30, 60)		47.5(15, 60)		55(25, 60)	

Table 1 (cont'd)

Continuous variable[2]											
Pre-pregnancy BMI	32.07±21.98	0.04(-0.15, 0.23)	0.67	0.06(-0.02, 0.15)	0.15	0.06(-0.07, 0.19)	0.34	0.10(-0.03, 0.24)	0.14	0.08(-0.06, 0.22)	0.23
Maternal age	29.64±4.66	-0.21(-1.09, 0.66)	0.63	-0.04(-0.46, 0.37)	0.84	-0.07(-0.69, 0.55)	0.82	-0.11(-0.77, 0.55)	0.75	0.12(-0.54, 0.78)	0.72
Gestational age	39.16±1.24	2.27(-0.98, 5.51)	0.17	0.33(-1.23, 1.89)	0.67	1.90(-0.37, 4.17)	0.10	0.62(-1.87, 3.11)	0.62	2.39 (-0.01, 4.79)	0.05

[1]Categorical variable data is presented as N (%) and median (min,max). Wilcoxon Rank-Sum test was used to determine the relationship between sex, race, delivery mode, and ASQ scales, respectively. The Kruskal-Wallis test was used to examine the associations between maternal education level, feeding methods and ASQ scales. [2]Continuous variable data is presented as Mean±SD and β (95% CI). Univariate linear regression models were used to examine the relationship between continuous variables and ASQ scales. *P-value < 0.05 is significant

2.6.2 Alpha diversity and ASQ

Though the overall model was not significant, the Chao 1 index, a measure of richness, was inversely associated with gross motor score (β=-0.38, p-value=0.02), adjusted by feeding practice, infant sex, antibiotic use since birth, delivery mode, infant race, maternal education level, gestational age at birth, pre-pregnancy BMI, and maternal age (Table 2). The relationships between Chao 1 index and ASQ varied depending on the infant's diet in the past 24 hours. In univariate analyses, a higher richness of gut microbiota was significantly associated with higher communication (β=0.57, p-value < 0.01) (Figure 1C) and personal-social (β=0.45, p-value=0.04) (Figure 1D) scores among formula-fed infants.

Shannon index, a measure of richness and evenness and weighs richness more, was positively associated with problem-solving scores (β=9.87, p-value=0.04) (Table 2). Infant diet influenced the relationship between the Shannon index and ASQ scales. Among formula-fed infants, Shannon index tended to be positively associated with fine motor (β=0.37, p-value=0.09) (Figure 2B) and communication (β=0.42, p-value=0.05) (Figure 2C) scores.

There was a trend that the inverse Simpson index, a measure of a measure of richness and evenness and weighs evenness more, was positively associated with communication (β=1.61, p-value=0.07) and problem-solving (β=1.84, p-value=0.07) scores (Table 2). The relationships between the inverse Simpson index and ASQ differed by infant diet. Inverse Simpson index tended to be positively associated with communication scores (β=0.39, p-value=0.07) among formula-fed infants (Figure 3C). For partially breastfed infants, there was a trend that inverse Simpson index was positively associated with personal-social scores (β=0.46, p-value=0.08) (Figure 3D).

Table 2. The associations between alpha diversity of gut microbiota at 3 months and each of the five ASQ scale measurements at 9 months

	B **(95% CI)**	***p*-value**	**Overall adjusted R-squared**	**Overall *p*-value**
Gross motor				
Chao1	-0.38(-0.72, -0.05)	0.02*	0	0.54
Shannon	-3.19(-17.73, 11.34)	0.66	0	0.92
inverse Simpson	-0.98(-3.95, 1.99)	0.51	0	0.91
Fine motor				
Chao1	-0.005(-0.14, 0.13)	0.93	0.27	0.005*
Shannon	3.17(-2.28, 8.61)	0.25	0.29	0.003*
inverse Simpson	0.42(-3.95, 1.99)	0.46	0.28	0.004*
Communication				
Chao1	0.03(-0.18, 0.24)	0.78	0.14	0.08
Shannon	7.24(-1.47, 15.95)	0.10	0.19	0.04*
inverse Simpson	1.61(-0.16, 3.39)	0.07	0.19	0.03*
Personal-social				
Chao1	0.21(-0.04, 0.46)	0.10	0	0.57
Shannon	7.11(-3.48, 17.71)	0.18	0	0.65
inverse Simpson	1.25(-0.94, 3.43)	0.26	0	0.70
Problem-solving				
Chao1	0.06(-0.18, 0.30)	0.61	0.08	0.19
Shannon	9.87(0.40, 19.34)	0.04*	0.15	0.07
inverse Simpson	1.84(-0.12, 3.80)	0.07	0.14	0.08

[1]Multivariate linear regression models were used, adjusted by feeding practice, infant sex, antibiotic use since birth, delivery mode, infant race, maternal education level, gestational age at birth, pre-pregnancy BMI, and maternal age.
*P-value < 0.05 is significant

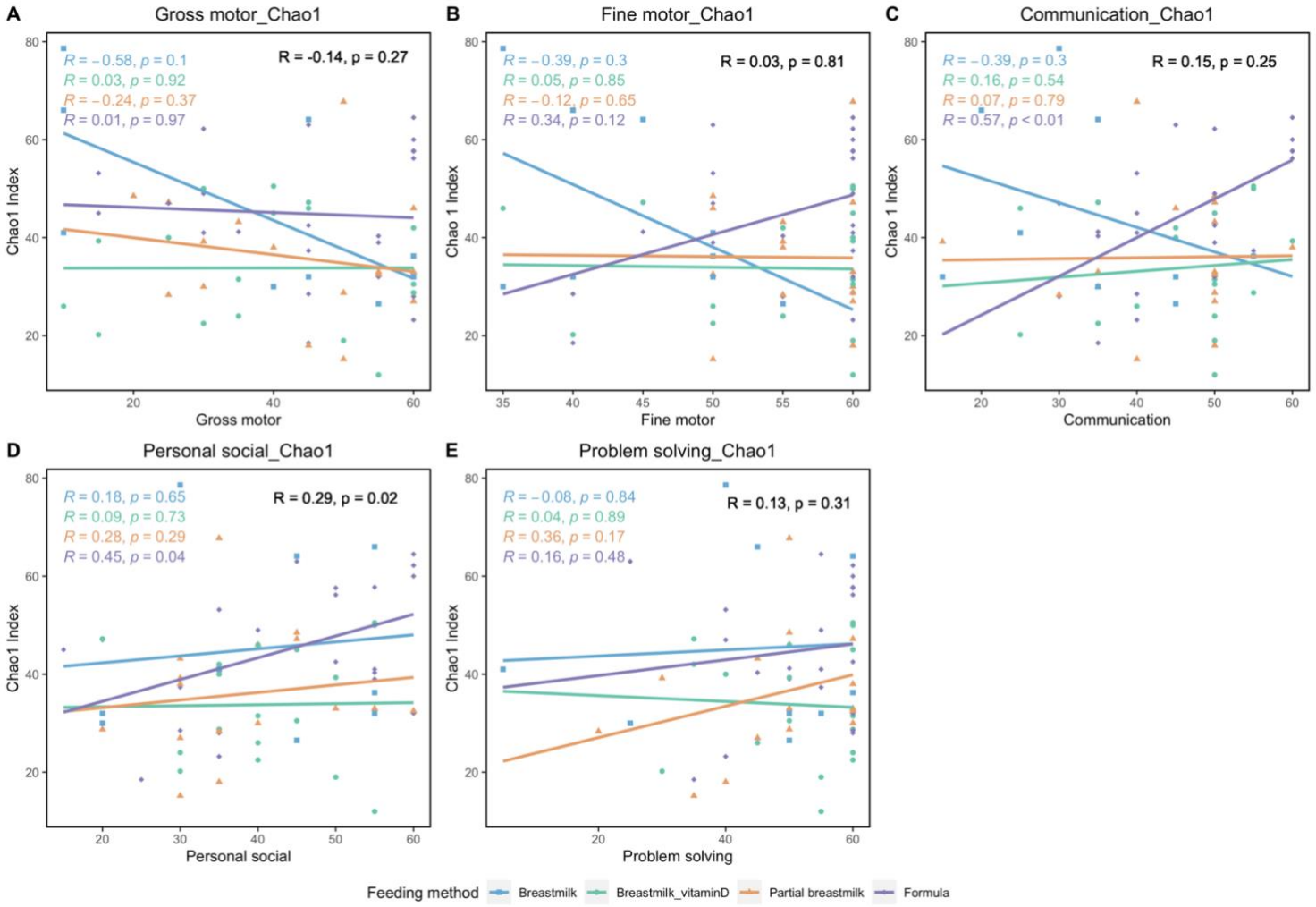

Figure 1. The associations between Chao 1 index and ASQ by different feeding methods at 3 months

Spearman correlations were used to test the association between the Chao1 index and ASQ scores for overall and individual tests. Data is presented as correlation coefficient (R) and p-value. Blue squared and regression line represent exclusively breastfed infants. Green dots and regression lines represent breastfed infants with vitamin D supplements. Orange triangles and regression lines represent partially breastfed infants. Purple diamonds and regression lines represent formula-fed infants. R and p-value in black color are the overall results. P-value < 0.05 is significant.

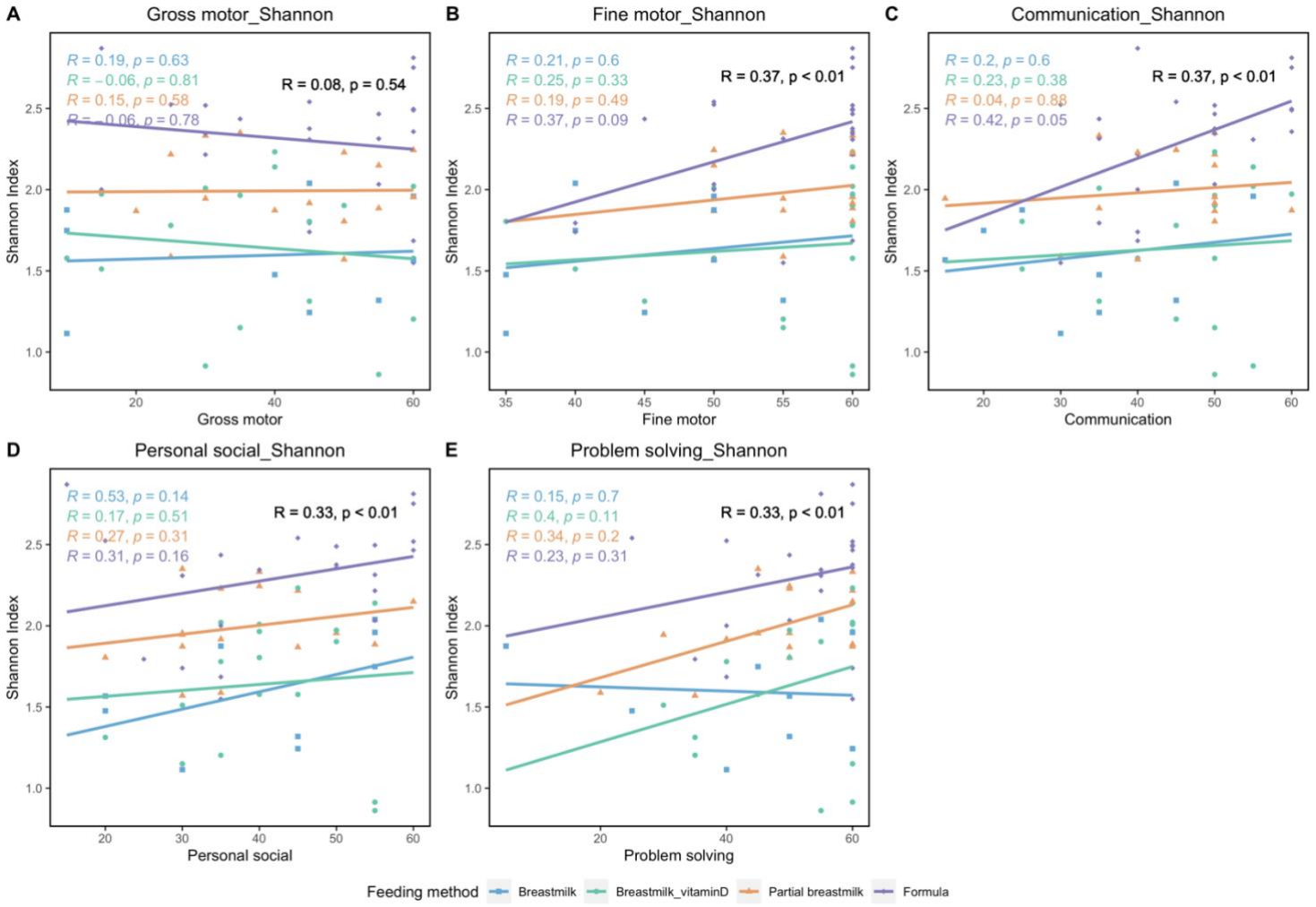

Figure 2. The associations between Shannon index and ASQ by different feeding methods at 3 months

Spearman correlations were used to test the association between the Shannon index and ASQ scores for overall and individual tests. Data is presented as correlation coefficient (R) and p-value. Blue squared and regression line represent exclusively breastfed infants. Green dots and regression lines represent breastfed infants with vitamin D supplements. Orange triangles and regression lines represent partially breastfed infants. Purple diamonds and regression lines represent formula-fed infants. R and p-value in black color are the overall results. P-value < 0.05 is significant.

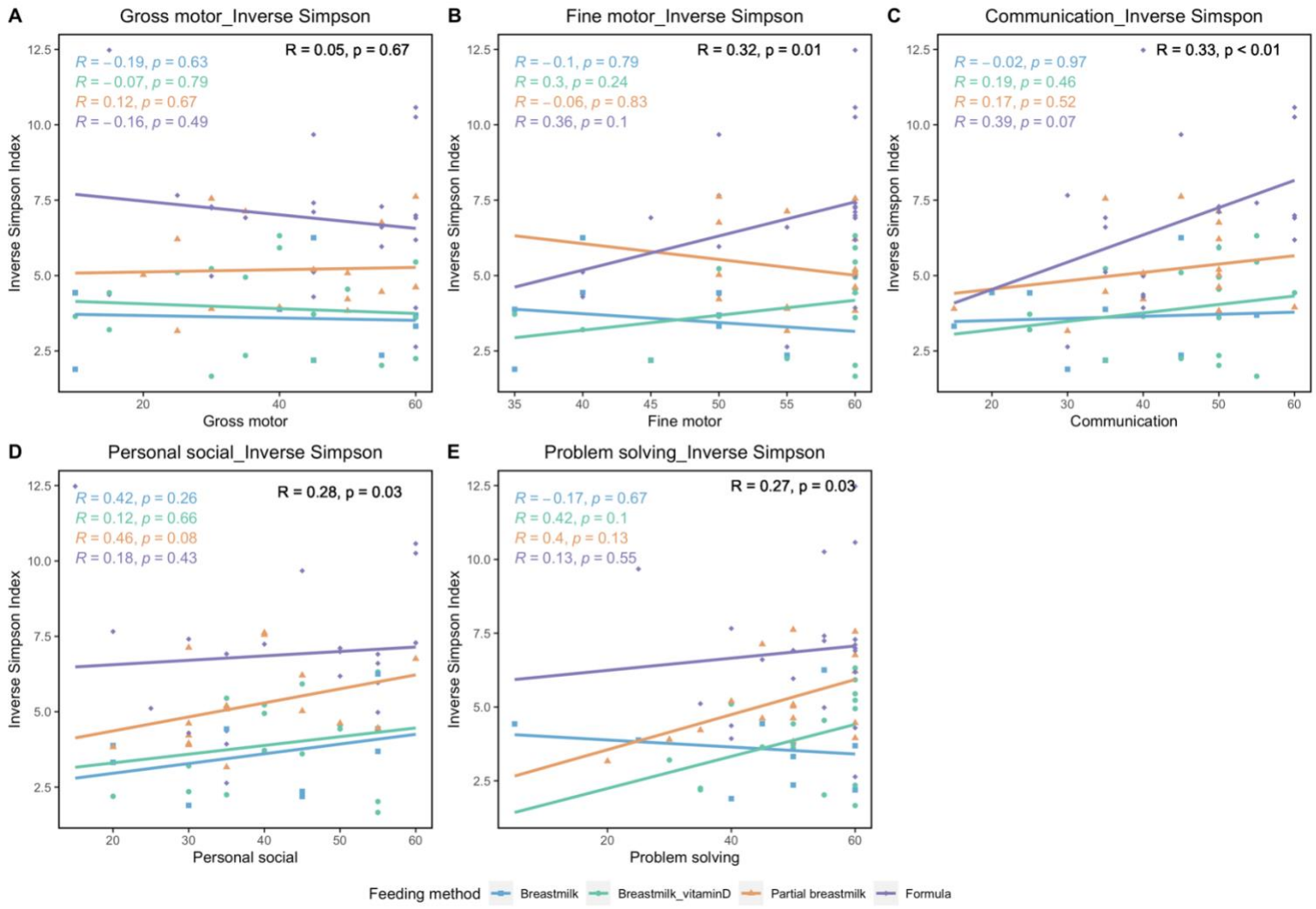

Figure 3. The associations between inverse Simpson index and ASQ by different feeding methods at 3 months

Spearman correlations were used to test the association between the inverse Simpson index and ASQ scores for overall and individual tests. Data is presented as correlation coefficient (R) and p-value. Blue squared and regression line represent exclusively breastfed infants. Green dots and regression lines represent breastfed infants with vitamin D supplements. Orange triangles and regression lines represent partially breastfed infants. Purple diamonds and regression lines represent formula-fed infants. R and p-value in black color are the overall results. P-value < 0.05 is significant.

2.6.3 Beta diversity and ASQ

For univariate analysis, the Bray-Curtis dissimilarity matrix was associated with fine motor (p-value < 0.01) and communication scores (p-value < 0.01) (Table 3, Figure 4). Bray-Curtis dissimilarity matrix was also significantly associated with fine motor (p-value < 0.01) and communication (p-value < 0.01) scores but tended to be related to problem-solving scores (p-value=0.05) after adjusting for feeding practice, infant sex, antibiotics use since birth, delivery mode, infant race, maternal education level, gestational age at birth, pre-pregnancy BMI, and maternal age (Table 3).

Table 3. The associations between beta diversity of the infant gut microbiota and each of the five ASQ scales

	Univariate analysis	**Multivariate analysis**
	***p*-value**	***p*-value**
Gross Motor		
Sorensen	0.47	0.35
Bray-Curtis	0.85	0.74
Fine Motor		
Sorensen	0.24	0.16
Bray-Curtis	<0.01*	<0.01*
Communication		
Sorensen	0.26	0.18
Bray-Curtis	0.01*	<0.01*
Personal-social		
Sorensen	0.20	0.13
Bray-Curtis	0.25	0.14
Problem-solving		
Sorensen	0.35	0.23
Bray-Curtis	0.11	0.05

PERMANOVA was performed. Multivariate analysis was adjusted by feeding practice, infant sex, antibiotic use since birth, delivery mode, infant race, maternal education level, gestational age at birth, pre-pregnancy BMI, and maternal age. *P-value < 0.05 is significant.

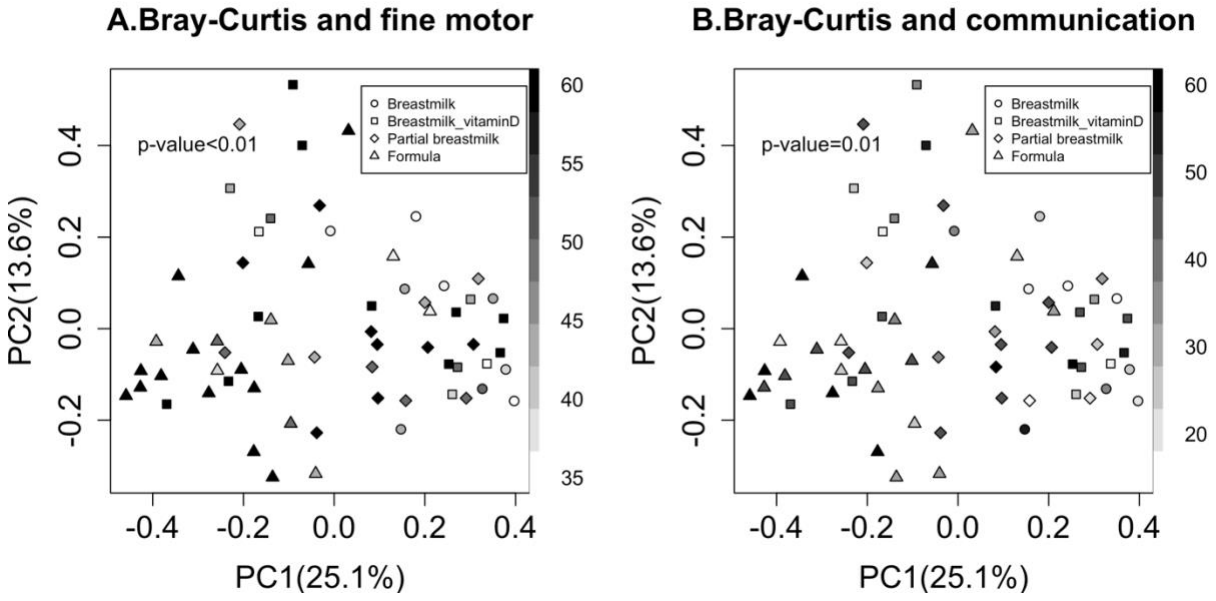

Figure 4. The significant associations between Bray-Curtis matrix and ASQ scales

PERMANOVA was performed to examine the relationships between beta diversity and ASQ scales. P-value < 0.05 is significant.

2.6.4 Cluster analysis

Upon using the PAM clustering algorithm and assessing using CH score to cluster infants into groups by their gut microbiota composition, three clusters emerged. The bacterial composition of each infant gut microbiota within each of the three clusters is shown in Figure 5. 31.25% of the infants fell into Cluster 1, 35.94% fell into Cluster 2, and 32.81% were clustered into Cluster 3. When only considering the top 5 most abundant taxa, Cluster 1 is dominated by *Lachnospiraceae* unclassified, Cluster 2 is dominated by *Bifidobacterium*, and *Bacteroides* is the most abundant taxa in Cluster 3 (Figure 6).

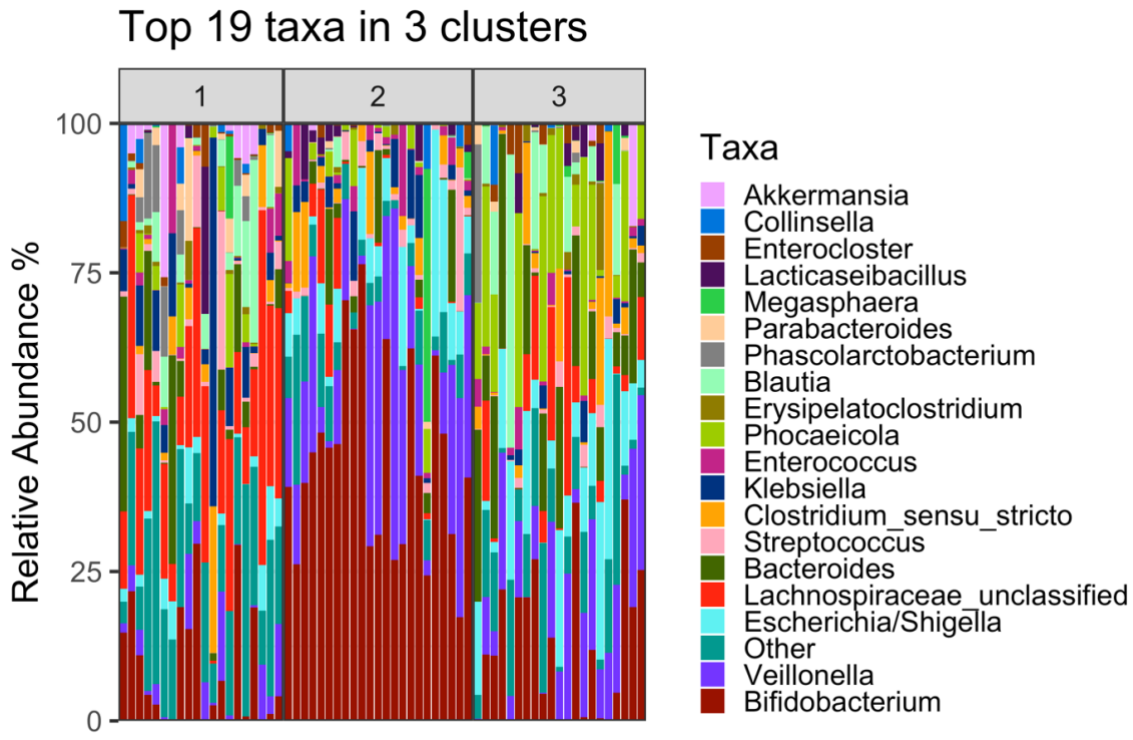

Figure 5. The gut microbiota composition of infant stool samples organized by cluster

Three clusters were determined by the PAM clustering algorithm assessed by the CH score based on the JSD of beta diversity.

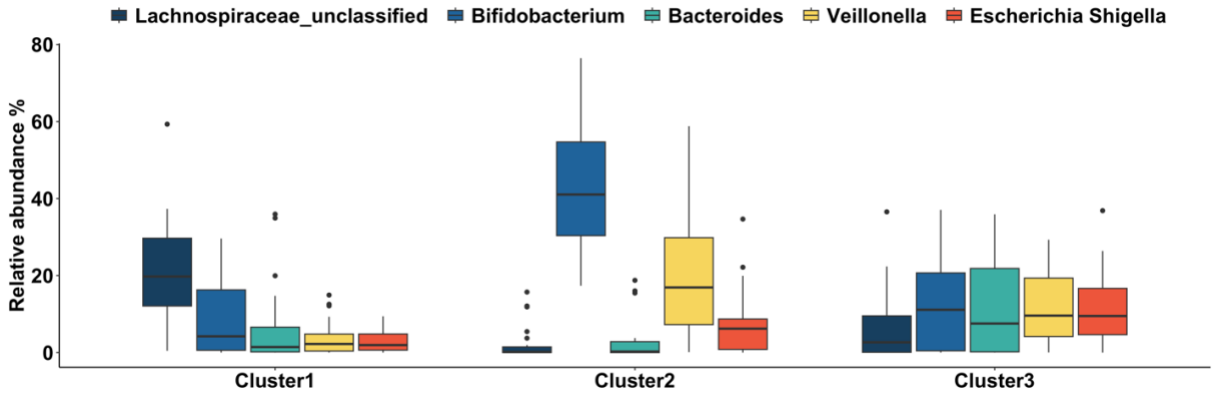

Figure 6. The composition of the top five overall most abundant taxa presented by cluster

The average abundance of each taxon was calculated. Only the top five most abundant taxa were selected and plotted.

The richness (Chao 1 index, p-value=0.11) of the 3-month infant gut microbiota was similar across the three clusters (Figure 7A). Shannon (p-value < 0.01) and inverse Simpson (p-

value < 0.01) indices differed by clusters (Figure 8B, 7C). Clusters 1 and 3 had similar gut microbiota richness and evenness as measured by Shannon and inverse Simpson indices (Figure 7B, 7C). Cluster 2 had significantly lower richness and was less even than Clusters 1 and 3 (Figure 7B, 7C). As expected, the three clusters had significantly different gut microbial membership and composition when measuring beta diversity (Figure 8).

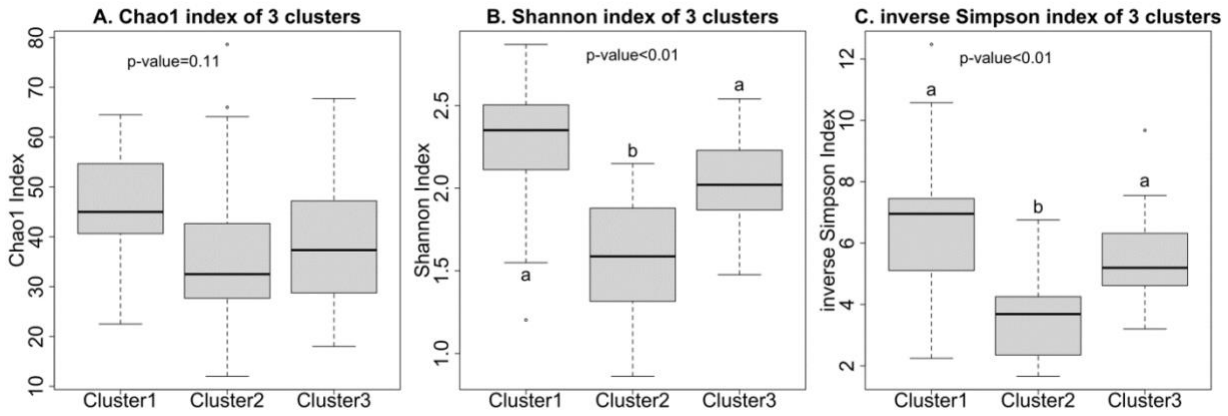

Figure 7. Shannon and inverse Simpson indices of gut microbial alpha diversity differs across the three clusters

Shapiro–Wilk test was used to test data normality. ANOVA tests were used to examine the relationships between Chao1 (A) and Shannon (B) indices and clusters. The relationship between inverse Simpson (C) and clusters was tested by the Kruskal-Wallis test. Tukey's HSD and Dunn's tests were performed for pairwise comparison. The median with the min and max was plotted. Different letters indicate significant differences in pairwise comparisons. P-value < 0.05 is significant.

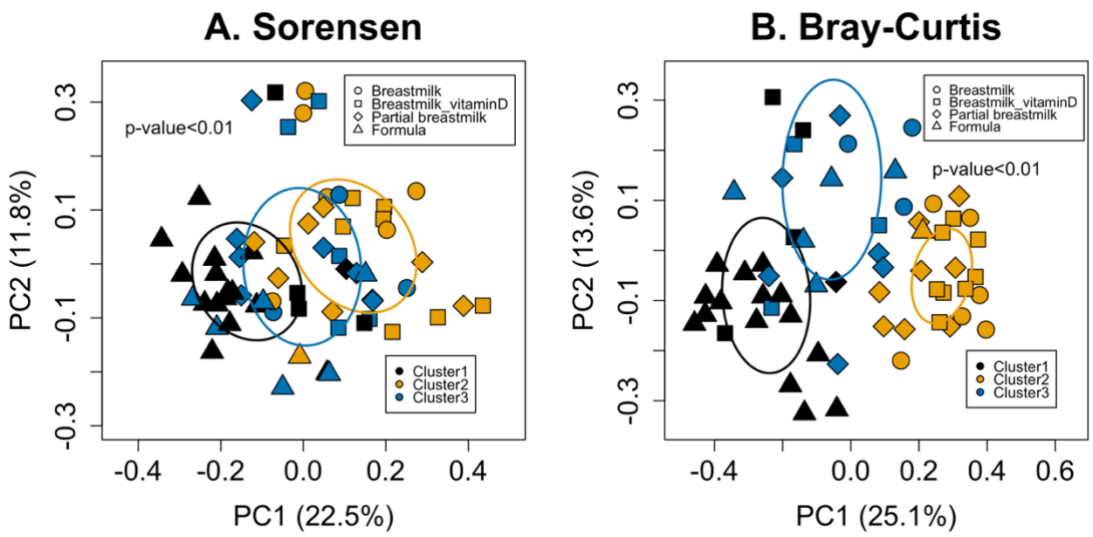

Figure 8. The gut microbiota beta diversity is differed by cluster

PERMANOVA was performed to examine the relationships between beta diversity and clusters. P-value < 0.05 is significant.

Infants whose gut microbiota were in Cluster 2 (*Bifidobacterium* dominated) had lower fine motor scores (β=-4.98, p-value=0.03) compared to infants whose gut microbiota were in Cluster 1 (*Lachnospiraceae* unclassified dominated) when conducting univariate analyses (Table 4). In both univariate and multivariate models, infants whose gut microbiota were in Cluster 3 (*Bacteroides* dominated) had lower problem-solving scores (univariate analysis: β=-9.01, p-value=0.02; multivariate analysis: β=-10.08, p-value=0.02) compared to infants whose gut microbiotas were in Cluster 1 (Table 4). There was a trend that fine motor was negatively associated with the relative abundance of *Bifidobacterium* (p-value=0.08) but positively associated with the relative abundance of *Lachnospiraceae* unclassified (p-value=0.09) (Figure 9A, 9B). Problem-solving tended to be positively associated with the relative abundance of *Lachnospiraceae* unclassified (p-value=0.06) (Figure 9D). The feeding method in the past 24 hours (Figure 10A) and the past week (Figure 10B) were significantly associated with gut microbiota clusters (p-values < 0.01). Infants whose gut microbiotas fell into Cluster 1 were more likely to have been fed formula in the past day and less likely to have been fed any human

milk in the past day than infants whose gut microbiotas fell into clusters 2 or 3 (Figure 10A). The gut microbiota of infants was clustered to Cluster 2 when infants were mostly fed 100% breastmilk or 80% breastmilk in the past week (Figure 10B)

Table 4. The associations between three clusters and ASQ scales

		Univariate model		**Multivariate model[1]**			
		β (95% CI)	***p*-value**	**β (95% CI)**	***p*-value**	**Overall adjusted R-squared**	**Overall *p*-value**
Gross motor	**Cluster 1**	Reference					
	Cluster 2	-1.91(-11.88, 8.05)	0.70	0.83(-14.57, 16.23)	0.91	0	0.96
	Cluster 3	-2.05(-12.23, 8.13)	0.69	-0.54(-13.76, 12.69)	0.94		
Fine motor	**Cluster 1**	Reference					
	Cluster 2	-4.98(-9.52, -0.43)	0.03*	-4.37(-10.06, 1.32)	0.13	0.30	0.004*
	Cluster 3	-2.69(-7.34, 1.95)	0.25	-3.29(-8.18, 1.60)	0.18		
Communication	**Cluster 1**	Reference					
	Cluster 2	-5.10(-11.98, 1.78)	0.14	-3.69(-13.03, 5.68)	0.43	0.14	0.08
	Cluster 3	-0.75(-7.78, 6.28)	0.83	0.02(-8.02, 8.06)	0.997		
Personal-social	**Cluster 1**	Reference					
	Cluster 2	-4.65(-12.10, 2.79)	0.22	-1.82(-13.22, 9.58)	0.75	0	0.85
	Cluster 3	-4(-11.61, 3.61)	0.30	-1.47(-11.26, 8.33)	0.77		
Problem-solving	**Cluster 1**	Reference					
	Cluster 2	-5.12(-12.26, 2.02)	0.16	-6.23(-16.10, 3.63)	0.21	0.16	0.06
	Cluster 3	-9.01(-16.31, -1.72)	0.02*	-10.08(-18.56, -1.61)	0.02*		

[1]Multivariate linear regression models were used, adjusted by feeding practice, infant sex, antibiotic use since birth, delivery mode, infant race, maternal education level, gestational age at birth, pre-pregnancy BMI, and maternal age.
*P-value<0.05 is significant

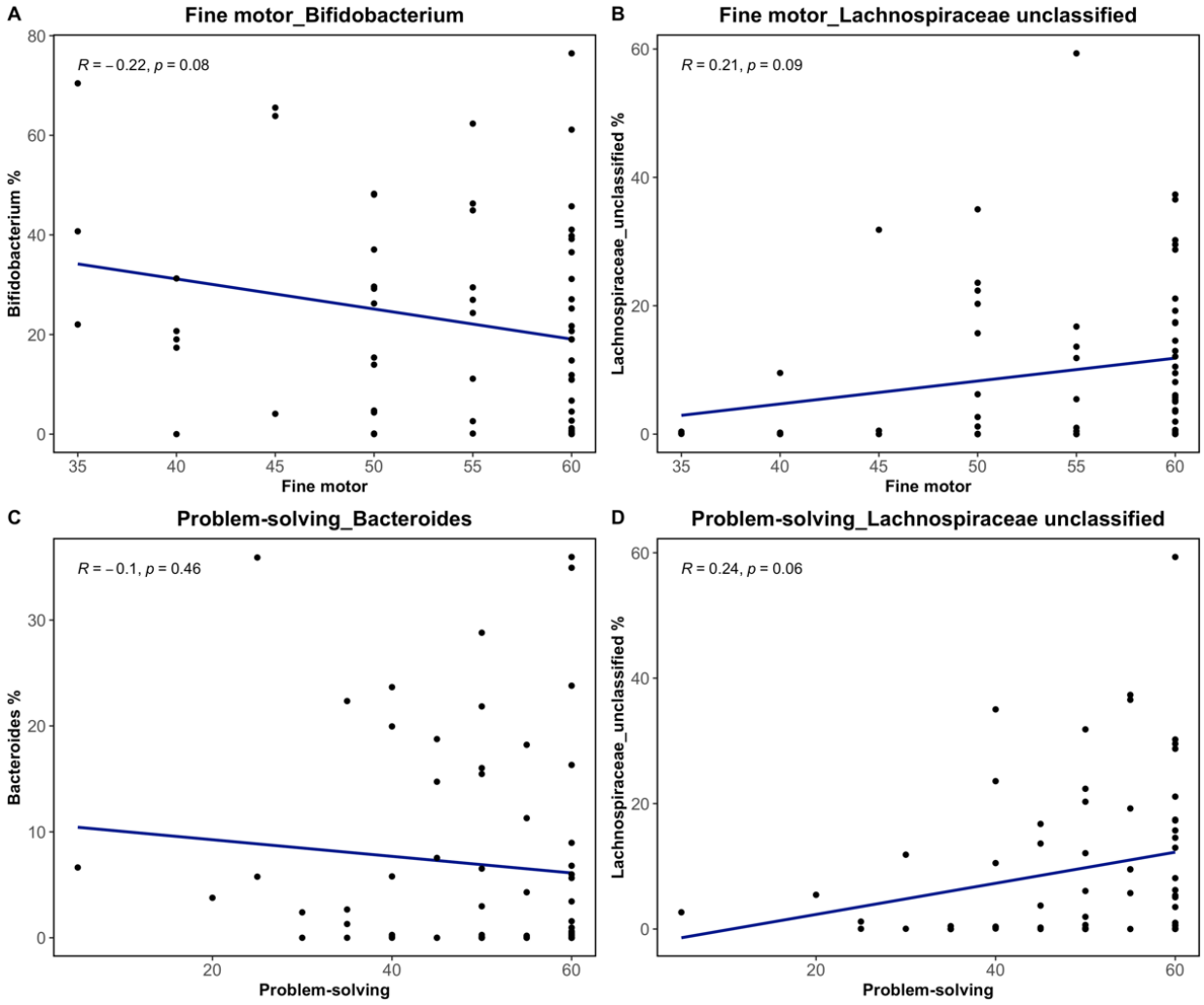

Figure 9. The relationships between ASQ and relative abundance of specific taxa

Spearman correlation was used. *P-value < 0.05 is significant.

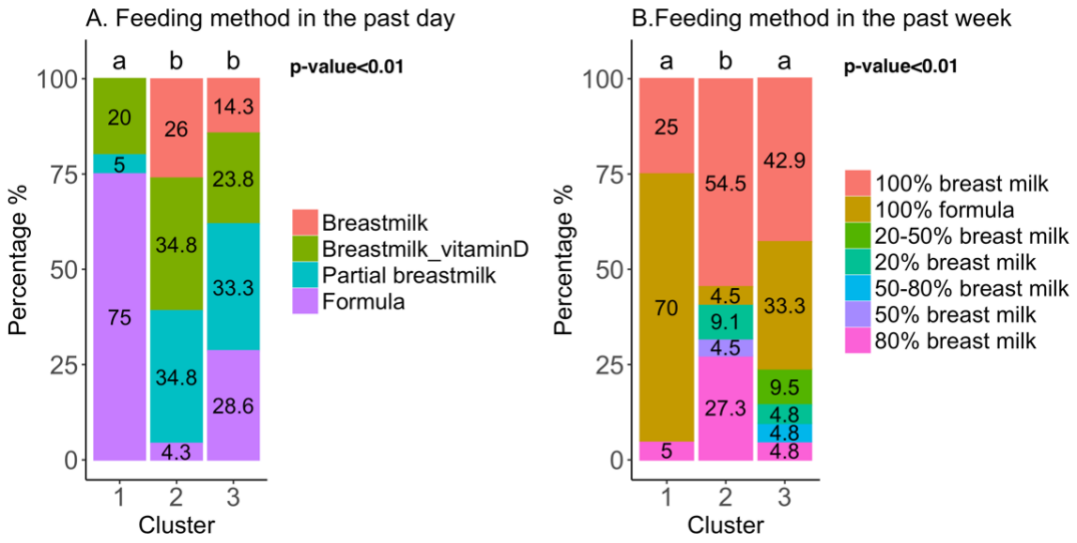

Figure 10. The frequency of feeding methods in the past 24 hours and past week at 3 months of age in each cluster

Data is presented as a percentage of infants within that cluster. Bars with different colors represent feeding groups. Chi-square was used to test if the proportion of infants in the various feeding groups differed across clusters. *P-value < 0.05 is significant

2.7 Discussion

We investigated the association between infant gut microbiota and later life neurodevelopment measured by ASQ-3. Several animal studies have shown that gut microbiota was related to brain development (Clarke et al., 2013; Pan et al., 2019). However, this connection has not been elucidated in human populations, especially among infants. Our results suggested infant gut microbiota at 3 months of age might be potentially associated with gross motor, fine motor, communication, and problem-solving skills later in life after adjustment for feeding practice, infant sex, antibiotic use since birth, delivery mode, infant race, maternal education level, gestational age at birth, pre-pregnancy BMI, and maternal age.

The current study found that the richness and evenness (Shannon index) of the gut microbiota were positively associated with problem-solving scores. Better-developed gut microbiota in infancy has a higher level of biological diversity, and decreased microbial diversity

is associated with adverse health outcomes among preterm infants (Jia et al., 2022; Warner et al., 2016). Alpha diversity of gut microbiota from full-term born infants was significantly higher than preterm born infants (< 32 weeks of gestation) at day 14 after birth (Jia et al., 2022). Necrotizing enterocolitis in very low birth weight infants might be attributed to the lack of gut microbial diversity (Claud & Walker, 2001). A more diverse gut microbiota in the first week of life was related to a reduced risk of eczema in infants at 12 months of age (Ismail et al., 2012). Moreover, the low diversity of gut microbiota during the first month of life was associated with asthma in 7-year-old children (Abrahamsson et al., 2014). However, the evidence of the relationship between gut microbial diversity and health status later in life is controversial. Carlson et al. reported that alpha diversity of the gut microbiota of 1-year-old children was negatively associated with Mullen Scales of Early Learning Composite (MSEL ELC), expressive language, and visual reception subscale scores at the age of 2 years (Vaher et al., 2022). A positive association was found between Chao 1 index and function connectivity between the supplementary motor area (SMA) and the inferior parietal lobule (IPL) in 1-year-old infants' brains. SMA-IPL connectivity was negatively related to the MSEL ELC at 2 years of age (Gao et al., 2019). A possible reason for these discrepancies is that the infant gut microbiota is susceptible to modulation by external factors such as infant feeding methods (O'Sullivan et al., 2015). In the U.S., 40% of mothers introduce solid foods to infants before 4 months of age and start feeding infants with solids at 12 weeks (Clayton et al., 2013). Increased alpha diversity of the infant gut microbiota was found when complementary foods were introduced to infants from 4 months until 12 months of age, where the shift occurred more significantly between 4 and 9 months of age (McKeen et al., 2022). Aside from diet, mode of delivery, antibiotic exposure, and maternal pre-pregnancy BMI also impact the development of infant gut microbiota (Ainonen et

al., 2022; Biasucci et al., 2010; Stanislawski et al., 2017). These factors contribute to three different phases of microbiome progression: a developmental phase (3-14 months of age), a transitional phase (15-30 months of age), and a stable phase (31-46 months of age) (Stewart et al., 2018). Shannon diversity index changed significantly during the developmental and transitional phases but remained stable during the stable phase (Stewart et al., 2018). Therefore, this evidence reinforces the notion that the directionality and strength of the associations between alpha diversity and health outcomes are different between ages due to exposure to external factors.

In addition to the alpha diversity, this study also demonstrated that Bray-Curtis dissimilarity matrix (gut microbial composition) was associated with fine motor. This was similar to other studies. Acuña et al. found that fine motor skills were strongly associated with gut microbial composition using weighted Unifrac metrics, which assesses membership and composition, measured by the Bayley-III questionnaire when the infants were at 18 months of age (Acuña et al., 2021).

The abundance of specific gut microbes in infancy prime influences the neurodevelopmental outcomes. In the univariate analysis, we observed that fine motor was negatively associated with Cluster 2 (*Bifidobacterium*-dominated) compared to Cluster 1 (*Lachnospiraceae* unclassified-dominated). Higher fine motor scores tended to be associated with a lower relative abundance of *Bifidobacterium*. It is somewhat surprising that *Bifidobacterium* was found to be more abundant in the above-median fine motor activity group compared to the below-median group in healthy full-term infants at 18 months of age (Acuña et al., 2021). However, another study reported that the relative abundance of Bifidobacteria wasn't associated with fine motor scores when infants aged 17-18 weeks (Wu et al., 2021). Our study

reported that problem-solving was negatively associated with Cluster 3 (*Bacteroides*-dominated) compared to Cluster 1(*Lachnospiraceae* unclassified-dominated) when conducting univariate and multivariate analyses. The problem-solving was negatively associated with the relative abundance of *Bacteroides*, but it was not statistically significant. This association varies due to the different sex, ages, and populations studied. The higher abundance of genus *Bacteroides* in gut microbiota was associated with better cognitive and language scores at age 2, predominantly among males (Tamana et al., 2021). An increased abundance of *Bacteroides* during the first year of life positively impacted communication development later in childhood (Vaher et al., 2022). Conversely, other studies reported that an increased abundance of *Bacteroides* may reflect delayed maturation of the gut microbiome in children, which further supports the adverse outcomes of *Bacteroides* on infant neurodevelopment (Carlson et al., 2018). The *Bacteroides*-dominated coabundance grouping of infants at ages 3 to 6 months was associated with poorer fine motor skills at age 3 years (Sordillo et al., 2019). The gut microbiota of infants with a *Bacteroides*-dominant community displayed poorer fine motor performance than other enterotypes (Acuña et al., 2021).

The present study has several strengths. We demonstrated prospective associations between the early-life infant gut microbiota and neurodevelopmental outcomes later in life in a longitudinal cohort of typically developing infants. In addition, we excluded pre-term born infants who typically have delayed neurodevelopment compared to full-term infants. There are several limitations in this study. The stool samples were kept and shipped at room temperature for the day, which might influence the gut microbiota composition. However, the stool collection tube used in our lab has preservatives that can retain the gut microbiota composition for up to two weeks at room temperature. ASQ-3 is a parent-reported measurement. Thus, there might be

some biases resulting from parental responses. For example, parents with low socioeconomic status have been shown to over- or underestimate their children's performance on the questions (Feldman et al., 2000). Some parents might be prone to social desirability bias (Bourdeaudhuij & Oost, 2000).

2.8 Conclusion

The current study suggests an association between infant gut microbiota composition at age 3 months and gross motor, fine motor, communication, and problem-solving skills at age 9 months. Our findings provide insights into the relationship between early-life gut microbiota alteration and neurodevelopmental outcomes through the gut-microbiota-brain axis.

CHAPTER 3: THE MEDIATING ROLE OF INFANT GUT MICROBIOTA AT THREE MONTHS OF AGE IN THE ASSOCIATIONS BETWEEN INFANT FEEDING METHODS AT THREE MONTHS OF AGE AND INFANT NEURODEVELOPMENT AT NINE MONTHS OF AGE

3.1 Abstract

Early life is crucial for brain development and the establishment of cognitive abilities. The gut microbiota that colonizes the gastrointestinal tract also develops rapidly after birth in response to external factors. The microbial colonization of the gastrointestinal tract appears to happen in parallel and interactively with brain development. The infant gut microbial composition is linked to the infant diet. Breastfeeding in infancy might improve long-term neurodevelopmental outcomes in childhood. However, it's unclear whether gut microbiota can mediate the association between infant feeding methods and neurodevelopmental outcomes. Aim 1 demonstrated a relationship between infant gut microbiota at 3 months of age and neurodevelopmental outcomes at 9 months of age. Therefore, this study aimed to identify the mediating role of infant gut microbiota at 3 months of age in the association between infant diet and infant neurodevelopment. DNA was extracted from 64 stool samples, 16S rRNA libraries were prepared, and libraries were sequenced by Illumina MiSeq. Sequences were processed using mothur, and data were analyzed in R. Infant diet information was provided by parental report at three and nine months of age. Neurodevelopment was assessed by parental completion of the Ages and Stages Questionnaire-3 (ASQ-3) when infants were 9 months old. Breastfed infants with vitamin D supplementation (p-value<0.01), partially breastfed infants (p-value<0.01), and formula-fed (p-value<0.01) infants at 3 months had higher fine-motor scores at 9 months than exclusively breastfed infants that were not supplemented. Infant feeding method was associated with infant gut microbial composition as measured by Bray-Curtis dissimilarity matrix. Bray-Curtis distance matrix of beta diversity mediated the associations between feeding method and fine-motor scores univariately (p-value=0.04). Our results support the potential mediating role of early-life gut microbiota in the association between infant feeding method and infant neurodevelopmental outcomes in late infancy.

3.2 Key words

vitamin D, breastfeeding, human milk, formula feeding, infant gut microbiota, mediation, neurodevelopment, problem-solving, Ages and Sages questionnaire

3.3 Introduction

Breastfeeding is a pathway to constantly transfer essential nutrients in appropriate amounts from mothers to infants (Hinde & German, 2012). In addition, breastfeeding has a more profound impact on infants' cognitive and behavioral development and mental health than simple nutrient transfer alone (Raju, 2011). In fact, a longer duration of exclusive breastfeeding was positively associated with memory performance, early language development, and motor skills at 14 months (Guxens et al., 2011) and 18 months of age (Leventakou et al., 2015) as measured by the Bayley Scales of Infant Development. Similarly, communication and global motor skills were more delayed in preschoolers who were breastfed for only 3 months compared to those with 6- and 12-month breastfeeding duration when using the ASQ-3 (Saliaj, 2015). Deoni et al. demonstrated that infant feeding practices influenced cognitive ability and white-matter development in children from 10 months through 4 years of age using magnetic resonance imaging (MRI) (Deoni et al., 2013). Breastfed infants had significantly better mental and motor development at 18 months than formula-fed infants as measured by Bayley Scales of Infant Development II (Morley et al., 2004). Exclusively breastfed infants had higher cognitive scores than formula-fed infants at 12 months, as assessed by the Bayley Scales of Infant and Toddler Development, Third Edition (Timby et al., 2014). Therefore, breastfeeding in early infancy might positively impact the infant neurodevelopment in late infancy and later life.

Infant feeding practices significantly influence the colonization and maturation of the infant gut microbiome (O'Sullivan et al., 2015). Human milk oligosaccharides (HMOs) are a

prominent constituent of human breast milk, and following partial digestion in the small intestine, then predominantly reach the colon. Once in the colon, they are metabolized by *Bifidobacterium* to produce short-chain fatty acids and other functional metabolites that are beneficial to our body (Le Huërou-Luron et al., 2010; Marcobal et al., 2010). Compared to partially or non-breastfed, exclusively breastfed infants exhibited reduced gut bacterial diversity, an increased prevalence of *Bifidobacterium*, and a decreased abundance of *Lachnospiraceae*. (Baumann-Dudenhoeffer et al., 2018; Forbes et al., 2018; Sugino, Ma, Kerver, et al., 2021). Infants exclusively fed with formula showed a higher diversity of gut microbiota with decreased prevalence of Bifidobacterium species and an increased prevalence of Clostridium species and Enterobacteriaceae species. This may be attributed to the absence of human milk oligosaccharides (HMOs) and higher protein content in infant formula, which contribute to the modulation of gut microbiota (Bäckhed et al., 2015; Benno et al., 1984; Penders et al., 2007).Thus, infant gut microbiota composition is tightly linked to the infant diet.

There has been limited evidence to determine if gut microbiota in early infancy mediates the association between infant feeding and neurodevelopment later in infancy. The prior aim has demonstrated that infant gut microbiota was associated with infant neurodevelopment later in life. Therefore, this study aimed to investigate whether there was an association between infant feeding practice at 3 months of age and infant neurodevelopment at 9 months of age and if gut microbiota at 3 months of age mediates this association.

3.4 Materials and methods

3.4.1 Study population

The study population was described in aim 1. For aim 2, data and samples from 64 Michigan Archive for Research on Child Health (MARCH) participants were used in the analyses. Mothers'

education level, mother's height, pre-pregnancy weight, and maternal age was collected via MARCH Prenatal 1 Survey. The birth certificate included the infant sex, mode of delivery, and estimated weeks of gestation. Infant race was obtained from MARCH 3-month questionnaire. The sample collection form, completed at the time of fecal sample collection and when the infants were approximately 3 months of age, included information about the antibiotics use since birth, infant diet in the past 24 hours, and the infant diet in the past week prior to fecal collection. Exclusive breastfeeding duration, any breastfeeding duration, and ASQ-3 information were obtained from the MARCH 9-Month Survey. Infants with gestational age less than 37 weeks were excluded from the analyses. Infants who had missing information were also removed. The Michigan State University Human Research Protection Program approved the study (IRB# 16-1429).

3.4.2 Classification of feeding methods

Infant feeding method in the past 24 hours prior to fecal collection was categorized into exclusive breastfeeding, partial breastfeeding, and formula feeding, which was the FED_PRAC_NEW variable in R codes. Infant feeding methods stratified with vitamin D supplementation included exclusive breastfeeding, exclusive breastfeeding with vitamin D supplementation, partial breastfeeding, and formula feeding in the past 24 hours prior to fecal collection, which was the FED_PRAC_LIGHT_NEW variable in R codes. Exclusive breastfeeding duration in days until 9 months of age was calculated. The end day was the last day the infants were fed exclusive breastmilk or the first day they were fed formula or complementary food. Any breastfeeding duration until 9 months includes the breastfeeding and formula feeding days, and the end day was the last day the infant stopped breastmilk feeding.

3.4.3 Ages and Stages Questionnaire

When the infants were approximately 9 months old, parents completed the ASQ-3

(Squires J, 2009) during a phone interview as part of the MARCH 9-Month Survey. The ASQ-3 is a parent-completed screening tool that pinpoints developmental progress in children. The ASQ-3 comprises 5 areas: communication, gross motor, fine motor, problem-solving, and personal-social for children from 1-66 months. Scores for each area fall between 0 and 60 points. Parents indicate for each item "yes" if the child performs the item and scores 10 points, "sometimes" indicating an occasional or emerging skill and the child scores 5 points, or "not yet" if the child doesn't perform the behavior and scores 0 points.

3.4.4 Stool sample collection

Sample collection was as described in Aim 1.

3.4.5 Laboratory Procedures

3.4.5.1 DNA Extraction and 16S rRNA Gene Amplification

DNA extraction, 16S rRNA gene amplification, and sequencing were carried out on stool samples as previously described in Aim 1.

3.4.5.2 Processing and analysis of sequence data

The processing of sequencing data was also described in Aim 1.

3.5 Statistical analysis

All data were analyzed using R (version 4.2.2). Data normality was tested using Shapiro–Wilk test (stats package). Kruskal-Wallis (stats package) with post-hoc Dunn's test (dunn.test package) was used to analyze the relationships between (1) infant feeding method (exclusive breastfeeding, partial breastfeeding, formula feeding), (2) infant feeding method stratified with vitamin D supplementation (exclusive breastfeeding, exclusive breastfeeding with vitamin D

supplementation, partial breastfeeding, formula feeding) and ASQ scales. Univariate and multivariate linear regression models (stats package) adjusted by antibiotics use since birth, infant sex, delivery mode, infant race, maternal education level, gestational age at birth, pre-pregnancy BMI, and maternal age were used to assess the associations between different feeding variables and ASQ scales. Chi-square (stats package) and Kruskal-Wallis tests were used to determine the associations between categorical and continuous variables of population characteristics and infant feeding methods stratified with vitamin D supplementation. Alpha diversity (Chao1, inverse Simpson, and Shannon indices) was calculated using the vegan package in R (Jari Oksanen et al., 2020). Analysis of variance (ANOVA) tests were used to examine the relationships between Chao1 and Shannon indices and feeding methods. The relationship between inverse Simpson and feeding methods was tested by the Kruskal-Wallis test. For beta diversity, Sorensen and Bray-Curtis dissimilarities were calculated using the vegan package in R and ordinated using principal coordinate analysis (PCoA). Permutational multivariate analysis of variance (PERMANOVA) was performed using the vegan package in R to test for significant differences in beta diversity between feeding methods. Simple mediation analysis was completed when the mediator was any of the alpha diversity indices (Shannon and inverse Simpson) using the MeMoBootR package (Buchanan, 2018), adjusted by infant sex, delivery mode, infant race, maternal education level, gestational age at birth, pre-pregnancy BMI, and maternal age. Mediation analysis was conducted using the LDM package to test the mediation effect of the Bray-Curtis dissimilarly matrix (Hu & Satten, 2020).

3.6 Results

3.6.1 The association between feeding methods and ASQ scales

Infant feeding methods (breastfeeding, partial breastfeeding, and formula feeding) in the

past 24 hours before sample collection was not associated with the score for any of the ASQ scales at 9 months of age (Table 5). However, partially breastfed infants had higher fine motor skill scores compared to exclusively breastfed infants when conducting univariate (β=5.1, p-value=0.03), but not multivariate (p-value=0.17), linear regression analysis (Table 6).

When stratified by vitamin D supplementation, infant feeding method was associated with fine motor skills (p-value < 0.01), where exclusively breastfed infants had lower fine motor scores compared to infants in the other three feeding groups (Table 5). Additionally, maternal education level (p-value=0.049) and mode of delivery (p-value=0.01) was associated with infant feeding method (Table 7). Maternal pre-pregnancy BMI (p-value=0.06) tended to be associated with infant feeding method when exclusively breastfed infants supplemented with vitamin D were included as a group distinct from the other exclusively breastfed infants (Table 7). The univariate linear regression models demonstrated that breastfeeding with vitamin D supplementation (fine motor: β=10.26, p-value < 0.01, communication: β=11.7, p-value=0.01), partial breastfeeding (fine motor: β=11.81, p-value < 0.01, communication: β=9.86, p-value=0.03), and formula feeding (fine motor: β=10.78, p-value < 0.01, communication: β=12.25, p-value < 0.01) were positively associated with fine motor and communication scores compared to exclusive breastfeeding (Table 8). Formula feeding tended to be positively associated with problem-solving scores compared to exclusive breastfeeding (β=8.71, p-value=0.07) (Table 8). Multivariate analyses adjusted by antibiotics use since birth, delivery mode, infant race, maternal education level, gestational age at birth, pre-pregnancy BMI, and maternal age were conducted. A significant association remained between breastfeeding with vitamin D intake compared to exclusive breastfeeding (β=10.27, p-value < 0.01), partial feeding compared to exclusive breastfeeding (β=10.19, p-value < 0.01), and formula feeding compared to exclusive breastfeeding (β=8.25, p-value=0.02) and fine motor scores. Breastfeeding with vitamin D intake was also positively associated with communication scores compared to exclusive breastfeeding (β=9.52, p-value=0.04) after controlling covariates (Table 8). Neither the duration of exclusively breastfeeding the infant (Table 9) nor the duration of any exposure to human milk (Table 10)

up to 9 months of age was associated with scores for any of the five ASQ scales. However, for each additional day of exposure to human milk, communication scores tended to decrease by 0.04 points (p-value=0.05) (Table 10).

Table 5. The associations between infant feeding methods of infants at 3 months of age and ASQ scores at 9 months of age

	N=64	**Gross motor**		**Fine motor**		**Communication**		**Personal-social**		**Problem-solving**	
	N (%)	Median(min, max)	*p*-value	Median(min ,max)	*p*-value	Median(min, max)	*p*-value	Median(min ,max)	*p*-value	Median(min ,max)	*p*-value
Feeding method											
Breastfeeding	26(40.6%)	40(10, 60)	0.33	52.5(35, 60)	0.15	45(15, 60)	0.53	40(20, 55)	0.34	50(5, 60)	0.44
Partial breastfeeding	16(25%)	47.5(20, 60)		57.5(50, 60)		50(15, 60)		35(20, 60)		50(20, 60)	
Formula	22(34.4%)	45(15, 60)		60(40, 60)		47.5(30, 60)		47.5(15, 60)		55(25, 60)	
Feeding method by vitamin D intake											
Breastfeeding	9(14.06%)	45(10, 60)	0.53	45(35, 55)[a]	<0.01 *	35(15, 55)	0.08	45(20, 55)	0.53	50(5, 60)	0.42
Breastfeeding with Vitamin D	17(26.56%)	40(10, 60)		60(35, 60)[b]		50(25, 60)		40(20, 55)		55(30, 60)	
Partial breastfeeding	16(25%)	47.5(20, 60)		57.5(50, 60)[b]		50(15, 60)		35(20, 60)		50(20, 60)	
Formula feeding	22(34.38%)	45(15, 60)		60(40, 60)[b]		47.5(30, 60)		47.5(15, 60)		55(25, 60)	

Infant feeding method was determined by parent responses on the 3-month stool sample information form questions which asked about infant feeding in the 24 hours just prior to stool sample collection. The Kruskal-Wallis test was used to examine the associations between feeding methods and ASQ scales. Dunn's test was performed to do the pairwise comparison. *P-value < 0.05 is significant

Table 6. Associations between feeding methods in the 24 hours prior to stool sample collection at 3 months and infant ASQ scales at 9 months of age

		Univariate model		Multivariate model[1]			
		β (95% CI)	***p*-value**	**β (95% CI)**	***p*-value**	**Overall adjusted R-squared**	**Overall *p*-value**
Gross motor	**Breastfeeding**	Reference		Reference		0	0.85
	Partial breastfeeding	5.24(-4.92, 15.40)	0.31	6.17(-6.19, 18.53)	0.32		
	Formula	7.12(-2.15, 16.38)	0.13	8.33(-3.43, 20.09)	0.16		
Fine motor	**Breastfeeding**	Reference		Reference		0.11	0.11
	Partial breastfeeding	5.10(0.41, 9.78)	0.03*	3.67(-1.62, 8.97)	0.17		
	Formula	4.07(-0.20, 8.34)	0.06	2.23(-2.80, 7.27)	0.38		
Communication	**Breastfeeding**	Reference		Reference		0.10	0.13
	Partial breastfeeding	2.21(-4.98, 9.4)	0.54	0.70(-7.23, 8.64)	0.86		
	Formula	4.60(-1.96, 11.15)	0.17	3.08(-4.46, 10.63)	0.42		
Personal-social	**Breastfeeding**	Reference		Reference		0	0.67
	Partial breastfeeding	-2.64(-10.38, 5.09)	0.50	-1.17(-10.33, 8.00)	0.80		
	Formula	2.64(-4.41, 9.69)	0.46	-1.5(-10.22, 7.22)	0.73		
Problem-solving	**Breastfeeding**	Reference		Reference		0.08	0.17
	Partial breastfeeding	-0.46(-8.14, 7.23)	0.91	2.90(-5.63, 11.42)	0.50		
	Formula	3.78(-3.23, 10.78)	0.29	2.59(-5.52, 10.71)	0.52		

[1]Multivariate linear regression models were used, adjusted by antibiotics use since birth, infant sex, delivery mode, infant race, maternal education level, gestational age at birth, pre-pregnancy BMI, and maternal age. *P-value < 0.05 is significant.

Table 7. Associations between infant feeding in the 24 hours prior to stool sample collection and population characteristics

	N=64	**Breastfeeding (N=9)**	**Breastfeeding with vitamin D (N=17)**	**Partial breastfeeding (N=16)**	**Formula feeding(N=22)**	***p*-value**
Categorical variable[1]	N (%) or Mean±SD	N (%) or Mean±SD	N (%) or Mean±SD	N (%) or Mean±SD	N (%) or Mean (SD)	
Infant sex						
Male	31(48.4%)	4(44.4%)	6(35.3%)	10(62.5%)	11(50%)	0.49
Female	33(51.6%)	5(55.6%)	11(64.7%)	6(37.5%)	11(50%)	
Infant race						
White	44(68.75%)	8(88.9%)	12(70.6%)	11(68.75%)	13(59.1%)	0.47
Non-White	20(31.25%)	1(11.1%)	5(29.4%)	5(31.25%)	9(40.9%)	
Maternal education level						
Did not finish high school	3(4.7%)	0(0%)	0(0%)	0(0%)	3(13.63%)	
High school graduate or GED	11(17.2%)	1(11.1%)	0(0%)	3(18.75%)	7(31.82%)	0.049*
Some college	13(20.3%)	2(22.2%)	3(17.6%)	3(18.75%)	5(22.73%)	
College graduate or more	37(57.8%)	6(66.7%)	14(82.4%)	10(62.5%)	7(31.82%)	
Delivery mode						
Vaginal	39(60.9%)	5(55.6%)	10(58.8%)	15(93.75%)	9(40.9%)	0.01*
C-section	25(39.1%)	4(44.4%)	7(41.2%)	1(6.25%)	13(59.1%)	
Continuous variable[2]						
Pre-pregnancy BMI	32.07±21.98	24.73±4.25	28.69±8.22	27.88±7.89	40.74±34.97	0.06
Maternal age	29.64±4.66	30±3.57	31.06±3.45	30.06±3.91	28.09±5.99	0.37
Gestational age	39.16±1.24	39±1.12	39.71±1.10	39.06±1.53	38.86±1.08	0.19

[1]Categorical variable data was present as N (%). Chi-square test was used to determine the associations between categorical variables and infant feeding method.
[2]Continuous variable data was present as Mean±SD. Kruskal-Wallis test was used to examine the relationship between continuous variables and infant feeding method. *P-value < 0.05 is significant.

Table 8. Associations between feeding methods after stratification by vitamin D supplementation in the 24 hours prior to stool sample collection at 3 months of age and infant ASQ scales at 9 months of age

		Univariate model		Multivariate model[1]			
		β(95% CI)	***p*-value**	**β (95% CI)**	***p*-value**	**Overall adjusted R-squared**	**Over all *p*-value**
Gross motor	Breastfeeding	Reference		Reference		0	0.89
	Breastfeeding with vitamin D	1.01(-12.28, 14.31)	0.88	-2.02(-16.65, 12.62)	0.78		
	Partial breastfeeding	5.90(-7.53, 19.34)	0.38	4.89(-10.66, 20.45)	0.53		
	Formula feeding	7.78(-4.98, 20.54)	0.23	7.15(-7.50, 21.79)	0.33		
Fine motor	Breastfeeding	Reference		Reference		0.29	<0.01*
	Breastfeeding with vitamin D	10.26(4.74, 15.79)	<0.01*	10.27(4.72, 15.82)	<0.01*		
	Partial breastfeeding	11.81(6.22, 17.39)	<0.01*	10.19(4.29, 16.08)	<0.01*		
	Formula feeding	10.78(5.48, 16.09)	<0.01*	8.25(2.70, 13.81)	<0.01*		
Communication	Breastfeeding	Reference		Reference		0.16	0.05
	Breastfeeding with vitamin D	11.70(2.79, 20.61)	0.01*	9.52(0.52, 18.52)	0.04*		
	Partial breastfeeding	9.86(0.86, 18.87)	0.03*	6.74(-2.82, 16.31)	0.16		
	Formula feeding	12.25(3.70, 20.80)	<0.01*	8.67(-0.34, 17.67)	0.06		
Personal-social	Breastfeeding	Reference		Reference		0	0.73
	Breastfeeding with vitamin D	1.18(-8.94, 11.30)	0.82	2.32(-8.53, 13.16)	0.67		
	Partial breastfeeding	-1.88(-12.11, 8.36)	0.72	0.30(-11.22, 11.82)	0.96		
	Formula feeding	3.41(-6.31, 13.12)	0.49	-0.15(-11.00, 10.70)	0.98		

Table 8 (cont'd)

Problem-Solving	Breastfeeding	Reference		Reference		0.09	0.15
	Breastfeeding with vitamin D	7.55(-2.32, 17.42)	0.13	6.45(-3.49, 16.38)	0.20		
	Partial breastfeeding	4.48(-5.50, 14.45)	0.37	6.98(-3.58, 17.54)	0.19		
	Formula feeding	8.71(-0.76, 18.18)	0.07	6.3 (-3.57, 16.32)	0.20		

[1]Multivariate linear regression models were used, adjusted by antibiotics use since birth, infant sex, delivery mode, infant race, maternal education level, gestational age at birth, pre-pregnancy BMI, and maternal age. *P-value < 0.05 is significant

Table 9. Associations between exclusive breastfeeding duration and infant ASQ scales at 9 months of age

	Univariate analysis		Multivariate analysis[1]			
	β (95% CI)	***p*-value**	**β (95% CI)**	***p*-value**	**Overall adjusted R-squared**	**Overall p-value**
Gross motor	-0.004 (-0.05, 0.04)	0.88	0.007 (-0.05, 0.06)	0.81	0	0.94
Fine motor	-0.003 (-0.03, 0.02)	0.77	0.004 (-0.02, 0.03)	0.72	0.09	0.13
Communication	-0.004 (-0.04, 0.03)	0.81	0.01 (-0.02, 0.05)	0.51	0.11	0.10
Personal-social	0.01 (-0.02, 0.05)	0.50	0.01 (-0.02, 0.05)	0.44	0	0.53
Problem-solving	-0.01 (-0.05, 0.02)	0.54	-0.003 (-0.04, 0.03)	0.88	0.09	0.14

[1]Multivariate linear regression models were used, adjusted by antibiotics use since birth, infant sex, delivery mode, infant race, maternal education level, gestational age at birth, pre-pregnancy BMI, and maternal age. *P-value < 0.05 is significant

Table 10. Associations between any breastfeeding duration and infant ASQ scales at 9 months of age

	Univariate analysis		Multivariate analysis[1]			
	β (95% CI)	***p*-value**	**β (95% CI)**	***p*-value**	**Overall adjusted R-squared**	**Overall p-value**
Gross motor	-0.03 (-0.07, 0.02)	0.25	-0.04 (-0.10, 0.01)	0.13	0	0.79
Fine motor	-0.02 (-0.04, 0.005)	0.13	-0.01 (-0.04, 0.01)	0.32	0.11	0.10
Communication	-0.02 (-0.06, 0.006)	0.11	-0.04 (-0.07, 0.0006)	0.05	0.17	0.03*
Personal-social	-0.02 (-0.06, 0.009)	0.16	-0.006 (-0.05, 0.04)	0.78	0	0.58
Problem-solving	-0.01(-0.05, 0.02)	0.44	-0.007 (-0.05, 0.03)	0.71	0.09	0.14

[1]Multivariate linear regression models were used, adjusted by antibiotics use since birth, infant sex, delivery mode, infant race, maternal education level, gestational age at birth, pre-pregnancy BMI, and maternal age. *P-value < 0.05 is significant

3.6.2 Alpha and beta diversity of infant gut microbiota and feeding method at 3 months of age

The gut microbiota richness was different in the four feeding groups (p-value=0.04) (Figure 11A). The diversity of infant gut microbiota differed from the four feeding methods after stratifying exclusively breastfed infants by vitamin D supplementation as measured by Shannon (p-value < 0.01) (Figure 11B) and inverse Simpson (p-value < 0.01) (Figure 11C) indices. When conducting the pairwise comparison, formula-fed infants had significantly higher gut microbial diversity compared to breastfed infants (Figure 11B, 11C). The membership (Sorensen, p-value<0.01) (Figure 12A) and composition (Bray-Curtis, p-value < 0.01) (Figure 12B) of the infant gut microbiota differed by feeding method. Formula-fed infants had different gut microbial membership and composition compared to exclusively breastfed, vitamin D supplemented exclusively breastfed, and partially breastfed infants.

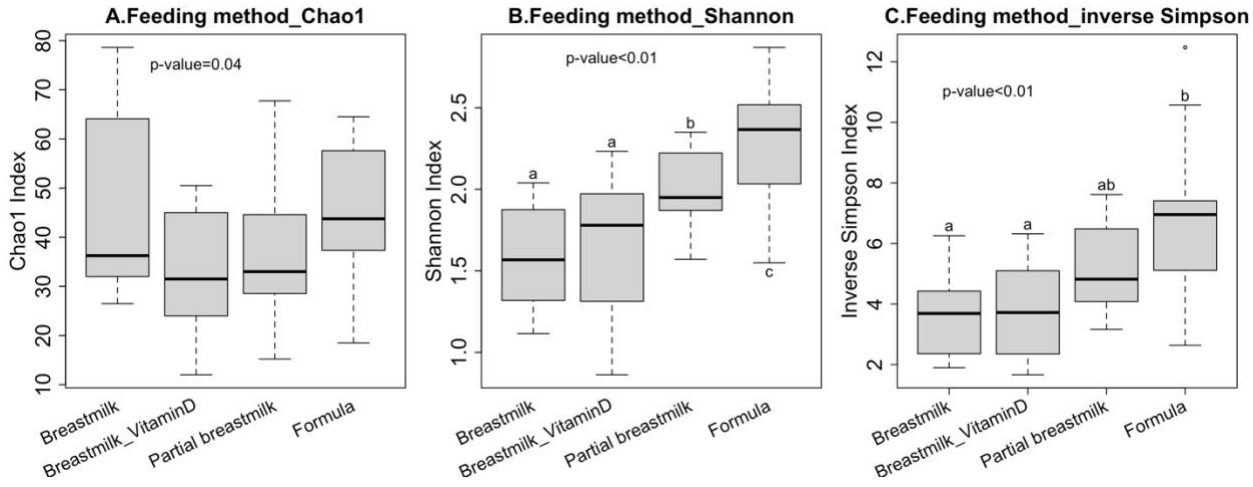

Figure 11. Associations between infant feeding method in the 24 hours prior to stool sample collection and infant gut microbiota alpha diversity at 3 months of age

Shapiro-Wilk test was used to test data normality. ANOVA tests were used to examine the relationships between Chao1 and Shannon indices with feeding methods. The relationship between inverse Simpson and feeding methods was tested by the Kruskal-Wallis test. Tukey's HSD and Dunn's tests were conducted for post hoc comparisons. Median with the min and max was plotted. Different letters indicate significant differences in pairwise comparisons. P-value < 0.05 is significant.

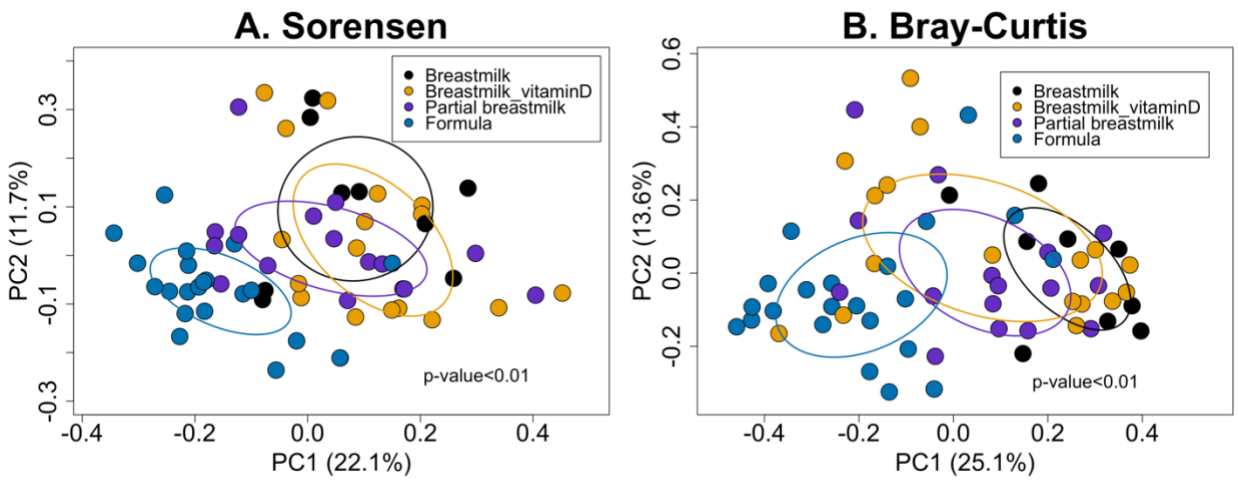

Figure 12. Associations between infant feeding methods in the 24 hours prior to stool sample collection and gut microbiota beta diversity at 3 months of age

PERMANOVA was performed to examine the relationships between beta diversity and clusters. P-value < 0.05 is significant.

3.6.3 Mediation analyses

In Aim 1 (Chapter 2), we reported that one measure of the alpha diversity of the gut microbiota, inverse Simpson, tended to be associated with communication (p-value=0.07) and problem-solving (p-value=0.07) scores (Table 2). Shannon index, another measure of the alpha diversity of the gut microbiota, was significantly associated with problem-solving score (p-value=0.04). The Bray-Curtis dissimilarity matrix, a measure of the beta diversity of gut bacterial communities, was associated with fine motor (p-value < 0.01) and communication (p-value < 0.01) scores (Table 3). In this chapter, the roles of inverse Simpson (alpha diversity), Shannon (alpha diversity), and Bray-Curtis (beta diversity) metrics as mediators in the association between infant feeding method (exposure) and ASQ scale scores (outcome) were evaluated. We reported the total, direct, and indirect effects. Direct effect indicates the effect from exposure to outcome after ignoring the mediating effect. Indirect effect is a measure of mediating effect.

Total effect indicates the effect from exposure to outcome, including the mediation effect of the mediator (Figure 13).

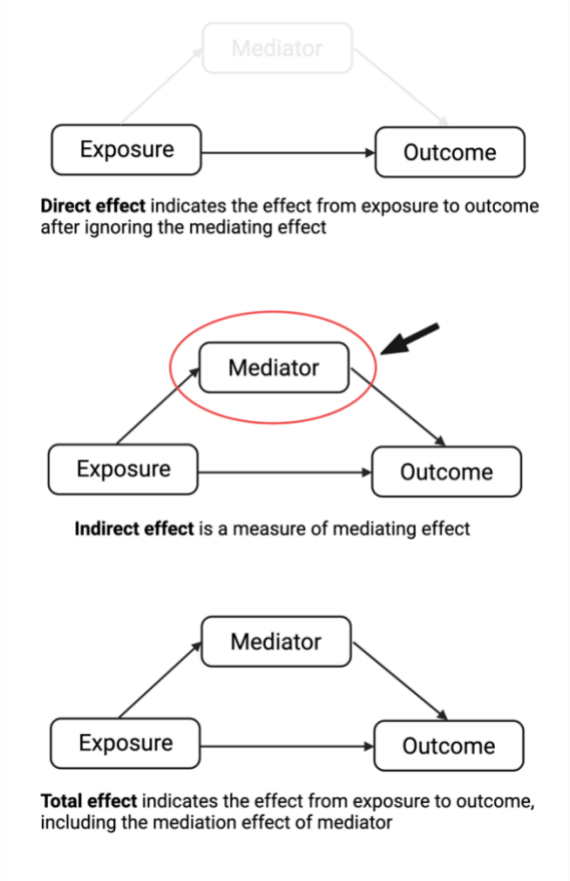

Figure 13. Direct effect, indirect effect, and total effect in mediation analysis

When considering the mediating effect of inverse Simpson of alpha diversity, the total effect of breastfeeding with vitamin D on the communication scales tended to score 9.51 units higher than breastfeeding (p-value=0.04) (Table 11). The total effect of formula feeding on the communication scales tended to score 8.66 units higher than breastfeeding (p-value=0.06). After ignoring the mediating effect, the direct effect of breastfeeding plus vitamin D supplementation on communication tended to increase by 8.22 units significantly compared to breastfeeding (p-value=0.07). The direct effect was insignificant when comparing the effect of partial

breastfeeding verse breastfeeding (p-value=0.34) and formula verse breastfeeding (p-value=0.32) on communication scores. The association of the infant feeding method at 3 months of age on communication at 9 months of age was not mediated by inverse Simpson of alpha diversity at 3 months of age (Table 11). Although the mediating effect was not statistically significant, one unit increased in breastfeeding with vitamin D intake, partial breastfeeding, and formula, the mediating effect increased by 1.29, 2.15, and 3.83 units, respectively, compared to the breastfeeding.

When the exposure was infant feeding method, the outcome was problem-solving scores, and the mediator was inverse Simpson (Table 12) or Shannon (Table 13). Neither the total effect nor the direct effect of infant feeding on problem-solving scores was statistically significant. Thus, the alpha diversity of the 3-month-old infant gut microbiota, as described by the inverse Simpson (Table 12) and Shannon (Table 13) indices, did not mediate the relationship between feeding method and problem-solving skills.

However, using the LDM package of Hu & Stratten (Hu & Satten, 2020) to test the mediation effect of the Bray-Curtis dissimilarly matrix, the Bray-Curtis distance matrix of beta diversity mediated the association of the feeding method and ASQ fine-motor (p-value=0.04) scores in univariate analysis (Table 14).

Table 11. Mediation effect of the inverse Simpson index on the association of feeding method with communication score

	Feeding method	β (95% CI)	*p*-value
Direct effect (exposure to outcome, ignoring the mediation effect)	Breastfeeding	Reference	
	Breastfeeding_vitaminD	8.22(-0.59, 17.03)	0.07
	Partial breastfeeding	4.57(-4.92, 14.06)	0.34
	Formula	4.82(-4.84, 14.49)	0.32
Indirect effect (a measure of mediating effect)	Breastfeeding	Reference	
	Breastfeeding_vitaminD	1.29(-1.75, 4.44)	0.37
	Partial breastfeeding	2.15(-1.46, 6.04)	0.23
	Formula	3.83(-1.75, 9.28)	0.12
Total effect (exposure to outcome, including the mediation effect)	Breastfeeding	Reference	
	Breastfeeding_vitaminD	9.51(0.62, 18.41)	0.04*
	Partial breastfeeding	6.72(-2.69, 16.12)	0.16
	Formula	8.66(-0.25, 17.57)	0.06

Simple mediation analysis was performed using the MeMoBootR package, adjusted by infant sex, delivery mode, infant race, maternal education level, gestational age at birth, pre-pregnancy BMI, and maternal age. *P-value < 0.05 is significant.

Table 12. Mediation effect of the inverse Simpson index on the association of feeding method with problem-solving score

	Feeding method	β (95% CI)	*p*-value
Direct effect (exposure to outcome, ignoring the mediation effect)	Breastfeeding	Reference	
	Breastfeeding_vitaminD	4.68(-5.16, 14.52)	0.34
	Partial breastfeeding	3.83(-6.77, 14.42)	0.47
	Formula	1.77(-9.01, 12.56)	0.74
Indirect effect (a measure of mediating effect)	Breastfeeding	Reference	
	Breastfeeding_vitaminD	1.49(-1.62, 4.90)	0.37
	Partial breastfeeding	2.48(-1.18, 6.65)	0.22
	Formula	4.42(-1.33, 10.14)	0.11
Total effect (exposure to outcome, including the mediation effect)	Breastfeeding	Reference	
	Breastfeeding_vitaminD	6.17(-3.78, 16.12)	0.22
	Partial breastfeeding	6.30(-4.22, 16.83)	0.23
	Formula	6.19(-3.78, 16.16)	0.22

Table 12 (cont'd)

Simple mediation analysis was performed using the MeMoBootR package, adjusted by infant sex, delivery mode, infant race, maternal education level, gestational age at birth, pre-pregnancy BMI, and maternal age. *P-value < 0.05 is significant.

Table 13. Mediation effect of the Shannon index on the association of feeding method with problem-solving score

	Feeding method	**β (95% CI)**	***p*-value**
Direct effect (exposure to outcome, ignoring the mediation effect)	Breastfeeding	Reference	
	Breastfeeding_vitaminD	4.93(-4.86, 14.71)	0.32
	Partial breastfeeding	2.74(-8.17, 13.65)	0.62
	Formula	1.42(-9.50, 12.33)	0.80
Indirect effect (a measure of mediating effect)	Breastfeeding	Reference	
	Breastfeeding_vitaminD	1.24(-2.55, 4.89)	0.44
	Partial breastfeeding	3.56(-0.52, 8.24)	0.14
	Formula	4.78(-1.66, 11.07)	0.10
Total Effect (exposure to outcome, including the mediation effect)	Breastfeeding	Reference	
	Breastfeeding_vitaminD	6.17(-3.78, 16.12)	0.22
	Partial breastfeeding	6.30(-4.22, 16.83)	0.23
	Formula	6.19(-3.78, 16.16)	0.22

Simple mediation analysis was performed using the MeMoBootR package, adjusted by infant sex, delivery mode, infant race, maternal education level, gestational age at birth, pre-pregnancy BMI, and maternal age. *P-value < 0.05 is significant.

Table 14. Mediation effect of the Bray-Curtis dissimilarity matrix on the association of feeding method with communication and fine motor scores

			Univariate analysis	Multivariate analysis[1]
Exposure	Mediator	Outcome	*p*-value	*p*-value
Infant feeding method	Bray-Curtis	Communication	0.16	0.55
Infant feeding method	Bray-Curtis	Fine motor	0.04	0.28

PERMANOVA-FL function from LDM package was used to test the mediation effect when the infant feeding method was the exposure, Bray-Curtis dissimilarity matric was the mediator, and ASQ scales were the outcomes.
[1]Multivariate linear regression models were used, adjusted by antibiotics use since birth, infant sex, delivery mode, infant race, maternal education level, gestational age at birth, pre-pregnancy BMI, and maternal age. *P-value < 0.05 is significant.

3.7 Discussion

The current study demonstrated that breastfed infants with vitamin D supplementation had higher fine motor and communication scores than those exclusively breastfed. Further, these results suggest that the feeding method in early infancy could potentially impact neurodevelopmental outcomes in later infancy. We also observed that the gut microbial membership and composition at 3 months, as measured by Bray-Curtis dissimilarity matrix mediates the association between infant feeding at 3 months and the infant neurodevelopmental outcomes of fine motor scores at 9 months. This study is a pilot study that investigated the mediating effect of gut microbiota in the association between infant feeding methods and neurodevelopmental outcomes. Further study with a larger and more diverse population will be analyzed when all necessary data collection has been completed by the cohort.

Vitamin D plays an important role in brain development during the early years of life (Schwarzenberg & Georgieff, 2018). Our study found that breastfed infants with vitamin D supplementation exhibited higher fine motor and communication scores than those non-supplemented infants who were exclusively breastfed. In an animal study, vitamin D deficiency in early life resulted in decreased social behavior, impaired learning, and memory problems among male adult rats (Yates et al., 2018). In humans, serum vitamin D level at birth was positively associated with communication and personal-social scores in 2-year-old infants, as measured by ASQ-3 (Juwita et al., 2021). Vitamin D supplementation in early life dose-dependently improved neurodevelopment in extremely preterm infants, but this was not statistically significant (Salas et al., 2018). However, other studies reported inconsistent results when describing the association between vitamin D supplementation in early life and neurodevelopment in childhood. Chowdhury et al. measured plasma vitamin D levels when infants were 6-30 months of age and observed that such levels were not associated with cognitive

development at 9 years of age (Chowdhury et al., 2020). There was no association between vitamin D status and motor performance when children were 5 years old (Filteau et al., 2016). The occurrence of these inconsistencies suggests that there could be an optimal time point in early life to examine the effect of vitamin D status on neurodevelopmental outcomes. The timing of vitamin D assessment, duration of vitamin D supplementation, the dose of vitamin D supplementation, the age of neurodevelopmental assessment, and tools for assessment might also influence the results.

Breastfeeding has long been considered to protect against adverse health outcomes, such as obesity and metabolic diseases, particularly when such outcomes are compared between breastfed infants and those infants fed formula (Armstrong & Reilly, 2002; Azad et al., 2018; Plagemann & Harder, 2005). Further, it is still debated whether breastfeeding is beneficial for cognitive development. Breastfeeding for more than nine months enhanced the cognitive development of Korean infants as measured by Bayley Scales of Infant Development II (Lee et al., 2016). This beneficial impact of breastfeeding remained evident until the children reached three years of age, even after accounting for other factors. Similarly, a meta-analysis of 20 studies reported that breastfeeding was linked to considerably enhanced cognitive development, spanning from infancy through to adolescence compared to formula feeding (Anderson et al., 1999). On the contrary, breastfeeding in the first after birth was found to have little or no effect on intelligence in 5-14 years old children using Peabody individual achievement test in the US (Der et al., 2006). There was no association between a long duration of breastfeeding and later cognitive development in 9- to 10-year-old children in South India, as measured by the Kaufman Assessment Battery for Children (Veena et al., 2010). In our study, we also found breastfeeding duration was not related to neurodevelopmental outcomes. The characteristics of individuals who

breastfeed their infants in the US have been extensively studied, and several factors such as maternal age (Colombo et al., 2018; Kitano et al., 2016), education level (Colombo et al., 2018), and household income (Temple Newhook et al., 2017), race and ethnicity(Jones et al., 2015) have been found to be associated with breastfeeding rates. However, it is important to note that these characteristics are not necessarily the driving factors in the observed improvements in neurodevelopment that have been linked to breastfeeding. Overall, cumulative evidence suggests whether breastfeeding can affect children's neurodevelopment is undetermined and deserves further analysis.

Interestingly, we found that infants fed formula at 3 months had higher fine motor and communication scores at 9 months compared to those fed with exclusive breast milk. The compositional difference in nutrients of breast milk and formula could possibly explain this. Formula-fed infants often have greater weight gains in infancy than breastfed infants because of the higher protein content in formula (Alexy et al., 1999; Dewey, 1998; Farrow et al., 2013; Kramer et al., 2004; Ren et al., 2022; Victora et al., 1998). Though some evidence suggests a positive association between protein intake and neurodevelopment in infancy, the evidence is mixed. In a cohort study, increased protein intake in the first month of life was not associated with better cognitive, language, and motor scores or decreased sensory impairments at 2 years of age (Cester et al., 2015). However, other studies reported the opposite results. Increased protein intake in the first week after birth was associated with higher Mental Development Index scores at 18 months in extremely low birth weight infants (Stephens et al., 2009). A positive association was demonstrated between protein intake during the first 28 days and cognitive and motor scores at 2 years in infants born at a gestational age of less than 31 weeks (Coviello et al., 2018). The current study excluded preterm-born infants with a gestational age of less than 37 weeks and

studied the relationship between the feeding method in the first 3 months of life (early infancy) and neurodevelopment at 9 months (late infancy). Thus, there is abundant room for further research in determining whether the feeding method in early infancy predicts neurodevelopment in late infancy.

Gut microbiota colonization and human brain development have similar developmental windows, and these windows occur during infancy (Ratsika et al., 2023). Gut-microbiota-axis (GBA), the bidirectional communication between the gut and brain, has been proposed (Carabotti et al., 2015). In the current study, we demonstrated that infant gut microbiota membership and composition (Bray-Curtis dissimilarity matrix) at 3 months of age mediated the association between infant feeding method at 3 months of age and infant neurodevelopment (fine motor scores) at 9 months. This result supports the assumption that nutritional intervention may be a novel strategy for initializing gut microbial colonization in early infancy with the aim of altering neurodevelopmental outcomes in late infancy. The extent to which and specific mechanisms by which the infant gut microbiota modulates neurodevelopment and how the infant feeding method mediates this association is still under investigation.

ASQ is generally reliable in identifying young children who may require an additional assessment to determine their eligibility for early intervention services. This screening tool has the advantages of being cost-effective, simple to administer, and efficient in terms of time. However, ASQ is a parent-reported measurement. Thus, some biases may result from this parental report. For example, parents with low socioeconomic status have been reported to over- or underestimate their children's performance on the questions from ASQ (Feldman et al., 2000). Some parents might be prone to social desirability bias (Bourdeaudhuij & Oost, 2000). In addition to ASQ, Bayley Scales of Infant and Toddler Development, a more formal and accurate

developmental assessment tool, is widely used to diagnose developmental delays in early childhood (Balasundaram & Avulakunta, 2022). Magnetic Resonance Imaging (MRI) can also be used if budget and time are allowed (Arulkumaran et al., 2020).

The present study has several strengths. We are the first study investigating the mediating effect of early-life gut microbiota in the association between infant feeding method and neurodevelopmental outcomes. Our study provides insights into the development of a nutritional intervention by manipulating gut microbiota in early life to help prevent or reverse neurodevelopmental disorders. In addition, we excluded preterm-born infants who typically have delayed neurodevelopment compared to full-term infants. Therefore, our findings are generalizable among full-term infants. There are several limitations to this study. Our sample size (n=64) is small, which could reduce the power of this study. The small sample size further limits the covariates which can be included in the statistical models. Additionally, the small sample size may lead to a poor representation of participants with specific characteristics, which could bias the results of these analyses. For example, a large proportion of exclusively breastfed infants who received a vitamin D supplement were non-White, whereas all but one non-supplemented exclusively breastfed infant was White. They might also have memory bias when collecting breastfeeding duration information until 9 months. Finally, we did not consider exposures at 9 months of age such as the contact with other infants during day care and feeding practices.

3.8 Conclusion

The evidence presented herein suggests that vitamin D supplementation could improve fine motor and communication skills among breastfed infants. Infants fed formula at 3 months had higher fine motor and communication scores at 9 months compared to those fed exclusive

breast milk. The Bray-Curtis dissimilarity matrix of gut microbiota at 3 months of age mediated the association between the infant feeding method at 3 months and fine motor scores at 9 months. Future studies with a more diverse population and more comprehensive neurodevelopment tools are needed to test the mediation effect of gut microbiota in the association of infant feeding on neurodevelopmental outcomes.

CHAPTER 4: THE RELATIONSHIPS BETWEEN BREAST MILK FEEDING PRACTICES AND INFANT GUT MICROBIOTA AT THREE MONTHS OF AGE

4.1 Abstract

Breastmilk plays a critical role in infant's growth and development. In addition to meeting the infant's direct nutritional needs, breastmilk can promote the growth of beneficial bacteria in infant's gut and maintain a healthy gut environment. Further, the act of feeding at the breast may also have beneficial effects on infant development. Currently, it's unknown how breastmilk feeding patterns (breastfeeding from breast, breastfeeding through a bottle, and breastfeeding through both breast and bottle) influence the infant gut microbial development. Therefore, this chapter aimed to investigate the relationship between breastfeeding patterns and infant gut microbiota among exclusively breastmilk-fed infants at 3 months of age. An additional aim was to compare gut microbes in infants of exclusively human milk fed groups to those in infants fed at least some formula. DNA was extracted, followed by the preparation of 16S rRNA libraries and sequencing on the Illumina MiSeq platform. Community sequencing data were processed using mothur, and data were analyzed in R. Bottle-fed infants had numerically lower alpha diversity of the gut microbiota than breast- and mixed-fed infants, but it was not statistically significant. Breast-fed infants had different gut microbial membership compared to bottle-fed and mixed-fed infants as measured by Sorensen dissimilarity matrix. Breast-fed infants had a lower abundance of *Bifidobacterium* but a higher abundance of *Enterobacteriaceae* unclassified compared to bottle- and mixed-fed infants. Infants in the groups fed some human milk had a higher abundance of *Lacticaseibacillus* compared to infants fed formula. These results suggest that breastfeeding patterns may play a role in shaping the composition and diversity of the gut microbiota in infants. Further research in analyzing the human milk bacteria is needed to better understand the mechanisms behind these differences and to determine the long-term implications for infant health.

4.2 Key words

breast milk, human milk, breastfeeding, exclusive breastmilk feeding, bottle-feeding, breastfeeding, mixed-feeding, infant feeding, *Bifidobacterium, Enterobacteriaceae* unclassified, *Escherichia-Shigella, Blautia, Parabacteroides*

4.3 Introduction

Breastfeeding profoundly influences the colonization and maturation of the infant gut microbiome (Li et al., 2021; O'Sullivan et al., 2015; Sugino, Ma, Paneth, et al., 2021). Breastmilk is recommended for the first six months of life as it provides the ideal energy and nutrients to support infants' growth and well-rounded development (Guittar et al., 2019). The human milk oligosaccharides (HMOs) are one of the main components of breast milk, which are partially digested in the small intestine and mostly reach the colon, where they are metabolized by *Bifidobacterium*, a beneficial bacteria, to produce metabolites that have physiological benefits and modulate immunological development (Donovan & Comstock, 2016; Le Huërou-Luron et al., 2010; Marcobal et al., 2010; Stuivenberg et al., 2022). In addition to the prebiotic effects of promoting the growth of beneficial bacteria, breast milk also contains diverse bacterial communities. It is recognized to be a potential source of bacteria that colonize the infant gut (Urbaniak et al., 2016). Exclusively breastfed infants had lower bacterial diversity, a higher abundance of *Bifidobacterium,* and a lower abundance of *Lachnospiraceae* compared to partially or non-breastfed infants (Baumann-Dudenhoeffer et al., 2018; Forbes et al., 2018; Sugino, Ma, Kerver, et al., 2021). Formula-fed infants had a distinct gut microbial composition from breastfed infants (Haddad et al., 2021; Ma et al., 2022; O'Sullivan et al., 2015; Yatsunenko et al., 2012). Exclusively formula-fed infants displayed a more diverse gut microbiota with a lower abundance of Bifidobacterium species and an increased abundance of Clostridium species and

Enterobacteriaceae species due to the lacking of HMOs and higher protein contents in infant formula (Bäckhed et al., 2015; Benno et al., 1984; Penders et al., 2007). The mode of breastfeeding includes direct breastfeeding, expressed breastfeeding, and mixed feeding (Pang et al., 2017; Pérez-Escamilla et al., 2023). Direct breastfeeding is when an infant feeds directly from the breast. In contrast, expressed breast milk is when an infant consumes human milk that has been manually or mechanically expressed via a pump and is provided through a bottle, cup, or spoon. Mixed feeding occurs when an infant is both fed directly at the breast and given expressed breast milk (Pang et al., 2017). In this chapter, direct breastfeeding is referred to as "breastfeeding" or "breast;" expressed breastfeeding is referred to as "bottle feeding" or "bottle," and mixed feeding is referred to as "mixed feeding" or "mix."

Pumping breast milk into a bottle can impact the bacterial composition of breast milk (Differding & Mueller, 2020; Moossavi & Azad, 2020; Weiss, 2005). However, the consequences of pumping and breastfeeding on infant gut microbiota have not been well studied. *Streptococcus spp.* and *Veillonella dispar* co-occurred in breast milk and infant's stool, but this co-occurrence was reduced when infants were fed with pumped breastmilk (Fehr et al., 2020). They also reported that infants fed exclusively with direct breastmilk and those fed some pumped breastmilk had similar gut microbial composition (Fehr et al., 2020).

It has yet to be fully examined whether there is a compositional difference in the infant gut microbiota when data are analyzed by breastfeeding patterns in the 24 hours immediately preceding fecal collection and the proportion of the human milk intake in the past week. There is little evidence on the association between breastfeeding patterns in the past 24 hours and infant gut microbiota when the infants are at 3 months of age. Therefore, this study aimed to investigate the relationship between breastmilk feeding patterns and infant gut microbiota in order to

determine how breastmilk feeding patterns and the proportion of the breastmilk affect the infant gut microbial composition.

4.4 Materials and methods

4.4.1 Study population

The study population was described in aim 1. For aim 3, a total of 299 3-month-old infants were included in the final analysis, in which 136 infants were exclusively breastfed. Population demographics information was obtained from MARCH Prenatal 1 Survey questionnaire that asks about mothers' education level, maternal age, mother's height, and pre-pregnancy weight. The birth certificate information includes the infant sex, estimated weeks of gestation, and mode of delivery. Infant race information was collected through MARCH 3-month survey dictionary. The sample collection form, completed at the time of fecal sample collection and when the infants were 3 months of age, included information about the infant diet in the past 24 hours and the infant dietary intake in the week prior to fecal collection, and breastfeeding patterns (at the breast, from a bottle, or mixed from breast and bottle). The Michigan State University Human Research Protection Program approved the study (IRB# 16-1429).

4.4.2 Classification of breastfeeding patterns in the past day and the proportion of breastmilk intake in the past week

The breastfeeding patterns in the past day among exclusively breastfed infants were classified as breastfeeding at the breast, bottle feeding, and mixed (at breast and from bottle) feeding. These infants were also reported to be fed 100% breastmilk in the past week. The additional categories of the proportion of breastmilk intake in the past week were breastmilk > 50%, breastmilk $\leqslant$ 50%, and exclusively formula.

4.4.3 Stool sample collection

Sample collection was as described in Aim 1.

4.4.4 Laboratory procedures

4.4.4.1 DNA extraction and 16S rRNA gene amplification

DNA extraction, 16S rRNA gene amplification, and sequencing were carried out on stool samples as described in Aim 1. The only alteration was: PCR amplicon purification and quantification were conducted using SequalPrep™ Normalization Plate Kit per the manufacturer's instructions (Thermo Fisher Scientific, Waltham, MA).

4.4.4.2 Processing and analysis of sequence data

The processing of sequencing data was also described in Aim 1. Samples were rarefied to 1383 reads per sample before further analysis. Rarefaction curves were generated to confirm adequate community coverage.

4.5 Statistical analysis

All data were analyzed using R (version 4.0.2). Data normality was tested using Shapiro – Wilk test from stats package. Chi-square (stats package) for categorical population characteristics and Kruskal-Wallis (stats package) for continuous variables were used to examine the relationships with breastfeeding patterns (breastfeeding, bottle feeding, and mixed feeding) among exclusively breastfed infants. Data is presented as N (%) for categorical variables and Mean±SD with median (min, max) for continuous variables. Alpha diversity (Chao1, Shannon, and inverse Simpson indices) were assessed using the vegan package (Jari Oksanen et al., 2020). For the analysis of breastfeeding patterns, the relationships between Chao1 and Shannon and

breastfeeding patterns were tested using Kruskal-Wallis. Analysis of variance (ANOVA) from stats package was used to determine the relationship between inverse Simpson and breastfeeding patterns. For the analysis of all infants in the six feeding groups, relationships between Chao1 and inverse Simpson and feeding groups were tested using Kruskal-Wallis. ANOVA was used to determine the relationship between Shannon and six groups. Dunn's test (dunn.test package) for Kruskal-Wallis and Tukey's HSD test (stats package) for ANOVA was used to conduct post hoc tests. Sorensen and Bray-Curtis dissimilarities of beta diversity were calculated using the vegan package and ordinated using principal coordinate analysis (PCoA). Permutational multivariate analysis of variance (PERMANOVA) was performed using the vegan package to test for significant differences in beta diversity. Post hoc pairwise comparison with FDR correction (Benjamini-Hochberg procedure, BH) was conducted to investigate the associations between two groups regarding beta diversity using pairwiseAdonis package. Average relative abundance for an OTU is calculated by summing all counts for that OTU and dividing by the total number of counts across all samples, then multiplying by 100 to get percent abundance. Taxa with an average relative abundance larger or equal to 1% were selected in the final analysis. Negative Binomial Generalized Linear Model from MASS package with FDR correction (BH procedure) was carried out to determine if the relative abundance of taxa differed by breastfeeding patterns or/and proportion of breastmilk intake. To validate the results from NB, Multivariate Association with Linear Models (MaAsLin) with FDR correction (BH procedure) from Maaslin2 package was used to investigate the associations between breastfeeding patterns and proportion of breastmilk intake and individual taxa (Mallick et al., 2021), adjusted by infant sex, infant race, mode of delivery, maternal education level, gestational age at birth, maternal pre-pregnancy BMI, maternal age and antibiotics use since birth. P-value<0.05 is significant. Associations are

considered significant when the q-value<0.1.

4.6 Results

4.6.1 Population characteristics

A total of 136 exclusively breastfed infants were included in the final analyses (Table 15). Of these, a majority of breastfed infants were female (53.7%) and White (87.5%). However, breastfeeding patterns were similar by infant sex and race. Maternal educational level tended to be associated with breastfeeding patterns (p-value=0.08). Around half of bottle-fed (54.5%) and mixed-fed infants (46.8%) were born to mothers with master's or PhD degrees, whereas almost the same numbers of breast-fed infants were born to mothers with some college (33.3%), bachelor's degree (30.2%), or master's or PhD degree (27%). Mode of delivery, maternal pre-pregnancy BMI, gestational age, and maternal age were not associated with breastfeeding patterns.

Table 15. Population characteristics and breastfeeding patterns among exclusively breastfed infants

	N=136	**Breastfeeding (N=63)**	**Bottle feeding (N=11)**	**Mixed feeding (N=62)**	***p*-value**
Categorical variable[1]	N(%) or Mean±SD	N(%) or Median(min,max)	N(%) or Median(min,max)	N(%) or Median(min,max)	
Infant sex					
Male	63(46.3%)	24(38.1%)	6(54.5%)	33(53.2%)	0.20
Female	73(53.7%)	39(61.9%)	5(45.5%)	29(46.8%)	
Infant race					
White	119(87.5%)	56(88.9%)	10(90.9%)	53(85.5%)	0.26
Black	4(2.9%)	0(0%)	0(0%)	4(6.4%)	
Others	13(9.6%)	7(11.1%)	1(9.1%)	5(8.1%)	
Maternal education level					
High school or some high school	10(7.4%)	6(9.5%)	0(0%)	4(6.5%)	0.08
Some college	31(22.8%)	21(33.3%)	1(9.1%)	9(14.5%)	
Bachelor's degree	43(31.6%)	19(30.2%)	4(36.4%)	20(32.2%)	
Master's or PhD degree	52(38.2%)	17(27%)	6(54.5%)	29(46.8%)	
Delivery mode					
Vaginal	99(72.8%)	50(79.4%)	9(81.8%)	40(64.5%)	0.15
C section	37(27.2%)	13(20.6%)	2(18.2%)	22(35.5%)	
No	114(83.8%)	54(85.7%)	11(100%)	49(79%)	
Continuous variable[2]					
Pre-pregnancy BMI	26.1±6.4	24.3 (17.6, 47.1)	23.5 (19, 39.5)	23.9(17, 46.5)	0.99
Gestational age	38.9±1.58	39(34, 41)	39(37, 40)	39(31, 41)	0.23
Maternal age	30.7±4.5	31(20, 51)	32(24, 34)	30.5(19, 42)	0.89

[1]Categorical variable data is presented as N (%). Chi-square was used to examine the associations between infant sex, infant race, maternal education level, mode of delivery and breastfeeding patterns. [2]Continuous variable data is presented as Mean±SD and Median(min,max). The Kruskal-Wallis test was used to examine the associations between maternal pre-pregnancy BMI, gestational age at birth, and maternal age and breastfeeding patterns. *P-value < 0.05 is significant

4.6.2 Alpha and beta diversity of the infant gut microbiota in relation to breastfeeding patterns

The gut microbial diversity of infants was similar between breastfeeding, bottle feeding, and mixed feeding (Figure 14). The gut microbiota richness (Chao1 index) and diversity (Shannon index) was numerically lower in bottle-fed infants compared to breastfed and mixed-

fed infants (Figure 14A, 14B). Similarly, mixed-fed infants seemed to have the lowest gut microbial richness and evenness among the more abundant microbiota compared to the other two groups (Figure 14C). However, it is important to note that these observed differences were not statistically significant.

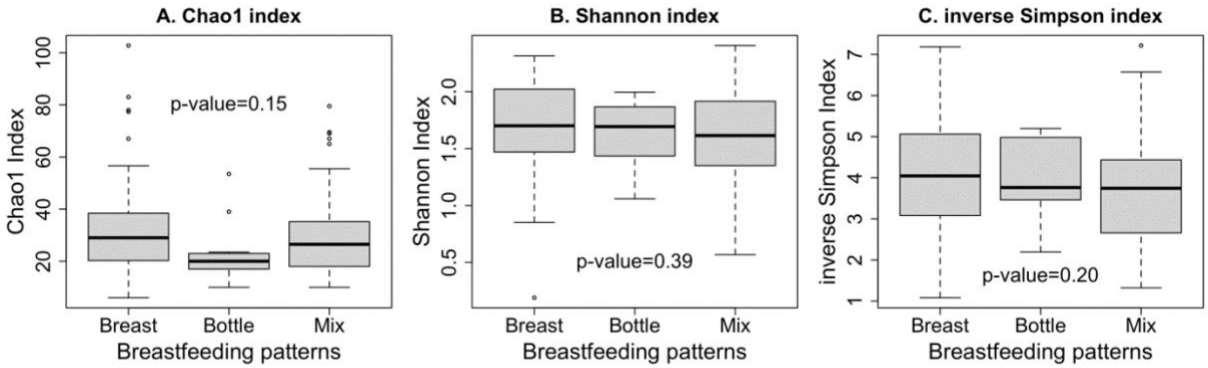

Figure 14. The associations between alpha diversity of the gut microbiota and infant breastfeeding patterns

Shapiro–Wilk test was used to test data normality. Kruskal-Wallis test was used to examine the relationships between Chao1 (A) and Shannon (B) indices and breastfeeding patterns. The relationship between inverse Simpson index (C) and breastfeeding patterns was tested by ANOVA. P-value < 0.05 is significant.

Breastfed, bottle-fed, and mixed-fed infants had significantly different gut microbial membership (p-value=0.03, Figure 15A) but similar gut microbial composition (Figure 15B). Gut microbiota of breastfed and mixed-fed infants had more similar richness (Figure 15A) compared to exclusively bottle-fed infants, explained by the closer ellipses and post hoc PERMANOVA tests (Breast vs Bottle, p-value=0.04; Breast vs Mix, p-value=0.4; Bottle vs Mix, p-value=0.3).

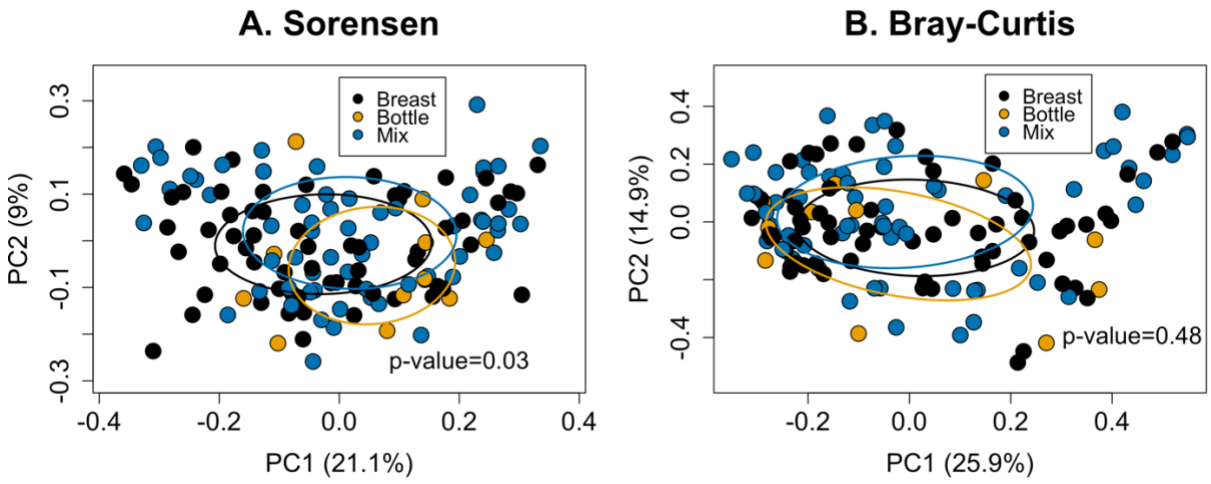

Figure 15. The associations between beta diversity of the gut microbiota and infant breastfeeding patterns

PERMANOVA was performed to test the relationships between beta diversity and breastfeeding patterns. P-value<0.05 is significant.

4.6.3 Associations of alpha and beta diversity with breastfeeding patterns in the past day and dietary intake in the past week

The alpha and beta diversity were compared between three breastfeeding patterns in the past day and three dietary intake groups in the past week (Figure 16, Figure 17). The infants in the three breastfeeding pattern groups (breast, bottle, and mix) were 100% breastmilk fed in the past day and were excluded from the other three infant feeding groups (> 50% breastmilk, ≤ 50% breastmilk, and formula).

Only breastfed infants had a similar gut microbial richness to those who fed breastmilk > 50% or ≤ 50% in the past week (Figure 16A). The richness of gut microbiota of infants fed with exclusive breastmilk from the bottle and from both bottle and breast (mix) was significantly lower than those fed with breastmilk > 50%, breastmilk ≤ 50%, and formula in the past week. Infants fed with breastmilk>50%, breastmilk≤50%, and formula had similar gut microbial richness (Figure 16A). Infants fed breastmilk from breast or bottle had similar gut microbial diversity to the infants fed with more than 50% breastmilk, however; infants fed with a mix of

breastmilk from both breast and bottle had lower gut microbial diversity compared to the infants fed with > 50%, ≤ 50% of the breastmilk, and formula (Figure 16B, 16C). Infants fed with breastmilk > 50% had significantly lower gut microbial richness and composition than those fed breastmilk ≤ 50% and formula.

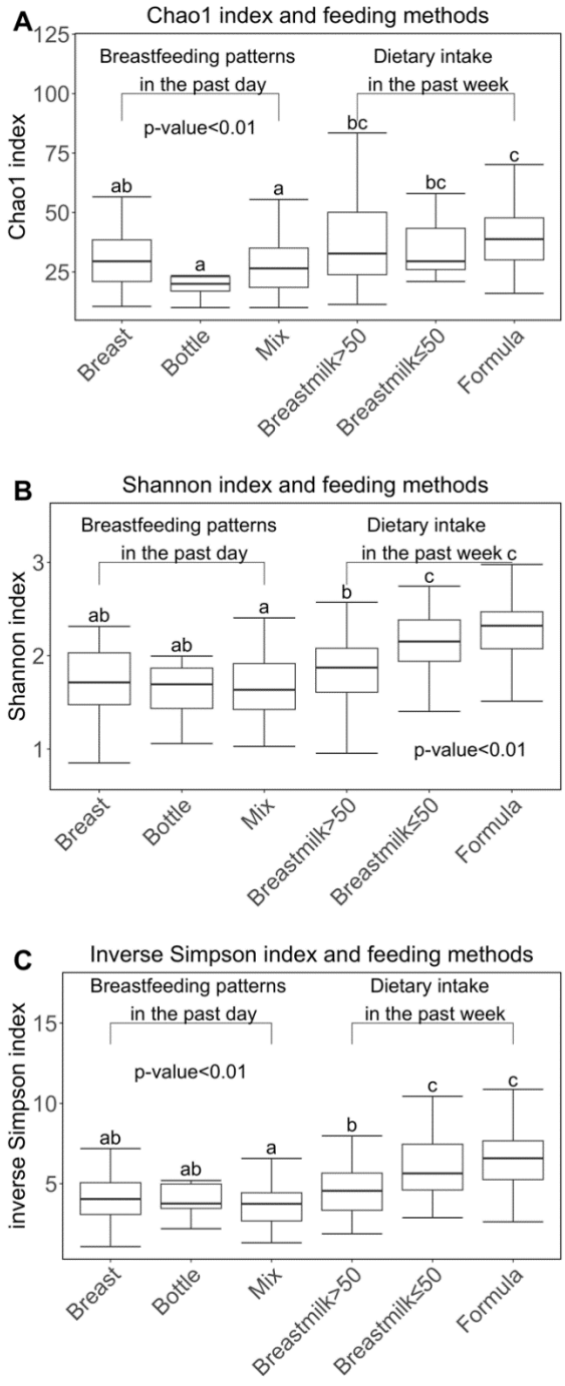

Figure 16. The associations between alpha diversity of the gut microbiota and breastfeeding patterns in the 24 hours immediately preceding stool sample collection for infants exclusively fed human milk and dietary intake in the past week for infants fed at least some formula

Shapiro–Wilk test was used to test data normality. Kruskal-Wallis tests were used to examine the relationships between Chao1(A) and inverse Simpson(C) indices and feeding groups. The relationship between Shannon (B) and feeding groups was tested by the ANOVA test. Dunn's and Tukey's HSD tests were performed for pairwise comparison. All infants in breastfeeding pattern groups were 100% breastmilk fed in the past week. They were excluded from the dietary intake groups. The boxplot shows the minimum, first quartile (Q1), median (Q2), third quartile (Q3), and maximum. P-value < 0.05 is significant.

For beta diversity, the gut microbial membership and composition differed across the six groups (Figure 17). Breastmilk from breast, bottle and mixed-fed infants had different gut microbial richness (Sorensen index) compared to > 50% and ≤ 50% breastmilk-fed infants (Figure 17A, Table 16). Bottle-fed infants with breastmilk had similar gut bacterial composition (Bray-Curtis index) to infants fed with > 50% and ≤ 50% breastmilk (Table 16). Formula-fed infants displayed a significantly different gut microbial membership and composition compared to the other 5 groups (breast-, bottle-, mixed-feeding, > 50% breastmilk, and ≤ 50% breastmilk) (Figure 17, Table 16).

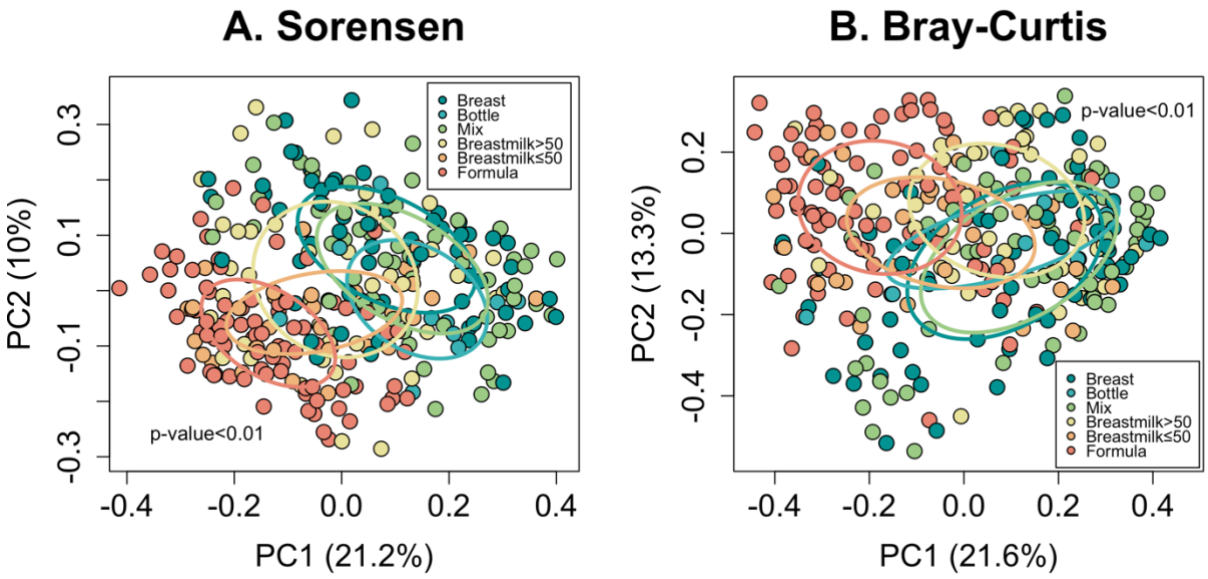

Figure 17. The associations between beta diversity of the gut microbiota and breastfeeding patterns in the past day for exclusively human milk fed infants and dietary intake in the past week for infants fed at least some formula

PERMANOVA was performed to examine the relationships between gut microbiota beta diversity and six feeding groups. All infants in breastfeeding pattern groups were 100% breastmilk fed in the past week. They were excluded from the dietary intake groups. P-value<0.05 is significant.

Table 16. Significant pairwise comparisons of the relationships between beta diversity of the gut microbiota and breastfeeding patterns and breastmilk intake

	Sorensen	**Bray-Curtis**
	Adjusted p-value	**Adjusted p-value**
Breast vs Breastmilk > 50%	0.05	0.02
Bottle vs Breastmilk > 50%	0.05	Not significant
Mixed vs Breastmilk > 50%	0.02	0.02
Breast vs Breastmilk ≤ 50%	0.02	0.02
Bottle vs Breastmilk ≤ 50%	0.03	Not significant
Mixed vs Breastmilk ≤ 50%	0.02	0.02
Breastmilk > 50% vs Breastmilk ≤ 50%	Not significant	0.08
Breastmilk ≤ 50% vs Formula	0.06	0.05
Breastmilk > 50% vs Formula	0.02	0.02
Breast vs Formula	0.02	0.02
Bottle vs Formula	0.02	0.02
Mixed vs Formula	0.02	0.02

Pairwise PERMANOA with FDR correction (BH procedure) was conducted to investigate the associations between two groups in terms of beta diversity. All infants in breastfeeding pattern groups were 100% breastmilk fed in the past week. They were excluded from the dietary intake groups. Adjusted p-value < 0.1 is significant.

4.6.4 The comparisons of the relative abundance of individual taxa in groups

4.6.4.1 Individual taxa and breastfeeding patterns, results from NB

Exclusively breastfed infants fed human milk at the breast had the lowest abundance of *Bifidobacterium* compared to bottle-fed and mixed-fed infants (Table 17). *Enterobacteriaceae* unclassified was more dominant in breastfed infants when compared to bottle-fed and mixed-fed infants (Table 17). The relative abundance of *Bifidobacterium* and *Enterobacteriaceae* unclassified was similar in the bottle and mixed feeding groups (Figure 18). The relative abundance of *Escherichia Shigella* was different across the three breastfeeding patterns, where bottle-fed infants had the highest abundance compared to breastfed and mixed-fed infants (Table

17, Figure 19). There was almost no *Blautia* or *Parabacteroides* present in the guts of bottle-fed infants compared to the other two groups (Table 17). Bottle-fed infants had a higher abundance of *Enterococcus* compared to breastfed infants but similar levels of this bacteria as mixed-fed infants (Table 17, Figure 18).

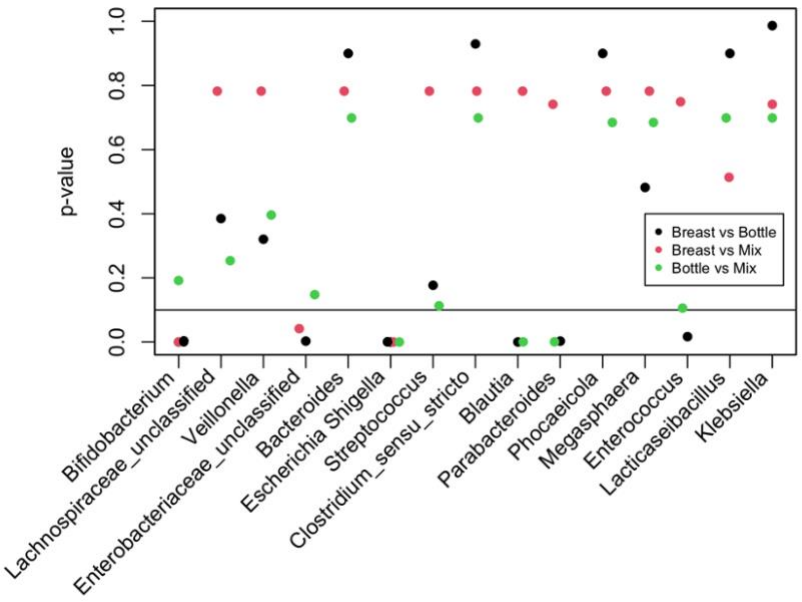

Figure 18. The comparisons of the relative abundance of taxa in three groups of breastfeeding patterns

The top 15 abundant taxa with overall relative abundance >1% were shown in the figure. Negative Binomial Generalized Linear Model was used to compare the relative abundance of taxa between breastfeeding patterns. P-values were FDR corrected with BH procedure. P-value < 0.1 is significant.

Table 17. The relative abundance of taxa in three groups of breastfeeding patterns

Taxa	Overall	Breastfeeding	Bottle feeding	Mixed feeding	*p*-value
Bifidobacterium	29.7 ± 21.5	27.3 ± 20.3a	31.1 ± 22.9b	31.9 ± 22.6b	<0.01*
Lachnospiraceae_unclassified	3.7 ± 8.1	3.7 ± 8.2	8.4 ± 11.4	3 ± 7.3	0.46
Veillonella	10.4 ± 12.4	11.3 ± 12.3	6 ± 8.4	10.4 ± 12.9	0.67
Enterobacteriaceae_unclassified	2.1 ± 11	3.4 ± 15.6a	0.2 ± 0.4b	1 ± 4b	0.01*
Bacteroides	6.5 ± 12.2	6.7 ± 12.9	8.2 ± 12.4	6 ± 11.6	0.90
Escherichia Shigella	9.9 ± 9.4	11.3 ± 10.8a	12.7 ± 9.4b	8 ± 7.5c	<0.01*
Streptococcus	3.8 ± 5.8	3.7 ± 5.7	1.8 ± 3.1	4.3 ± 6.3	0.36
Clostridium_sensu_stricto	6 ± 10.4	5.7 ± 8.5	5.1 ± 6.5	6.6 ± 12.5	0.90
Blautia	1.7 ± 6.8	2 ± 6.8a	0 ± 0b	1.6 ± 7.4a	0.02*
Parabacteroides	1.2 ± 5.1	1 ± 2.7a	0 ± 0b	1.7 ± 7a	0.05*
Phocaeicola	5.1 ± 9.4	5.5 ± 9.7	7.1 ± 12.2	4.3 ± 8.6	0.82
Megasphaera	1.2 ± 6.4	1.6 ± 8.3	0.3 ± 1	1 ± 4.4	0.80
Enterococcus	1.7 ± 3.6	1.2 ± 2.1a	5.1 ± 10.1b	1.6 ± 2.2ab	0.02*
Lacticaseibacillus	1.8 ± 3.4	2.3 ± 3.9	1.7 ± 2.7	1.2 ± 3	0.56
Klebsiella	6.5 ± 12.7	5.1 ± 11.8	5.2 ± 11.7	8 ± 11.8	0.76

The top 15 abundant taxa with overall relative abundance >1% are shown in the table. Negative Binomial Generalized Linear Model was used to compare the relative abundance of taxa between breastfeeding patterns. P-values were FDR corrected with BH procedure. Data is presented as Mean±SD. P-value < 0.1 is significant.

4.6.4.2 Individual taxa and six feeding groups, results from NB

Formula-fed infants had a similar abundance of *Bifidobacterium* with infants fed with breastmilk ≤50% but lower than the rest of the four groups (Table 18). The relative abundance of *Lachnospiraceae* unclassified was similar in infants exclusively fed human milk through a bottle to those fed with breastmilk >50%, breastmilk≤50%, and formula. Infants exclusively fed human milk at the breast had a higher relative abundance of *Enterobacteriaceae* unclassified compared to the other groups. Formula-fed infants had a significantly lower abundance of *Escherichia Shigella* than infants fed with breast, bottle, and mixed. Breastfed, bottle-fed, and mixed-fed infants had a similar abundance of *Streptococcus* to formula-fed infants. *Clostridium sensu stricto* was more prevalent in breastfed and mixed-fed infants as compared to formula-fed

infants. *Blautia* and *Parabacteroides* were the least abundant in bottle-fed infants than the others. Breastfed infants had less abundance of *Enterococcus* than bottle-fed infants but a similar abundance to the other groups. *Lacticaseibacillus* was the least abundant in formula-fed infants in contrast to the other feeding groups.

Table 18. The relative abundance of taxa in six feeding groups, results from NB

	Overall	Breastfeeding	Bottle feeding	Mixed feeding	Breastmilk >50	Breastmilk ≤50	Formula	*p*-value
Bifidobacterium	24.1 ± 19.3	27.3 ± 20.3a	31.1 ± 22.9a	31.9 ± 22.6a	30.6 ± 17.2a	20.4 ± 14.7ab	13.1 ± 11.7b	<0.01*
Lachnospiraceae_unclassified	8.1 ± 10.6	3.7 ± 8.2a	8.4 ± 11.4ab	3 ± 7.3a	8.2 ± 8.7b	9.1 ± 9.4b	14.7 ± 12b	<0.01*
Veillonella	10.6 ± 11.4	11.3 ± 12.3	6 ± 8.4	10.4 ± 12.9	11.4 ± 12.7	13.4 ± 8.5	9.2 ± 9.9	0.66
Enterobacteriaceae_unclassifie	1.2 ± 7.5	3.4 ± 15.6a	0.2 ± 0.4b	1 ± 4b	0.4 ± 1.9b	0.3 ± 0.8b	0.5 ± 1.4b	<0.01*
Bacteroides	6.4 ± 10.6	6.7 ± 12.9	8.2 ± 12.4	6 ± 11.6	6.9 ± 10.4	7.4 ± 10.6	5.6 ± 7.8	0.97
Escherichia Shigella	7.3 ± 8.6	11.3 ± 10.8a	12.7 ± 9.4a	8 ± 7.5a	6.6 ± 9.6ab	6.5 ± 6.6ab	3.8 ± 5.8b	<0.01*
Streptococcus	3.1 ± 5.6	3.7 ± 5.7a	1.8 ± 3.1ab	4.3 ± 6.3a	1.5 ± 2.5b	2.7 ± 6.7ab	3 ± 5.9a	0.01*
Clostridium_sensu_stricto	4 ± 7.9	5.7 ± 8.5a	5.1 ± 6.5ab	6.6 ± 12.5a	2.4 ± 4.5b	2.6 ± 4.8ab	2.1 ± 4.3b	<0.01*
Blautia	3 ± 6.8	2 ± 6.8a	0 ± 0b	1.6 ± 7.4a	1.5 ± 3.4a	4.8 ± 7.6ac	5.3 ± 6.9c	<0.01*
Parabacteroides	1.5 ± 4.9	1 ± 2.7a	0 ± 0b	1.7 ± 7a	1.3 ± 3.2a	2.1 ± 4.7a	1.9 ± 5.3a	0.07*
Phocaeicola	5.6 ± 9.2	5.5 ± 9.7	7.1 ± 12.2	4.3 ± 8.6	4.5 ± 7.5	6.5 ± 10	6.6 ± 9.4	0.91
Megasphaera	2.8 ± 8.7	1.6 ± 8.3ab	0.3 ± 1ab	1 ± 4.4a	7.2 ± 15.7b	1.7 ± 4.5ab	3.3 ± 7.1ab	0.13
Enterococcus	1.7 ± 3.3	1.2 ± 2.1a	5.1 ± 10.1b	1.6 ± 2.2ab	1.2 ± 2.6a	2.4 ± 3.6ab	1.8 ± 2.9ab	0.03*
Lacticaseibacillus	1 ± 2.6	2.3 ± 3.9a	1.7 ± 2.7a	1.2 ± 3a	0.9 ± 1.8a	0.6 ± 1.6a	0.1 ± 0.8b	<0.01*
Klebsiella	4.8 ± 9.5	5.1 ± 11.8ac	5.2 ± 11.7abc	8 ± 13.7a	2.2 ± 4.5b	3.3 ± 4.5ab	4 ± 6.2bc	0.04*

Note that Breastfeeding, Bottle Feeding, and Mixed feeding were all exclusively fed human milk in the week preceding stool sample collection. Those infants in the remaining three groups were fed at least some formula in the week preceding stool sample collection. The top 15 abundant taxa with overall relative abundance>1% were shown in the table. Negative Binomial Generalized Linear Model was used to compare the relative abundance of taxa between breastfeeding patterns. P-values were FDR corrected with BH procedure. Data is presented as Mean±SD. P-value < 0.1 is significant.

4.6.4.3 Individual taxa and six feeding groups, results from MaAsLin

The relative abundance of *Bifidobacterium* was significantly higher in the infants fed with breastmilk > 50% compared to formula-fed infants (Figure 19). Infants fed with breastmilk from the breast, breastmilk from a bottle, breastmilk from both breast and bottle, and breastmilk > 50% had a lower abundance of *Blautia* compared to those fed with formula. Similarly, infants fed breast milk from a bottle or mixed-fed had a lower abundance of *Blautia* than infants fed less than or equal to 50% breastmilk. *Lachnospiraceae* unclassified was lower in breast and mixed-fed infants than in formula-fed infants. Infants fed with breastmilk > 50% or ≤ 50% had a higher relative abundance of *Lachnospiraceae* unclassified than those fed with breast and mixed patterns. Less *Lacticaseibacillus* was present in formula-fed infants than in the other five groups. Breastfed infants had a higher abundance of *Lacticaseibacillus* than those fed with breastmilk > 50% or ≤ 50%. The relative abundance of *Streptococcus* was higher in mixed-fed infants when compared to infants who received breastmilk > 50%.

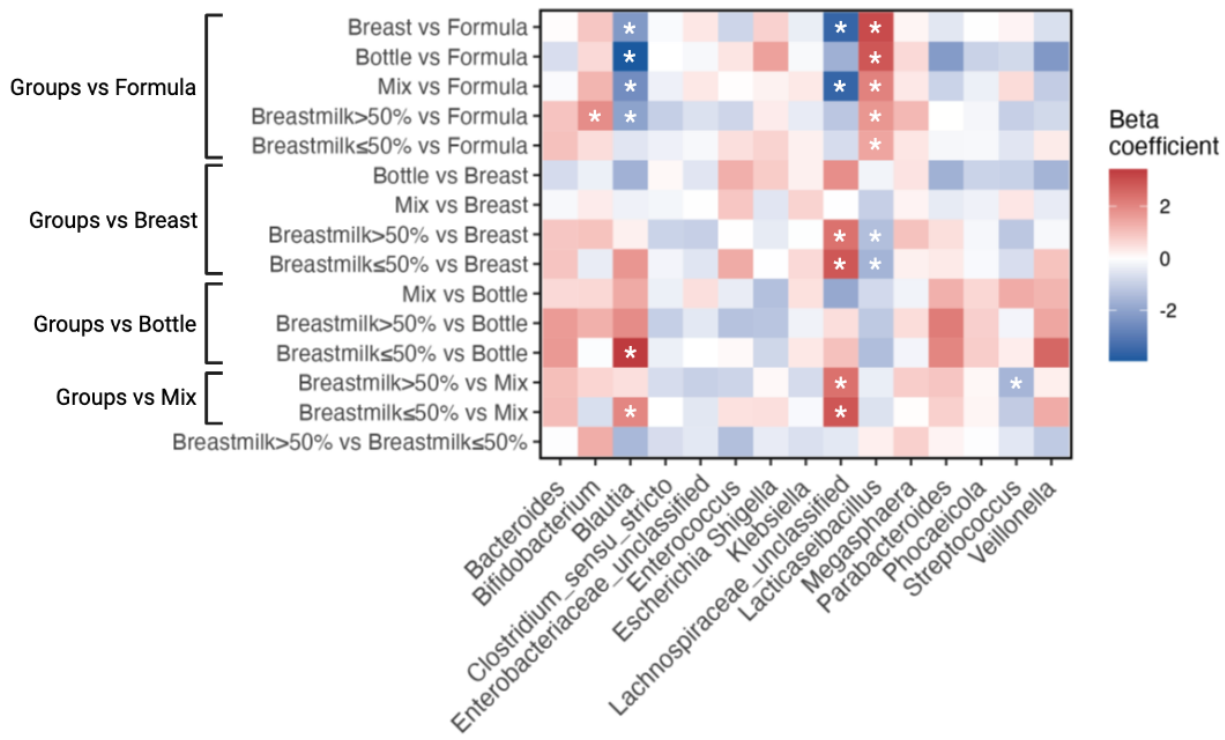

Figure 19. The comparisons of the relative abundance of taxa in six feeding groups, results from MaAsLin

Note that Breastfeeding, Bottle Feeding, and Mixed were all exclusively fed human milk in the week preceding stool sample collection. Those infants in the remaining three groups were fed at least some formula in the week preceding stool sample collection. The top 15 abundant taxa with overall relative abundance >1% were shown in the table. MaAsLin with FDR correction by BH procedure was used to compare the relative abundance of taxa between feeding groups, adjusted by infant sex, infant race, mode of delivery, maternal education level, gestational age at birth, maternal pre-pregnancy BMI, maternal age and antibiotics use since birth. q-value < 0.1 is significant.

4.7 Discussion

The current study demonstrated that the infant gut microbiota differed when fed human milk at the breast, from a bottle, and from both breast and bottle. However, the difference was small compared to the difference in the gut microbiota of infants fed breastmilk, partial breastmilk, and formula. Although bottle-fed infants were 100% breastmilk fed, they have similar microbiota composition with > 50% and ⩽ 50% breastmilk intake infants. These results indicate that the mode of feeding (breastfeeding, bottle feeding, and mixed feeding) and the proportion of breastmilk intake may have an impact on the composition of infant gut

microbiome. The differences in gut microbial richness, diversity, and specific bacterial taxa among the groups could have potential implications for the infant's health and development. It is important to note that these results do not establish a causal relationship between the mode of feeding and gut microbiome composition. Further research is needed to understand the potential health implications of these differences, as well as the factors driving the observed differences in gut microbial composition among the feeding groups.

Breast milk contains a rich microbiota, a potential source of microbes that colonize the infant's gut (Corona-Cervantes et al., 2020; Pannaraj et al., 2017). Previous studies have shown that 68% of the infant gut bacteria within the first six days postpartum originated from human milk among Mexican newborns (Corona-Cervantes et al., 2020). Additionally, microbiota from mother's areolar skin was transferred to exclusively breastfed infants' guts. Breastfed infants had 27.7% of their gut microbiota colonized from breastmilk and 10.4% from areolar skin of their mothers during the first-month life among American infants (Pannaraj et al., 2017). Breastmilk feeding patterns could potentially influence the bacterial transfer from human milk or skin to the infant gut microbiota. In our study, we found that the bottle-fed infants had numerically lower richness (Chao1 index) and diversity (Shannon and inverse Simpson indices) of gut microbiota compared to breastfed infants. However, this difference was not statistically significant. Gut microbiota of breastfed and mix-fed infants had more similar membership (Sorensen) compared to bottle-fed infants. Our results are similar to those reported by Fehr et al., where consumption of pumped milk was associated with depletion of some shared bacteria milk, but they didn't report that there was a significant compositional and taxonomic difference (Fehr et al., 2020). In another study, human milk microbiota in pumped breastmilk was associated with lower alpha diversity (Observed OTUs and inverse Simpson index) compared to manually expressed breast

milk. In the same study, the milk bacterial richness was significantly lower in some indirect breastfeeding compared to all direct breastfeeding (Moossavi et al., 2019). Therefore, it is possibly explained by whether the breastmilk bacteria can remain alive and active during pumping (e.g., sanitating), storing (e.g., freezing, heating, thawing), and bottle feeding (e.g., indirect contact with mothers), which weren't assessed in our study. Additionally, we did not research the associations between human milk bacteria and infant gut microbiota by different breastfeeding patterns. Therefore, future research is needed to investigate how breastfeeding patterns, considering these potential factors mentioned above, would influence and shape the infant gut microbiota.

Human milk oligosaccharides (HMOs) are comprised of complex and unconjugated glycans that are present in human breast milk (Austin & Bénet, 2018; Bode, 2012). They have recognized prebiotics that can promote the growth of beneficial gut microbiota in infants, such as *Bifidobacterium* (Akkerman et al., 2019; Fabiano et al., 2021; Ferro et al., 2021; Karav et al., 2016; Rahman et al., 2023). Regardless of breastfeeding patterns, infants fed with more than 50% breastmilk in the past week had a significantly higher abundance of *Bifidobacterium* than infants fed with less than or equal to 50% breastmilk and formula. This finding was consistent with the previous literature (Hascoët et al., 2011; Ma et al., 2020). Surprisingly, we observed that infants fed with breastmilk from the breast had the lowest abundance of *Bifidobacterium* compared to bottle-fed and mixed-fed infants. A possible explanation might be the compositional changes in breast milk during feeding, as explained next. Foremilk refers to the milk at the beginning of a feed, and it is lower in fat and higher in lactose than hindmilk. Hindmilk is the milk at the end of a feed with higher fats (Gidrewicz & Fenton, 2014; Khan et al., 2013; Slusher et al., 2003). This natural change in milk composition during a single feeding session exposes the

infant's gut to a range of nutrient concentrations and osmolarity levels. As a result, different microenvironments may be created in the infant's gut, which could impact the growth and development of certain gut microbiota. On the other hand, when breast milk is expressed and fed to the infant from a bottle, the foremilk, and hindmilk are mixed together, creating a more uniform milk composition. This means that the infant receives a consistent mixture of nutrients throughout the feeding, with no gradual transition between foremilk and hindmilk, which may affect the infant's gut microbiota differently compared to the gradual transition experienced during direct breastfeeding.

Our study has several strengths. We researched the infant gut microbial variation based on the short-term (one day) breastfeeding patterns and long-term (one week) dietary history and investigated how the variations in infant diet relate to infant gut microbiota composition and diversity. This study offers insights into the fact that, although infants who were fed breastmilk from a bottle in the past day are still considered breast milk-fed, their gut microbial diversity might differ from those who have been fed more than 50% breastmilk in the past week, potentially due to variations in feeding practices. There are some limitations in our study as well. Only 11 out of 136 infants were fed breastmilk from the bottle, and 87.5% were White. This might reduce the statistical power and limit the generalizability of the results. Additionally, we did not collect information on how caregivers sanitized the pumping supplies or bottles, nor did we collect information about how the pumped breastmilk was stored. These factors could potentially affect the milk microbial composition and, thereby, the gut microbial composition and, consequently, influence the results.

4.8 Conclusions

In conclusion, this study provides evidence that the mode of feeding human milk, specifically breastfeeding, bottle feeding, or mixed feeding, may have an impact on the composition of an infant's gut microbiome. We identified variations in the abundance of specific bacterial taxa among the groups, such as *Bifidobacterium*, *Enterobacteriaceae* unclassified, *Escherichia-Shigella*, *Blautia*, and *Parabacteroides*. These results highlight the importance of further research to better understand the potential health implications of these differences and to inform healthcare professionals in providing personalized feeding recommendations for infants that promote optimal gut microbiome development and overall health.

CHAPTER 5: CONCLUSIONS AND FUTURE DIRECTION

5.1 Conclusion

The results of the studies demonstrated the relationships between the infant feeding method in early infancy and neurodevelopmental outcomes in late infancy and how gut microbiota in early infancy mediated this relationship. It also provided evidence on whether breastfeeding patterns (breastmilk fed at breast, breastmilk fed from bottle, and breastmilk fed from both breast and bottle) can shape the infant gut microbiota composition. The following research aims were examined in each chapter (Figure 20):

- Chapter 2 (Aim 1). The associations between infant gut microbiota and neurodevelopmental outcomes.
- Chapter 3 (Aim 2). 1) The associations between infant feeding method and neurodevelopmental outcomes. 2) The relationship between infant feeding method and infant gut microbiota. 3) The mediating role of the infant gut microbiota in the associations of infant feeding method on neurodevelopmental outcomes.
- Chapter 4 (Aim 3). The associations between infant breastfeeding patterns (breast, bottle, and mixed feeding) and infant gut microbiota

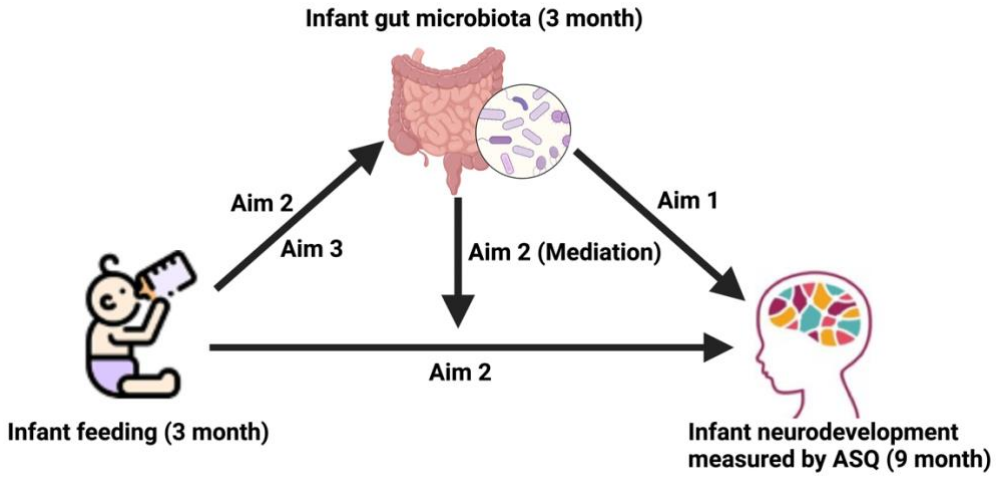

Figure 20. An overview of the study design

The covariates were adjusted in each aim (Table 19):

Table 19. Covariates adjusted in each aim

Questionnaires	Time	Variables used	Aim
MARCH Prenatal 1 Survey questionnaire	During pregnancy	Maternal education level	1,2,3
		Maternal height	1,2,3
		Pre-pregnancy weight	1,2,3
		Maternal age	1,2,3
Birth certificate information	Infants were born	Infant sex	1,2,3
		Estimated weeks of gestation	1,2,3
		Mode of delivery	1,2,3
MARCH 3-month survey	Infants were 3 months of age, before sending fecal collection kit	Infant race	1,2,3
Sample collection form	Fecal collection at 3 months	Infant feeding method in the past 24 hours	1,2
		Infant feeding method in the past week	1,3
		Infant breastfeeding patterns in the past 24 hours	3
		Antibiotics intake since birth	1,2,3
MARCH 9-month survey dictionary	Infants were 9 months of age	Breastfeeding duration	2
		Any breastfeeding duration	
		ASQ	1,2

The following are the most important results:

- A higher Chao 1 index was associated with lower gross motor skills. Shannon index was positively related to problem-solving. The Bray-Curtis dissimilarity matrix was associated with fine motor and communication.
- Infants with gut microbiotas that grouped into Cluster 3 (Bacteroides-dominant) had lower problem-solving scores than those with gut microbiotas that grouped into Cluster 1 (Lachnospiraceae unclassified-dominated).
- Formula-fed infants had more diverse gut microbiota than breastfed infants at 3 months of age.

- Breastfed infants who had been given a vitamin D supplement in the past 24 hours prior to sampling had higher fine motor and communication scores than those exclusively breastfed infants.
- Infants fed formula at 3 months had higher fine motor and communication scores at 9 months compared to those fed exclusive breast milk.
- The Bray-Curtis dissimilarity matrix of gut microbiota at 3 months of age mediated the association between the infant feeding method at 3 months and fine motor scores at 9 months.
- Infant fed exclusively breastmilk from a bottle had lower alpha diversity of the gut microbiota than those fed from breast and both breast and bottle, but it was not statistically significant.
- Infant fed exclusively breastmilk at breast had different gut microbial membership than bottle-fed infants as measured by the Sorensen dissimilarity matrix.
- Infant fed exclusively breastmilk at breast had a lower abundance of Bifidobacterium but the higher abundance of *Enterobacteriaceae*_unclassified compared to bottle- and mixed-fed infants.

The study in Chapter 2 (Aim 1) indicated that the richness measured by Chao 1 index of infant gut microbiota at 3 months was negatively associated with gross motor scores at 9 months. Richness and evenness, as measured by Shannon index of the gut microbiota, were positively associated with problem-solving scores. Bray-Curtis dissimilarity matrix was associated with fine motor and communication scores. These results suggest that the gut microbiota in early life plays a role in cognitive development later in life, which supports the growing body of evidence linking gut microbial diversity to brain development. Additionally, the positive association

between gut microbial diversity and problem-solving scores highlights the importance of maintaining a diverse gut microbiota in early life. This could have implications for developing interventions, such as probiotics or prebiotics, aimed at promoting gut health in infants. Finally, these results suggest that it may be possible to use gut microbiota measures as a predictor of infant development. However, further validation studies are required.

Based on the results presented in Chapter 3 (Aim 2) demonstrated that breastfed infants given a vitamin D supplement in the 24 hours prior to stool sampling had higher fine motor and communication scores than those exclusively breastfed at 3 months, suggesting that supplementing breastfed infants with vitamin D may have a positive impact on infant brain development. However, we did not track the dose of the supplemented vitamin D. Nor did we measure vitamin D consumption status of the infants or their mothers. Therefore, the duration of vitamin D supplementation and the dose of vitamin D supplementation should also be considered to affirm this result. We found that infants fed formula at 3 months had higher fine motor and communication scores at 9 months than those fed exclusive breast milk. This result indicates that formula feeding may positively impact fine motor and communication development in some infants when the formula is provided beginning at 3 months of age specifically.

A mediating role of gut microbiota in the associations between infant feeding method and neurodevelopment was reported by this study. The Bray-Curtis dissimilarity matrix of gut microbiota at 3 months of age mediated the association between the infant feeding method at 3 months and fine motor scores at 9 months, but this mediation disappeared after controlling covariates. This result suggests that gut microbiota in early infancy plays a key role in mediating the impact of feeding practices on some aspects of infant neurodevelopment. In conclusion, this chapter provides insights into the importance of gut health in early life for infant

neurodevelopment. It provides evidence that the gut microbiome may play a key role in mediating the impact of feeding practices on infant neurodevelopment. These results confirm the hypothesis that utilizing nutritional intervention as a new approach to initiate gut microbial colonization in the early stages of infancy has the potential to change neurodevelopmental outcomes in later infancy.

For Chapter 4 (Aim 3), it is important to note that all the infants were exclusively breastmilk-fed in the past day before fecal collection. The results presented in this chapter suggested that, for those infants exclusively fed human milk, the gut microbiota of bottle-fed infants had lower alpha diversity compared to breast- and mixed-fed infants; however, it was not statistically different. Breastfed infants, on the other hand, exhibited distinct gut microbial composition when compared to those who were bottle-fed, as indicated by the Sorensen dissimilarity matrix. Therefore, these results suggest that breastmilk feeding patterns play a crucial role in shaping the gut microbiota of infants, and infants fed human milk via a bottle may impact the richness and composition of the gut microbiota. Additionally, infants fed human milk exclusively at the breast had lower levels of *Bifidobacterium* but higher levels of *Enterobacteriaceae* unclassified than bottle- and mixed-fed infants. It has been studied that *Bifidobacterium* is a beneficial bacteria in infant gut that can help modulate the immune response, strengthen the gut barrier, etc (Stuivenberg et al., 2022). However, some species of *Enterobacteriaceae* are pathogenic (Zhang et al., 2020). Therefore, our results may have important implications for infant health and development. Although this result is opposed to conventional wisdom that breastfeeding from the breast is more beneficial, it highlights the complex relationship between the mode of breastfeeding and gut microbial composition in infants.

5.2 Future directions

In Chapter 2 (Aim 1), we present evidence of associations between infant gut microbiota at 3 months of age and later life neurodevelopment measured by ASQ-3 at 9 months of age. In Chapter 3 (Aim 2), we present evidence for the mediating role of the early-life gut microbiota composition (Bray-Curtis matrix) in the association between infant feeding method at 3 months of age and fine motor scores at 9 months of age. The neurodevelopmental assessment instrument, ASQ-3, used in these experiments, requires parents or caregivers to complete the questionnaire. Thus, there might be some biases by their own perceptions, expectations, and cultural beliefs. Additionally, parents with low socioeconomic status have been shown to over- or underestimate their children's performance on the questions (Feldman et al., 2000). Future study could use another standardized and comprehensive tool to assess infant neurodevelopment to obtain consistent results. The Bayley Scales of Infant and Toddler Development, which requires a trained evaluator to directly interact with the infants and score development using standardized tasks, would be a good option to assess the infant neurodevelopment (Balasundaram & Avulakunta, 2022). For example, a significant association was observed between infant gut microbiota and fine motor skills in 18-month-old full-term infants using Bayley Scales of Infant and Toddler Development, Third Edition (Acuña et al., 2021). In our study, we only assessed the neurodevelopment outcome when the infants were at 9 months of age. We did not extract the clinical diagnosis of neurodevelopment delays or follow the infants longitudinally to check if ASQ accurately captures the neurodevelopment delays. Therefore, conducting a longitudinal study to collect neurodevelopment information at different time points or using medical records of neurodevelopment is needed to analyze the relationship between infant gut microbial development and infant neurodevelopment at different ages to obtain potentially more consistent

results and develop a causal relationship. In our study, we observed that breastfed infants with vitamin supplementation had higher fine motor and communication scores than exclusively breastfed infants based on the parental reports on vitamin D intake. Therefore, future studies should include an accurate vitamin D assessment. For example, collecting infant blood samples and testing the serum or plasma vitamin D levels should be done in the future.

In Chapter 4 (Aim 3), we identified the potential influences of breastfeeding patterns on gut microbial development among exclusively breast milk-fed infants at 3 months of age. Infants were determined to be "exclusively human milk-fed" based on parental reports of infant dietary intake in the past week. We also collected breastfeeding patterns information (breastfeeding at breast, breastfeeding from the bottle and breastfeeding from both breast and bottle) in the past 24 hours before fecal collection. However, based on this information, we can't establish a causal relationship between breastfeeding patterns and infant gut microbiota. It has been shown that breastmilk bacteria can be affected by breastfeeding patterns (Moossavi et al., 2019). Future study can compare the survival rates of live breastmilk bacteria between breastfeeding patterns in combination with the infant gut microbiota. Additionally, other exposures during pumping, such as sanitation for the bottles and pumping supplies and breast milk storage conditions (e.g., heating, freezing, thawing), such as "How did you store the rest of the pumped breastmilk if you pump a lot of milk at once?" and "How often do you sanitize the pumping supplies?", can also be assessed along with the breastfeeding patterns in future work. Through such research, the external bacteria contributed by the three different breastfeeding patterns on the infant gut microbiota will be better understood.

Our study has several strengths. Our longitudinal study of typically developing infants found evidence of a relationship between the gut microbiota during infancy and neurodevelopmental

outcomes later in life. Our study is the first to examine the potential mediating role of early-life gut microbiota in the relationship between infant feeding practices and later neurodevelopmental outcomes. To reduce the potential impact of confounding factors, we excluded pre-term infants known to have delayed neurodevelopment compared to full-term infants. Our study also provides insights into whether breastfeeding patterns can affect the infant gut microbial composition among exclusively breastmilk-fed infants. This study is subject to several limitations that should be taken into consideration. Firstly, the stool samples were stored and transported at room temperature for a day, which may have affected the gut microbiota composition. However, we used stool collection tubes with preservatives that can maintain the gut microbiota composition for up to two weeks at room temperature, reducing the impact of this limitation. Secondly, the ASQ-3 measurements used in this study were parent-reported, possibly introducing some biases in the results. Thirdly, we did not consider exposures at 9 months of age such as the contact with other infants during day care and feeding practices. The limited size of our study sample may result in an inadequate representation of specific participant characteristics, potentially leading to biased results in our analyses. Additionally, we cannot research the "real neurodevelopmental delays" because a majority of the kids were appropriately developing as measured by the ASQ. Further study with a larger and more diverse population will be analyzed when the recruitment of participants is done. Shotgun metagenomics analysis could be conducted to investigate the functions and other organisms of our interest. The findings from the studies included in this dissertation provide a better understanding of the complex relationship between infant feeding practices, gut microbiota, and neurodevelopmental outcomes. Primarily, the work demonstrated that early-life gut microbiota plays a significant role in cognitive development, highlighting the importance of modulation of gut microbiota in early life. Additionally, this research noted that

gut microbiota composition at 3 months of age mediates the association between infant feeding at 3 months of age and fine motor scores at 9 months of age. Finally, this study suggests that breastmilk feeding patterns play a crucial role in shaping the gut microbiota of infants, with distinct gut microbial composition found in infants fed breastmilk from the breast compared to those fed breastmilk from a bottle. Overall, the findings provide important implications for healthcare providers and parents to promote optimal gut health and cognitive development in early life through nutritional intervention and suggest the need for further research to confirm and expand upon these findings.

BIBLIOGRAPHY

Abrahamsson, T. R., Jakobsson, H. E., Andersson, A. F., Björkstén, B., Engstrand, L., & Jenmalm, M. C. (2014). Low gut microbiota diversity in early infancy precedes asthma at school age. *Clin Exp Allergy*, *44*(6), 842-850. https://doi.org/10.1111/cea.12253

Acuña, I., Cerdó, T., Ruiz, A., Torres-Espínola, F. J., López-Moreno, A., Aguilera, M., Suárez, A., & Campoy, C. (2021). Infant Gut Microbiota Associated with Fine Motor Skills. *Nutrients*, *13*(5). https://doi.org/10.3390/nu13051673

Adlerberth, I., Strachan, D. P., Matricardi, P. M., Ahrné, S., Orfei, L., Aberg, N., Perkin, M. R., Tripodi, S., Hesselmar, B., Saalman, R., Coates, A. R., Bonanno, C. L., Panetta, V., & Wold, A. E. (2007). Gut microbiota and development of atopic eczema in 3 European birth cohorts. *J Allergy Clin Immunol*, *120*(2), 343-350. https://doi.org/10.1016/j.jaci.2007.05.018

Ainonen, S., Tejesvi, M. V., Mahmud, M. R., Paalanne, N., Pokka, T., Li, W., Nelson, K. E., Salo, J., Renko, M., Vänni, P., Pirttilä, A. M., & Tapiainen, T. (2022). Antibiotics at birth and later antibiotic courses: effects on gut microbiota. *Pediatric Research*, *91*(1), 154-162. https://doi.org/10.1038/s41390-021-01494-7

Akagawa, S., Tsuji, S., Onuma, C., Akagawa, Y., Yamaguchi, T., Yamagishi, M., Yamanouchi, S., Kimata, T., Sekiya, S. I., Ohashi, A., Hashiyada, M., Akane, A., & Kaneko, K. (2019). Effect of Delivery Mode and Nutrition on Gut Microbiota in Neonates. *Ann Nutr Metab*, *74*(2), 132-139. https://doi.org/10.1159/000496427

Akkerman, R., Faas, M. M., & de Vos, P. (2019). Non-digestible carbohydrates in infant formula as substitution for human milk oligosaccharide functions: Effects on microbiota and gut maturation. *Crit Rev Food Sci Nutr*, *59*(9), 1486-1497. https://doi.org/10.1080/10408398.2017.1414030

Alexy, U., Kersting, M., Sichert-Hellert, W., Manz, F., & Schöch, G. (1999). Macronutrient intake of 3- to 36-month-old German infants and children: results of the DONALD Study. Dortmund Nutritional and Anthropometric Longitudinally Designed Study. *Ann Nutr Metab*, *43*(1), 14-22. https://doi.org/10.1159/000012762

Anderson, J. W., Johnstone, B. M., & Remley, D. T. (1999). Breast-feeding and cognitive development: a meta-analysis. *Am J Clin Nutr*, *70*(4), 525-535. https://doi.org/10.1093/ajcn/70.4.525

Armstrong, J., & Reilly, J. J. (2002). Breastfeeding and lowering the risk of childhood obesity. *The Lancet*, *359*(9322), 2003-2004. https://doi.org/https://doi.org/10.1016/S0140-6736(02)08837-2

Arulkumaran, S., Tusor, N., Chew, A., Falconer, S., Kennea, N., Nongena, P., Hajnal, J. V., Counsell, S. J., Rutherford, M. A., & Edwards, A. D. (2020). MRI Findings at Term-Corrected Age and Neurodevelopmental Outcomes in a Large Cohort of Very Preterm Infants. *AJNR Am J Neuroradiol*, *41*(8), 1509-1516. https://doi.org/10.3174/ajnr.A6666

Austin, S., & Bénet, T. (2018). Quantitative determination of non-lactose milk oligosaccharides. *Anal Chim Acta*, *1010*, 86-96. https://doi.org/10.1016/j.aca.2017.12.036

Azad, M., Konya, T., Persaud, R., Guttman, D., Chari, R., Field, C., Sears, M., Mandhane, P., Turvey, S., Subbarao, P., Becker, A., Scott, J., Kozyrskyj, A., & Investigators, t. C. S. (2016). Impact of maternal intrapartum antibiotics, method of birth and breastfeeding on gut microbiota during the first year of life: a prospective cohort study. *BJOG: An International Journal of Obstetrics & Gynaecology*, *123*(6), 983-993. https://doi.org/https://doi.org/10.1111/1471-0528.13601

Azad, M. B., Vehling, L., Chan, D., Klopp, A., Nickel, N. C., McGavock, J. M., Becker, A. B., Mandhane, P. J., Turvey, S. E., Moraes, T. J., Taylor, M. S., Lefebvre, D. L., Sears, M. R., & Subbarao, P. (2018). Infant Feeding and Weight Gain: Separating Breast Milk From Breastfeeding and Formula From Food. *Pediatrics*, *142*(4). https://doi.org/10.1542/peds.2018-1092

Bäckhed, 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., . . . Wang, J. (2015). Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host Microbe*, *17*(5), 690-703. https://doi.org/10.1016/j.chom.2015.04.004

Balasubramanian, S. (2011). Vitamin D deficiency in breastfed infants & the need for routine vitamin D supplementation. *Indian J Med Res*, *133*(3), 250-252.

Balasundaram, P., & Avulakunta, I. D. (2022). Bayley Scales Of Infant and Toddler Development. In *StatPearls*.

Baumann-Dudenhoeffer, A. M., D'Souza, A. W., Tarr, P. I., Warner, B. B., & Dantas, G. (2018). Infant diet and maternal gestational weight gain predict early metabolic maturation of gut microbiomes. Nat Med, 24(12), 1822-1829. https://doi.org/10.1038/s41591-018-0216-2

Belfort, M. B., Anderson, P. J., Nowak, V. A., Lee, K. J., Molesworth, C., Thompson, D. K., Doyle, L. W., & Inder, T. E. (2016). Breast Milk Feeding, Brain Development, and Neurocognitive Outcomes: A 7-Year Longitudinal Study in Infants Born at Less Than 30 Weeks' Gestation. *The Journal of Pediatrics*, *177*, 133-139.e131. https://doi.org/10.1016/j.jpeds.2016.06.045

Belkaid, Y., & Hand, T. W. (2014). Role of the microbiota in immunity and inflammation. *Cell*, *157*(1), 121-141. https://doi.org/10.1016/j.cell.2014.03.011

Benno, Y., Sawada, K., & Mitsuoka, T. (1984). The intestinal microflora of infants: composition of fecal flora in breast-fed and bottle-fed infants. *Microbiol Immunol*, *28*(9), 975-986. https://doi.org/10.1111/j.1348-0421.1984.tb00754.x

Bernard, J. Y., De Agostini, M., Forhan, A., Alfaiate, T., Bonet, M., Champion, V., Kaminski, M., de Lauzon-Guillain, B., Charles, M. A., & Heude, B. (2013). Breastfeeding duration and cognitive development at 2 and 3 years of age in the EDEN mother-child cohort. *J Pediatr*, *163*(1), 36-42.e31. https://doi.org/10.1016/j.jpeds.2012.11.090

Bian, X., Yao, G., Squires, J., Hoselton, R., Chen, C.-I., Murphy, K., Wei, M., & Fang, B. (2012). Translation and use of parent-completed developmental screening test in Shanghai. *Journal of Early Childhood Research*, *10*(2), 162-175. https://doi.org/10.1177/1476718x11430071

Biasucci, G., Rubini, M., Riboni, S., Morelli, L., Bessi, E., & Retetangos, C. (2010). Mode of delivery affects the bacterial community in the newborn gut. *Early Hum Dev*, *86 Suppl 1*, 13-15. https://doi.org/10.1016/j.earlhumdev.2010.01.004

Björkstén, B., Sepp, E., Julge, K., Voor, T., & Mikelsaar, M. (2001). Allergy development and the intestinal microflora during the first year of life. *J Allergy Clin Immunol*, *108*(4), 516-520. https://doi.org/10.1067/mai.2001.118130

Bode, L. (2012). Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology*, *22*(9), 1147-1162. https://doi.org/10.1093/glycob/cws074

Bokulich, N. A., Chung, J., Battaglia, T., Henderson, N., Jay, M., Li, H., A, D. L., Wu, F., Perez-Perez, G. I., Chen, Y., Schweizer, W., Zheng, X., Contreras, M., Dominguez-Bello, M. G., & Blaser, M. J. (2016). Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Sci Transl Med*, *8*(343), 343ra382. https://doi.org/10.1126/scitranslmed.aad7121

Bonaz, B., Bazin, T., & Pellissier, S. (2018). The Vagus Nerve at the Interface of the Microbiota-Gut-Brain Axis. *Front Neurosci*, *12*, 49. https://doi.org/10.3389/fnins.2018.00049

Bourdeaudhuij, I. D., & Oost, P. V. (2000). Personal and family determinants of dietary behaviour in adolescents and their parents. *Psychology & Health*, *15*(6), 751-770. https://doi.org/10.1080/08870440008405579

Braniste, V., Al-Asmakh, M., Kowal, C., Anuar, F., Abbaspour, A., Tóth, M., Korecka, A., Bakocevic, N., Ng, L. G., Kundu, P., Gulyás, B., Halldin, C., Hultenby, K., Nilsson, H.,

Hebert, H., Volpe, B. T., Diamond, B., & Pettersson, S. (2014). The gut microbiota influences blood-brain barrier permeability in mice. *Sci Transl Med*, *6*(263), 263ra158. https://doi.org/10.1126/scitranslmed.3009759

Brown, A. J., Goldsworthy, S. M., Barnes, A. A., Eilert, M. M., Tcheang, L., Daniels, D., Muir, A. I., Wigglesworth, M. J., Kinghorn, I., Fraser, N. J., Pike, N. B., Strum, J. C., Steplewski, K. M., Murdock, P. R., Holder, J. C., Marshall, F. H., Szekeres, P. G., Wilson, S., Ignar, D. M., . . . Dowell, S. J. (2003). The Orphan G protein-coupled receptors GPR41 and GPR43 are activated by propionate and other short chain carboxylic acids. *J Biol Chem*, *278*(13), 11312-11319. https://doi.org/10.1074/jbc.M211609200

Buchanan, E. M. (2018). MeMoBootR [Computer Program]. Avaliable at: https://github.com/doomlab/MeMoBootR.

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, *3*(1), 1-27. https://doi.org/10.1080/03610927408827101

Carabotti, M., Scirocco, A., Maselli, M. A., & Severi, C. (2015). The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems. *Annals of gastroenterology*, *28*(2), 203-209. https://pubmed.ncbi.nlm.nih.gov/25830558

Carlson, A. L., Xia, K., Azcarate-Peril, M. A., Goldman, B. D., Ahn, M., Styner, M. A., Thompson, A. L., Geng, X., Gilmore, J. H., & Knickmeyer, R. C. (2018). Infant Gut Microbiome Associated With Cognitive Development. Biological psychiatry, 83(2), 148-159. https://doi.org/10.1016/j.biopsych.2017.06.021

Cester, E. A., Bloomfield, F. H., Taylor, J., Smith, S., & Cormack, B. E. (2015). Do recommended protein intakes improve neurodevelopment in extremely preterm babies? *Arch Dis Child Fetal Neonatal Ed*, *100*(3), F243-247. https://doi.org/10.1136/archdischild-2014-307521

Chaudhari, S., & Kadam, S. (2012). Ages and Stages Questionnaire--a developmental screening test. *Indian Pediatr*, *49*(6), 440-441.

Chelimo, C., Camargo, C. A., Jr., Morton, S. M. B., & Grant, C. C. (2020). Association of Repeated Antibiotic Exposure Up to Age 4 Years With Body Mass at Age 4.5 Years. *JAMA Netw Open*, *3*(1), e1917577. https://doi.org/10.1001/jamanetworkopen.2019.17577

Chowdhury, R., Taneja, S., Kvestad, I., Hysing, M., Bhandari, N., & Strand, T. A. (2020). Vitamin D status in early childhood is not associated with cognitive development and linear growth at 6-9 years of age in North Indian children: a cohort study. *Nutr J*, *19*(1), 14. https://doi.org/10.1186/s12937-020-00530-2

Cioni, G., Inguaggiato, E., & Sgandurra, G. (2016). Early intervention in neurodevelopmental disorders: underlying neural mechanisms. *Dev Med Child Neurol*, *58 Suppl 4*, 61-66. https://doi.org/10.1111/dmcn.13050

Clapp, M., Aurora, N., Herrera, L., Bhatia, M., Wilen, E., & Wakefield, S. (2017). Gut microbiota's effect on mental health: The gut-brain axis. *Clin Pract*, *7*(4), 987. https://doi.org/10.4081/cp.2017.987

Clarke, G., Grenham, S., Scully, P., Fitzgerald, P., Moloney, R. D., Shanahan, F., Dinan, T. G., & Cryan, J. F. (2013). The microbiome-gut-brain axis during early life regulates the hippocampal serotonergic system in a sex-dependent manner. *Molecular psychiatry*, *18*(6), 666-673. https://doi.org/10.1038/mp.2012.77

Claud, E. C., & Walker, W. A. (2001). Hypothesis: inappropriate colonization of the premature intestine can cause neonatal necrotizing enterocolitis. *Faseb j*, *15*(8), 1398-1403. https://doi.org/10.1096/fj.00-0833hyp

Clayton, H. B., Li, R., Perrine, C. G., & Scanlon, K. S. (2013). Prevalence and reasons for introducing infants early to solid foods: variations by milk feeding type. *Pediatrics*, *131*(4), e1108-1114. https://doi.org/10.1542/peds.2012-2265

Colombo, L., Crippa, B. L., Consonni, D., Bettinelli, M. E., Agosti, V., Mangino, G., Bezze, E. N., Mauri, P. A., Zanotta, L., Roggero, P., Plevani, L., Bertoli, D., Giannì, M. L., & Mosca, F. (2018). Breastfeeding Determinants in Healthy Term Newborns. *Nutrients*, *10*(1). https://doi.org/10.3390/nu10010048

Corona-Cervantes, K., García-González, I., Villalobos-Flores, L. E., Hernández-Quiroz, F., Piña-Escobedo, A., Hoyo-Vadillo, C., Rangel-Calvillo, M. N., & García-Mena, J. (2020). Human milk microbiota associated with early colonization of the neonatal gut in Mexican newborns. *PeerJ*, *8*, e9205. https://doi.org/10.7717/peerj.9205

Corvaglia, L., Tonti, G., Martini, S., Aceti, A., Mazzola, G., Aloisio, I., Di Gioia, D., & Faldella, G. (2016). Influence of Intrapartum Antibiotic Prophylaxis for Group B Streptococcus on Gut Microbiota in the First Month of Life. *J Pediatr Gastroenterol Nutr*, *62*(2), 304-308. https://doi.org/10.1097/mpg.0000000000000928

Coviello, C., Keunen, K., Kersbergen, K. J., Groenendaal, F., Leemans, A., Peels, B., Isgum, I., Viergever, M. A., de Vries, L. S., Buonocore, G., Carnielli, V. P., & Benders, M. (2018). Effects of early nutrition and growth on brain volumes, white matter microstructure, and neurodevelopmental outcome in preterm newborns. *Pediatr Res*, *83*(1-1), 102-110. https://doi.org/10.1038/pr.2017.227

Del Chierico, F., Vernocchi, P., Petrucca, A., Paci, P., Fuentes, S., Praticò, G., Capuani, G., Masotti, A., Reddel, S., Russo, A., Vallone, C., Salvatori, G., Buffone, E., Signore, F.,

Rigon, G., Dotta, A., Miccheli, A., de Vos, W. M., Dallapiccola, B., & Putignani, L. (2015). Phylogenetic and Metabolic Tracking of Gut Microbiota during Perinatal Development. *PLoS One*, *10*(9), e0137347. https://doi.org/10.1371/journal.pone.0137347

Deoni, S. C., Dean, D. C., 3rd, Piryatinsky, I., O'Muircheartaigh, J., Waskiewicz, N., Lehman, K., Han, M., & Dirks, H. (2013). Breastfeeding and early white matter development: A cross-sectional study. *Neuroimage*, *82*, 77-86. https://doi.org/10.1016/j.neuroimage.2013.05.090

Der, G., Batty, G. D., & Deary, I. J. (2006). Effect of breast feeding on intelligence in children: prospective study, sibling pairs analysis, and meta-analysis. *Bmj*, *333*(7575), 945. https://doi.org/10.1136/bmj.38978.699583.55

Dewey, K. G. (1998). Growth characteristics of breast-fed compared to formula-fed infants. *Biol Neonate*, *74*(2), 94-105. https://doi.org/10.1159/000014016

Differding, M. K., Doyon, M., Bouchard, L., Perron, P., Guérin, R., Asselin, C., Massé, E., Hivert, M. F., & Mueller, N. T. (2020). Potential interaction between timing of infant complementary feeding and breastfeeding duration in determination of early childhood gut microbiota composition and BMI. *Pediatr Obes*, *15*(8), e12642. https://doi.org/10.1111/ijpo.12642

Differding, M. K., & Mueller, N. T. (2020). Human Milk Bacteria: Seeding the Infant Gut? *Cell Host & Microbe*, *28*(2), 151-153. https://doi.org/https://doi.org/10.1016/j.chom.2020.07.017

Dominguez-Bello, M. G., Costello, E. K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., & Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(26), 11971-11975. https://doi.org/10.1073/pnas.1002601107

Donovan, S. M., & Comstock, S. S. (2016). Human Milk Oligosaccharides Influence Neonatal Mucosal and Systemic Immunity. *Ann Nutr Metab*, *69 Suppl 2*(Suppl 2), 42-51. https://doi.org/10.1159/000452818

Dydensborg Sander, S., Nybo Andersen, A. M., Murray, J. A., Karlstad, Ø., Husby, S., & Størdal, K. (2019). Association Between Antibiotics in the First Year of Life and Celiac Disease. *Gastroenterology*, *156*(8), 2217-2229. https://doi.org/10.1053/j.gastro.2019.02.039

Eck, A., Rutten, N. B. M. M., Singendonk, M. M. J., Rijkers, G. T., Savelkoul, P. H. M., Meijssen, C. B., Crijns, C. E., Oudshoorn, J. H., Budding, A. E., & Vlieger, A. M. (2020). Neonatal microbiota development and the effect of early life antibiotics are

determined by two distinct settler types. *PLoS One*, *15*(2), e0228133. https://doi.org/10.1371/journal.pone.0228133

Fabiano, V., Indrio, F., Verduci, E., Calcaterra, V., Pop, T. L., Mari, A., Zuccotti, G. V., Cullu Cokugras, F., Pettoello-Mantovani, M., & Goulet, O. (2021). Term Infant Formulas Influencing Gut Microbiota: An Overview. *Nutrients*, *13*(12). https://doi.org/10.3390/nu13124200

Fallani, M., Amarri, S., Uusijarvi, A., Adam, R., Khanna, S., Aguilera, M., Gil, A., Vieites, J. M., Norin, E., Young, D., Scott, J. A., Doré, J., Edwards, C. A., & The Infabio, T. (2011). Determinants of the human infant intestinal microbiota after the introduction of first complementary foods in infant samples from five European centres. *Microbiology (Reading)*, *157*(Pt 5), 1385-1392. https://doi.org/10.1099/mic.0.042143-0

Farrow, C., Haycraft, E., & Mitchell, G. (2013). Milk Feeding, Solid Feeding, and Obesity Risk: A Review of the Relationships Between Early Life Feeding Practices and Later Adiposity. *Current Obesity Reports*, *2*(1), 58-64. https://doi.org/10.1007/s13679-012-0034-7

Farzi, A., Fröhlich, E. E., & Holzer, P. (2018). Gut Microbiota and the Neuroendocrine System. *Neurotherapeutics*, *15*(1), 5-22. https://doi.org/10.1007/s13311-017-0600-5

Fehr, K., Moossavi, S., Sbihi, H., Boutin, R. C. T., Bode, L., Robertson, B., Yonemitsu, C., Field, C. J., Becker, A. B., Mandhane, P. J., Sears, M. R., Khafipour, E., Moraes, T. J., Subbarao, P., Finlay, B. B., Turvey, S. E., & Azad, M. B. (2020). Breastmilk Feeding Practices Are Associated with the Co-Occurrence of Bacteria in Mothers' Milk and the Infant Gut: the CHILD Cohort Study. *Cell Host & Microbe*, *28*(2), 285-297.e284. https://doi.org/https://doi.org/10.1016/j.chom.2020.06.009

Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Lasky, M., Janosky, J. E., & Paradise, J. L. (2000). Measurement properties of the MacArthur communicative development inventories at ages one and two years. *Child development*, *71*(2), 310-322. https://doi.org/10.1111/1467-8624.00146

Ferro, L. E., Sugino, K. Y., Klepac-Ceraj, V., & Comstock, S. S. (2021). The Abundance of Human Milk Oligosaccharide (HMO)-Metabolizing Genes in Fecal Samples from Six-Month-Old Human Infants. *Microorganisms*, *9*(7). https://doi.org/10.3390/microorganisms9071352

Filteau, S., Rehman, A. M., Yousafzai, A., Chugh, R., Kaur, M., Sachdev, H. P. S., & Trilok-Kumar, G. (2016). Associations of vitamin D status, bone health and anthropometry, with gross motor development and performance of school-aged Indian children who were born at term with low birth weight. *BMJ Open*, *6*(1), e009268. https://doi.org/10.1136/bmjopen-2015-009268

Forbes, J. D., Azad, M. B., Vehling, L., Tun, H. M., Konya, T. B., Guttman, D. S., Field, C. J., Lefebvre, D., Sears, M. R., Becker, A. B., Mandhane, P. J., Turvey, S. E., Moraes, T. J., Subbarao, P., Scott, J. A., & Kozyrskyj, A. L. (2018). Association of Exposure to Formula in the Hospital and Subsequent Infant Feeding Practices With Gut Microbiota and Risk of Overweight in the First Year of Life. *JAMA Pediatr*, *172*(7), e181161. https://doi.org/10.1001/jamapediatrics.2018.1161

Fouhy, F., Ross, R. P., Fitzgerald, G. F., Stanton, C., & Cotter, P. D. (2012). Composition of the early intestinal microbiota: knowledge, knowledge gaps and the use of high-throughput sequencing to address these gaps. *Gut Microbes*, *3*(3), 203-220. https://doi.org/10.4161/gmic.20169

Gao, W., Salzwedel, A. P., Carlson, A. L., Xia, K., Azcarate-Peril, M. A., Styner, M. A., Thompson, A. L., Geng, X., Goldman, B. D., Gilmore, J. H., & Knickmeyer, R. C. (2019). Gut microbiome and brain functional connectivity in infants-a preliminary study focusing on the amygdala. *Psychopharmacology (Berl)*, *236*(5), 1641-1651. https://doi.org/10.1007/s00213-018-5161-8

Gareau, M. G., Wine, E., Rodrigues, D. M., Cho, J. H., Whary, M. T., Philpott, D. J., Macqueen, G., & Sherman, P. M. (2011). Bacterial infection causes stress-induced memory dysfunction in mice. *Gut*, *60*(3), 307-317. https://doi.org/10.1136/gut.2009.202515

Gidrewicz, D. A., & Fenton, T. R. (2014). A systematic review and meta-analysis of the nutrient content of preterm and term breast milk. *BMC Pediatr*, *14*, 216. https://doi.org/10.1186/1471-2431-14-216

Gilmore, J. H., Knickmeyer, R. C., & Gao, W. (2018). Imaging structural and functional brain development in early childhood. *Nature Reviews Neuroscience*, *19*(3), 123-137. https://doi.org/10.1038/nrn.2018.1

Greenwood, C., Morrow, A. L., Lagomarcino, A. J., Altaye, M., Taft, D. H., Yu, Z., Newburg, D. S., Ward, D. V., & Schibler, K. R. (2014). Early empiric antibiotic use in preterm infants is associated with lower bacterial diversity and higher relative abundance of Enterobacter. *J Pediatr*, *165*(1), 23-29. https://doi.org/10.1016/j.jpeds.2014.01.010

Guittar, J., Shade, A., & Litchman, E. (2019). Trait-based community assembly and succession of the infant gut microbiome. *Nature Communications*, *10*(1), 512. https://doi.org/10.1038/s41467-019-08377-w

Guxens, M., Mendez, M. A., Moltó-Puigmartí, C., Julvez, J., García-Esteban, R., Forns, J., Ferrer, M., Vrijheid, M., López-Sabater, M. C., & Sunyer, J. (2011). Breastfeeding, Long-Chain Polyunsaturated Fatty Acids in Colostrum, and Infant Mental Development. *Pediatrics*, *128*(4), e880-e889. https://doi.org/10.1542/peds.2010-1633

Haddad, E. N., Sugino, K. Y., Kerver, J. M., Paneth, N., & Comstock, S. S. (2021). The infant gut microbiota at 12 months of age is associated with human milk exposure but not with maternal pre-pregnancy body mass index or infant BMI-for-age z-scores. *Curr Res Physiol*, *4*, 94-102. https://doi.org/10.1016/j.crphys.2021.03.004

Hadders-Algra, M. (2021). Early Diagnostics and Early Intervention in Neurodevelopmental Disorders-Age-Dependent Challenges and Opportunities. *J Clin Med*, *10*(4). https://doi.org/10.3390/jcm10040861

Hascoët, J. M., Hubert, C., Rochat, F., Legagneur, H., Gaga, S., Emady-Azar, S., & Steenhout, P. G. (2011). Effect of formula composition on the development of infant gut microbiota. *J Pediatr Gastroenterol Nutr*, *52*(6), 756-762. https://doi.org/10.1097/MPG.0b013e3182105850

Heijtz, R. D., Wang, S., Anuar, F., Qian, Y., Björkholm, B., Samuelsson, A., Hibberd, M. L., Forssberg, H., & Pettersson, S. (2011). Normal gut microbiota modulates brain development and behavior. *Proceedings of the National Academy of Sciences*, *108*(7), 3047-3052. https://doi.org/doi:10.1073/pnas.1010529108

Hill, C. J., Lynch, D. B., Murphy, K., Ulaszewska, M., Jeffery, I. B., O'Shea, C. A., Watkins, C., Dempsey, E., Mattivi, F., Tuohy, K., Ross, R. P., Ryan, C. A., PW, O. T., & Stanton, C. (2017). Evolution of gut microbiota composition from birth to 24 weeks in the INFANTMET Cohort. *Microbiome*, *5*(1), 4. https://doi.org/10.1186/s40168-016-0213-y

Hinde, K., & German, J. B. (2012). Food in an evolutionary context: insights from mother's milk. *J Sci Food Agric*, *92*(11), 2219-2223. https://doi.org/10.1002/jsfa.5720

Hoang, D. M., Levy, E. I., & Vandenplas, Y. (2021). The impact of Caesarean section on the infant gut microbiome. *Acta paediatrica (Oslo, Norway : 1992)*, *110*(1), 60-67. https://doi.org/10.1111/apa.15501

Hu, Y. J., & Satten, G. A. (2020). Testing hypotheses about the microbiome using the linear decomposition model (LDM). *Bioinformatics*, *36*(14), 4106-4115. https://doi.org/10.1093/bioinformatics/btaa260

Ismail, I. H., Oppedisano, F., Joseph, S. J., Boyle, R. J., Licciardi, P. V., Robins-Browne, R. M., & Tang, M. L. (2012). Reduced gut microbial diversity in early life is associated with later development of eczema but not atopy in high-risk infants. *Pediatr Allergy Immunol*, *23*(7), 674-681. https://doi.org/10.1111/j.1399-3038.2012.01328.x

Jakobsson, H. E., Abrahamsson, T. R., Jenmalm, M. C., Harris, K., Quince, C., Jernberg, C., Björkstén, B., Engstrand, L., & Andersson, A. F. (2014). Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants

delivered by caesarean section. *Gut*, *63*(4), 559-566. https://doi.org/10.1136/gutjnl-2012-303249

Jari Oksanen, F. G. B., Michael Friendly, Roeland Kindt, Pierre Legendre,, Dan McGlinn, P. R. M., R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry, & H. Stevens, E. S. a. H. W. (2020). *Vegan: Community Ecology Package.* https://CRAN.R-project.org/package=vegan

Jena, A., Montoya, C. A., Mullaney, J. A., Dilger, R. N., Young, W., McNabb, W. C., & Roy, N. C. (2020). Gut-Brain Axis in the Early Postnatal Years of Life: A Developmental Perspective [Review]. *Frontiers in Integrative Neuroscience*, *14*(44). https://doi.org/10.3389/fnint.2020.00044

Jia, Q., Yu, X., Chang, Y., You, Y., Chen, Z., Wang, Y., Liu, B., Chen, L., Ma, D., Xing, Y., & Tong, X. (2022). Dynamic Changes of the Gut Microbiota in Preterm Infants With Different Gestational Age. *Front Microbiol*, *13*, 923273. https://doi.org/10.3389/fmicb.2022.923273

Jones, K. M., Power, M. L., Queenan, J. T., & Schulkin, J. (2015). Racial and ethnic disparities in breastfeeding. *Breastfeed Med*, *10*(4), 186-196. https://doi.org/10.1089/bfm.2014.0152

Julvez, J., Guxens, M., Carsin, A. E., Forns, J., Mendez, M., Turner, M. C., & Sunyer, J. (2014). A cohort study on full breastfeeding and child neuropsychological development: the role of maternal social, psychological, and nutritional factors. *Dev Med Child Neurol*, *56*(2), 148-156. https://doi.org/10.1111/dmcn.12282

Juwita, F., Gumilang, L., Risan, N. A., & Dhamayanti, M. (2021). The Association of Vitamin D and Neurodevelopmental Status Among 2 Years Old Infants. *Global Pediatric Health*, *8*, 2333794X211034075. https://doi.org/10.1177/2333794x211034075

Kalliomäki, M., Collado, M. C., Salminen, S., & Isolauri, E. (2008). Early differences in fecal microbiota composition in children may predict overweight. *Am J Clin Nutr*, *87*(3), 534-538. https://doi.org/10.1093/ajcn/87.3.534

Kalliomäki, M., Kirjavainen, P., Eerola, E., Kero, P., Salminen, S., & Isolauri, E. (2001). Distinct patterns of neonatal gut microflora in infants in whom atopy was and was not developing. *J Allergy Clin Immunol*, *107*(1), 129-134. https://doi.org/10.1067/mai.2001.111237

Karav, S., Le Parc, A., Leite Nobrega de Moura Bell, J. M., Frese, S. A., Kirmiz, N., Block, D. E., Barile, D., & Mills, D. A. (2016). Oligosaccharides Released from Milk Glycoproteins Are Selective Growth Substrates for Infant-Associated Bifidobacteria. *Appl Environ Microbiol*, *82*(12), 3622-3630. https://doi.org/10.1128/aem.00547-16

Kaufman, L. a. R., P.J. (1990). Partitioning Around Medoids (Program PAM). In *Finding Groups in Data* (pp. 68-125). https://doi.org/https://doi.org/10.1002/9780470316801.ch2

Khan, S., Hepworth, A. R., Prime, D. K., Lai, C. T., Trengove, N. J., & Hartmann, P. E. (2013). Variation in fat, lactose, and protein composition in breast milk over 24 hours: associations with infant feeding patterns. *J Hum Lact*, *29*(1), 81-89. https://doi.org/10.1177/0890334412448841

Kitano, N., Nomura, K., Kido, M., Murakami, K., Ohkubo, T., Ueno, M., & Sugimoto, M. (2016). Combined effects of maternal age and parity on successful initiation of exclusive breastfeeding. *Prev Med Rep*, *3*, 121-126. https://doi.org/10.1016/j.pmedr.2015.12.010

Koenig, J. E., Spor, A., Scalfone, N., Fricker, A. D., Stombaugh, J., Knight, R., Angenent, L. T., & Ley, R. E. (2011). Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, *108 Suppl 1*(Suppl 1), 4578-4585. https://doi.org/10.1073/pnas.1000081107

Korpela, K., Salonen, A., Saxen, H., Nikkonen, A., Peltola, V., Jaakkola, T., de Vos, W., & Kolho, K. L. (2020). Antibiotics in early life associate with specific gut microbiota signatures in a prospective longitudinal infant cohort. *Pediatr Res*, *88*(3), 438-443. https://doi.org/10.1038/s41390-020-0761-5

Kostic, A. D., Gevers, D., Siljander, H., Vatanen, T., Hyötyläinen, T., Hämäläinen, A. M., Peet, A., Tillmann, V., Pöhö, P., Mattila, I., Lähdesmäki, H., Franzosa, E. A., Vaarala, O., de Goffau, M., Harmsen, H., Ilonen, J., Virtanen, S. M., Clish, C. B., Orešič, M., . . . Xavier, R. J. (2015). The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe*, *17*(2), 260-273. https://doi.org/10.1016/j.chom.2015.01.001

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol*, *79*(17), 5112-5120. https://doi.org/10.1128/aem.01043-13

Krajmalnik-Brown, R., Ilhan, Z. E., Kang, D. W., & DiBaise, J. K. (2012). Effects of gut microbes on nutrient absorption and energy regulation. *Nutr Clin Pract*, *27*(2), 201-214. https://doi.org/10.1177/0884533611436116

Kramer, M. S., Guo, T., Platt, R. W., Vanilovich, I., Sevkovskaya, Z., Dzikovich, I., Michaelsen, K. F., & Dewey, K. (2004). Feeding effects on growth during infancy. *J Pediatr*, *145*(5), 600-605. https://doi.org/10.1016/j.jpeds.2004.06.069

Kronman, M. P., Zaoutis, T. E., Haynes, K., Feng, R., & Coffin, S. E. (2012). Antibiotic exposure and IBD development among children: a population-based cohort study. *Pediatrics*, *130*(4), e794-803. https://doi.org/10.1542/peds.2011-3886

Kundu, P., Blacher, E., Elinav, E., & Pettersson, S. (2017). Our Gut Microbiome: The Evolving Inner Self. *Cell*, *171*(7), 1481-1493. https://doi.org/10.1016/j.cell.2017.11.024

Le Doare, K., & Heath, P. T. (2013). An overview of global GBS epidemiology. *Vaccine*, *31 Suppl 4*, D7-12. https://doi.org/10.1016/j.vaccine.2013.01.009

Le Huërou-Luron, I., Blat, S., & Boudry, G. (2010). Breast- v. formula-feeding: impacts on the digestive tract and immediate and long-term health effects. *Nutr Res Rev*, *23*(1), 23-36. https://doi.org/10.1017/s0954422410000065

Lee, H., Park, H., Ha, E., Hong, Y. C., Ha, M., Park, H., Kim, B. N., Lee, B., Lee, S. J., Lee, K. Y., Kim, J. H., Jeong, K. S., & Kim, Y. (2016). Effect of Breastfeeding Duration on Cognitive Development in Infants: 3-Year Follow-up Study. *J Korean Med Sci*, *31*(4), 579-584. https://doi.org/10.3346/jkms.2016.31.4.579

Leventakou, V., Roumeliotaki, T., Koutra, K., Vassilaki, M., Mantzouranis, E., Bitsios, P., Kogevinas, M., & Chatzi, L. (2015). Breastfeeding duration and cognitive, language and motor development at 18 months of age: Rhea mother-child cohort in Crete, Greece. *J Epidemiol Community Health*, *69*(3), 232-239. https://doi.org/10.1136/jech-2013-202500

Li, N., Xie, Q., Zhao, L., Shi, J., Evivie, S. E., Lv, X., Huo, G., & Li, B. (2021). Human milk and infant formula modulate the intestinal microbiota and immune systems of human microbiota-associated mice. *Food Funct*, *12*(6), 2784-2798. https://doi.org/10.1039/d0fo03004j

Li, W. Z., Stirling, K., Yang, J. J., & Zhang, L. (2020). Gut microbiota and diabetes: From correlation to causality and mechanism. *World J Diabetes*, *11*(7), 293-308. https://doi.org/10.4239/wjd.v11.i7.293

Liu, B. N., Liu, X. T., Liang, Z. H., & Wang, J. H. (2021). Gut microbiota in obesity. *World J Gastroenterol*, *27*(25), 3837-3850. https://doi.org/10.3748/wjg.v27.i25.3837

Lockyer, F., McCann, S., & Moore, S. E. (2021). Breast Milk Micronutrients and Infant Neurodevelopmental Outcomes: A Systematic Review. *Nutrients*, *13*(11), 3848. https://www.mdpi.com/2072-6643/13/11/3848

Longo, S., Caporali, C., Pisoni, C., Borghesi, A., Perotti, G., Tritto, G., Olivieri, I., La Piana, R., Tonduti, D., Decio, A., Ariaudo, G., Spairani, S., Naboni, C., Gardella, B., Spinillo, A., Manzoni, F., Tinelli, C., Stronati, M., & Orcesi, S. (2021). Neurodevelopmental outcome

of preterm very low birth weight infants admitted to an Italian tertiary center over an 11-year period. *Scientific Reports*, *11*(1), 16316. https://doi.org/10.1038/s41598-021-95864-0

Loughman, A., Ponsonby, A. L., O'Hely, M., Symeonides, C., Collier, F., Tang, M. L. K., Carlin, J., Ranganathan, S., Allen, K., Pezic, A., Saffery, R., Jacka, F., Harrison, L. C., Sly, P. D., & Vuillermin, P. (2020). Gut microbiota composition during infancy and subsequent behavioural outcomes. *EBioMedicine*, *52*, 102640. https://doi.org/10.1016/j.ebiom.2020.102640

Luczynski, P., McVey Neufeld, K.-A., Oriach, C. S., Clarke, G., Dinan, T. G., & Cryan, J. F. (2016). Growing up in a Bubble: Using Germ-Free Animals to Assess the Influence of the Gut Microbiota on Brain and Behavior. *International Journal of Neuropsychopharmacology*, *19*(8). https://doi.org/10.1093/ijnp/pyw020

Ma, J., Li, Z., Zhang, W., Zhang, C., Zhang, Y., Mei, H., Zhuo, N., Wang, H., Wang, L., & Wu, D. (2020). Comparison of gut microbiota in exclusively breast-fed and formula-fed babies: a study of 91 term infants. *Sci Rep*, *10*(1), 15792. https://doi.org/10.1038/s41598-020-72635-x

Ma, T., Bu, S., Paneth, N., Kerver, J. M., & Comstock, S. S. (2022). Vitamin D Supplementation in Exclusively Breastfed Infants Is Associated with Alterations in the Fecal Microbiome. *Nutrients*, *14*(1). https://doi.org/10.3390/nu14010202

Mallick, H., Rahnavard, A., McIver, L. J., Ma, S., Zhang, Y., Nguyen, L. H., Tickle, T. L., Weingart, G., Ren, B., Schwager, E. H., Chatterjee, S., Thompson, K. N., Wilkinson, J. E., Subramanian, A., Lu, Y., Waldron, L., Paulson, J. N., Franzosa, E. A., Bravo, H. C., & Huttenhower, C. (2021). Multivariable association discovery in population-scale meta-omics studies. *PLOS Computational Biology*, *17*(11), e1009442. https://doi.org/10.1371/journal.pcbi.1009442

Marcobal, A., Barboza, M., Froehlich, J. W., Block, D. E., German, J. B., Lebrilla, C. B., & Mills, D. A. (2010). Consumption of human milk oligosaccharides by gut-related microbes. *J Agric Food Chem*, *58*(9), 5334-5340. https://doi.org/10.1021/jf9044205

Martin, C. R., Ling, P. R., & Blackburn, G. L. (2016). Review of Infant Feeding: Key Features of Breast Milk and Infant Formula. *Nutrients*, *8*(5). https://doi.org/10.3390/nu8050279

Martin, R., Makino, H., Cetinyurek Yavuz, A., Ben-Amor, K., Roelofs, M., Ishikawa, E., Kubota, H., Swinkels, S., Sakai, T., Oishi, K., Kushiro, A., & Knol, J. (2016). Early-Life Events, Including Mode of Delivery and Type of Feeding, Siblings and Gender, Shape the Developing Gut Microbiota. *PLoS One*, *11*(6), e0158498. https://doi.org/10.1371/journal.pone.0158498

Matijašić, M., Meštrović, T., Paljetak, H., Perić, M., Barešić, A., & Verbanac, D. (2020). Gut Microbiota beyond Bacteria-Mycobiome, Virome, Archaeome, and Eukaryotic Parasites in IBD. *Int J Mol Sci*, *21*(8). https://doi.org/10.3390/ijms21082668

Mazzola, G., Murphy, K., Ross, R. P., Di Gioia, D., Biavati, B., Corvaglia, L. T., Faldella, G., & Stanton, C. (2016). Early Gut Microbiota Perturbations Following Intrapartum Antibiotic Prophylaxis to Prevent Group B Streptococcal Disease. *PLoS One*, *11*(6), e0157527. https://doi.org/10.1371/journal.pone.0157527

McKeen, S., Roy, N. C., Mullaney, J. A., Eriksen, H., Lovell, A., Kussman, M., Young, W., Fraser, K., Wall, C. R., & McNabb, W. C. (2022). Adaptation of the infant gut microbiome during the complementary feeding transition. *PLoS One*, *17*(7), e0270213. https://doi.org/10.1371/journal.pone.0270213

Mitchell, C. M., Mazzoni, C., Hogstrom, L., Bryant, A., Bergerat, A., Cher, A., Pochan, S., Herman, P., Carrigan, M., Sharp, K., Huttenhower, C., Lander, E. S., Vlamakis, H., Xavier, R. J., & Yassour, M. (2020). Delivery Mode Affects Stability of Early Infant Gut Microbiota. *Cell Rep Med*, *1*(9), 100156. https://doi.org/10.1016/j.xcrm.2020.100156

Moore, M. R., Schrag, S. J., & Schuchat, A. (2003). Effects of intrapartum antimicrobial prophylaxis for prevention of group-B-streptococcal disease on the incidence and ecology of early-onset neonatal sepsis. *Lancet Infect Dis*, *3*(4), 201-213. https://doi.org/10.1016/s1473-3099(03)00577-2

Moossavi, S., & Azad, M. B. (2020). Origins of human milk microbiota: new evidence and arising questions. *Gut Microbes*, *12*(1), 1667722. https://doi.org/10.1080/19490976.2019.1667722

Moossavi, S., Sepehri, S., Robertson, B., Bode, L., Goruk, S., Field, C. J., Lix, L. M., de Souza, R. J., Becker, A. B., Mandhane, P. J., Turvey, S. E., Subbarao, P., Moraes, T. J., Lefebvre, D. L., Sears, M. R., Khafipour, E., & Azad, M. B. (2019). Composition and Variation of the Human Milk Microbiota Are Influenced by Maternal and Early-Life Factors. *Cell Host & Microbe*, *25*(2), 324-335.e324. https://doi.org/10.1016/j.chom.2019.01.011

Morley, R., Fewtrell, M. S., Abbott, R. A., Stephenson, T., MacFadyen, U., & Lucas, A. (2004). Neurodevelopment in children born small for gestational age: a randomized trial of nutrient-enriched versus standard formula and comparison with a reference breastfed group. *Pediatrics*, *113*(3 Pt 1), 515-521. https://doi.org/10.1542/peds.113.3.515

Mueller, N. T., Shin, H., Pizoni, A., Werlang, I. C., Matte, U., Goldani, M. Z., Goldani, H. A. S., & Dominguez-Bello, M. G. (2016). Birth mode-dependent association between pre-pregnancy maternal weight status and the neonatal intestinal microbiome. *Scientific Reports*, *6*(1), 23133. https://doi.org/10.1038/srep23133

Munyaka, P. M., Khafipour, E., & Ghia, J. E. (2014). External influence of early childhood establishment of gut microbiota and subsequent health implications. *Front Pediatr*, *2*, 109. https://doi.org/10.3389/fped.2014.00109

Murk, W., Risnes, K. R., & Bracken, M. B. (2011). Prenatal or early-life exposure to antibiotics and risk of childhood asthma: a systematic review. *Pediatrics*, *127*(6), 1125-1138. https://doi.org/10.1542/peds.2010-2092

Nelson, C. A., 3rd, Zeanah, C. H., Fox, N. A., Marshall, P. J., Smyke, A. T., & Guthrie, D. (2007). Cognitive recovery in socially deprived young children: the Bucharest Early Intervention Project. *Science (New York, N.Y.)*, *318*(5858), 1937-1940. https://doi.org/10.1126/science.1143921

Neu, J., & Rushing, J. (2011). Cesarean versus vaginal delivery: long-term infant outcomes and the hygiene hypothesis. *Clin Perinatol*, *38*(2), 321-331. https://doi.org/10.1016/j.clp.2011.03.008

Neufeld, K. M., Kang, N., Bienenstock, J., & Foster, J. A. (2011). Reduced anxiety-like behavior and central neurochemical change in germ-free mice. *Neurogastroenterol Motil*, *23*(3), 255-264, e119. https://doi.org/10.1111/j.1365-2982.2010.01620.x

Nogacka, A., Salazar, N., Suárez, M., Milani, C., Arboleya, S., Solís, G., Fernández, N., Alaez, L., Hernández-Barranco, A. M., de Los Reyes-Gavilán, C. G., Ventura, M., & Gueimonde, M. (2017). Impact of intrapartum antimicrobial prophylaxis upon the intestinal microbiota and the prevalence of antibiotic resistance genes in vaginally delivered full-term neonates. *Microbiome*, *5*(1), 93. https://doi.org/10.1186/s40168-017-0313-3

O'Sullivan, A., Farver, M., & Smilowitz, J. T. (2015). The Influence of Early Infant-Feeding Practices on the Intestinal Microbiome and Body Composition in Infants. *Nutr Metab Insights*, *8*(Suppl 1), 1-9. https://doi.org/10.4137/nmi.S29530

Oliphant, K., & Allen-Vercoe, E. (2019). Macronutrient metabolism by the human gut microbiome: major fermentation by-products and their impact on host health. *Microbiome*, *7*(1), 91. https://doi.org/10.1186/s40168-019-0704-8

Onyszkiewicz, M., Jaworska, K., & Ufnal, M. (2020). Short chain fatty acids and methylamines produced by gut microbiota as mediators and markers in the circulatory system. *Experimental Biology and Medicine*, *245*(2), 166-175. https://doi.org/10.1177/1535370219900898

Pan, J.-X., Deng, F.-L., Zeng, B.-H., Zheng, P., Liang, W.-W., Yin, B.-M., Wu, J., Dong, M.-X., Luo, Y.-Y., Wang, H.-Y., Wei, H., & Xie, P. (2019). Absence of gut microbiota during early life affects anxiolytic Behaviors and monoamine neurotransmitters system in the

hippocampal of mice. *Journal of the Neurological Sciences*, *400*, 160-168. https://doi.org/https://doi.org/10.1016/j.jns.2019.03.027

Pang, W. W., Bernard, J. Y., Thavamani, G., Chan, Y. H., Fok, D., Soh, S. E., Chua, M. C., Lim, S. B., Shek, L. P., Yap, F., Tan, K. H., Gluckman, P. D., Godfrey, K. M., van Dam, R. M., Kramer, M. S., & Chong, Y. S. (2017). Direct vs. Expressed Breast Milk Feeding: Relation to Duration of Breastfeeding. *Nutrients*, *9*(6). https://doi.org/10.3390/nu9060547

Pannaraj, P. S., Li, F., Cerini, C., Bender, J. M., Yang, S., Rollie, A., Adisetiyo, H., Zabih, S., Lincez, P. J., Bittinger, K., Bailey, A., Bushman, F. D., Sleasman, J. W., & Aldrovandi, G. M. (2017). Association Between Breast Milk Bacterial Communities and Establishment and Development of the Infant Gut Microbiome. *JAMA Pediatr*, *171*(7), 647-654. https://doi.org/10.1001/jamapediatrics.2017.0378

Penders, J., Thijs, C., van den Brandt, P. A., Kummeling, I., Snijders, B., Stelma, F., Adams, H., van Ree, R., & Stobberingh, E. E. (2007). Gut microbiota composition and development of atopic manifestations in infancy: the KOALA Birth Cohort Study. *Gut*, *56*(5), 661-667. https://doi.org/10.1136/gut.2006.100164

Penders, J., Thijs, C., Vink, C., Stelma, F. F., Snijders, B., Kummeling, I., van den Brandt, P. A., & Stobberingh, E. E. (2006). Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics*, *118*(2), 511-521. https://doi.org/10.1542/peds.2005-2824

Pérez-Escamilla, R., Tomori, C., Hernández-Cordero, S., Baker, P., Barros, A. J. D., Bégin, F., Chapman, D. J., Grummer-Strawn, L. M., McCoy, D., Menon, P., Ribeiro Neves, P. A., Piwoz, E., Rollins, N., Victora, C. G., & Richter, L. (2023). Breastfeeding: crucially important, but increasingly challenged in a market-driven world. *The Lancet*, *401*(10375), 472-485. https://doi.org/https://doi.org/10.1016/S0140-6736(22)01932-8

Plagemann, A., & Harder, T. (2005). Breast feeding and the risk of obesity and related metabolic diseases in the child. *Metab Syndr Relat Disord*, *3*(3), 222-232. https://doi.org/10.1089/met.2005.3.222

Qiu, P., Ishimoto, T., Fu, L., Zhang, J., Zhang, Z., & Liu, Y. (2022). The Gut Microbiota in Inflammatory Bowel Disease. *Front Cell Infect Microbiol*, *12*, 733992. https://doi.org/10.3389/fcimb.2022.733992

Questionnaires, A. S. (2022). *ASQ-3*. https://agesandstages.com/products-pricing/asq3/

Rahman, T., Sarwar, P. F., Potter, C., Comstock, S. S., & Klepac-Ceraj, V. (2023). Role of human milk oligosaccharide metabolizing bacteria in the development of atopic dermatitis/eczema [Review]. *Frontiers in Pediatrics*, *11*. https://doi.org/10.3389/fped.2023.1090048

Raju, T. N. (2011). Breastfeeding is a dynamic biological process--not simply a meal at the breast. *Breastfeed Med*, *6*(5), 257-259. https://doi.org/10.1089/bfm.2011.0081

Ratsika, A., Cruz Pereira, J. S., Lynch, C. M. K., Clarke, G., & Cryan, J. F. (2023). Microbiota-immune-brain interactions: A lifespan perspective. *Current Opinion in Neurobiology*, *78*, 102652. https://doi.org/https://doi.org/10.1016/j.conb.2022.102652

Ren, Q., Li, K., Sun, H., Zheng, C., Zhou, Y., Lyu, Y., Ye, W., Shi, H., Zhang, W., Xu, Y., & Jiang, S. (2022). The Association of Formula Protein Content and Growth in Early Infancy: A Systematic Review and Meta-Analysis. *Nutrients*, *14*(11). https://doi.org/10.3390/nu14112255

Reyman, M., van Houten, M. A., Watson, R. L., Chu, M. L. J. N., Arp, K., de Waal, W. J., Schiering, I., Plötz, F. B., Willems, R. J. L., van Schaik, W., Sanders, E. A. M., & Bogaert, D. (2022). Effects of early-life antibiotics on the developing infant gut microbiome and resistome: a randomized trial. *Nature Communications*, *13*(1), 893. https://doi.org/10.1038/s41467-022-28525-z

Rinne, M., Kalliomaki, M., Arvilommi, H., Salminen, S., & Isolauri, E. (2005). Effect of probiotics and breastfeeding on the bifidobacterium and lactobacillus/enterococcus microbiota and humoral immune responses. *J Pediatr*, *147*(2), 186-191. https://doi.org/10.1016/j.jpeds.2005.03.053

Rodríguez, J. M., Murphy, K., Stanton, C., Ross, R. P., Kober, O. I., Juge, N., Avershina, E., Rudi, K., Narbad, A., Jenmalm, M. C., Marchesi, J. R., & Collado, M. C. (2015). The composition of the gut microbiota throughout life, with an emphasis on early life. *Microb Ecol Health Dis*, *26*, 26050. https://doi.org/10.3402/mehd.v26.26050

Rozé, J.-C., Ancel, P.-Y., Marchand-Martin, L., Rousseau, C., Montassier, E., Monot, C., Le Roux, K., Butin, M., Resche-Rigon, M., Aires, J., Neu, J., Lepage, P., Butel, M.-J., & Group, f. t. E. S. (2020). Assessment of Neonatal Intensive Care Unit Practices and Preterm Newborn Gut Microbiota and 2-Year Neurodevelopmental Outcomes. *JAMA Network Open*, *3*(9), e2018119-e2018119. https://doi.org/10.1001/jamanetworkopen.2020.18119

Saari, A., Virta, L. J., Sankilampi, U., Dunkel, L., & Saxen, H. (2015). Antibiotic exposure in infancy and risk of being overweight in the first 24 months of life. *Pediatrics*, *135*(4), 617-626. https://doi.org/10.1542/peds.2014-3407

Salas, A. A., Woodfin, T., Phillips, V., Peralta-Carcelen, M., Carlo, W. A., & Ambalavanan, N. (2018). Dose-Response Effects of Early Vitamin D Supplementation on Neurodevelopmental and Respiratory Outcomes of Extremely Preterm Infants at 2 Years of Age: A Randomized Trial. *Neonatology*, *113*(3), 256-262. https://doi.org/10.1159/000484399

Saliaj, A. Z., Majlinda; Pura, Bruna (2015). Long-Term Effect of Breastfeeding in Preschooler's Psychomotor Development. *International Journal of Medical and Health Sciences*, *108*.

Saturio, S., Nogacka, A. M., Alvarado-Jasso, G. M., Salazar, N., de Los Reyes-Gavilán, C. G., Gueimonde, M., & Arboleya, S. (2021). Role of Bifidobacteria on Infant Health. *Microorganisms*, *9*(12). https://doi.org/10.3390/microorganisms9122415

Savino, F., Cresi, F., Pautasso, S., Palumeri, E., Tullio, V., Roana, J., Silvestro, L., & Oggero, R. (2004). Intestinal microflora in breastfed colicky and non-colicky infants. *Acta paediatrica (Oslo, Norway : 1992)*, *93*(6), 825-829.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J., & Weber, C. F. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol*, *75*(23), 7537-7541. https://doi.org/10.1128/aem.01541-09

Schrag, S. J., Zywicki, S., Farley, M. M., Reingold, A. L., Harrison, L. H., Lefkowitz, L. B., Hadler, J. L., Danila, R., Cieslak, P. R., & Schuchat, A. (2000). Group B streptococcal disease in the era of intrapartum antibiotic prophylaxis. *N Engl J Med*, *342*(1), 15-20. https://doi.org/10.1056/nejm200001063420103

Schwarzenberg, S. J., & Georgieff, M. K. (2018). Advocacy for Improving Nutrition in the First 1000 Days to Support Childhood Development and Adult Health. *Pediatrics*, *141*(2). https://doi.org/10.1542/peds.2017-3716

Shore, R. M., & Chesney, R. W. (2013). Rickets: Part I. *Pediatr Radiol*, *43*(2), 140-151. https://doi.org/10.1007/s00247-012-2532-x

Siller, M., Morgan, L., Turner-Brown, L., Baggett, K. M., Baranek, G. T., Brian, J., Bryson, S. E., Carter, A. S., Crais, E. R., Estes, A., Kasari, C., Landa, R. J., Lord, C., Messinger, D. S., Mundy, P., Odom, S. L., Reznick, J. S., Roberts, W., Rogers, S. J., . . . Zwaigenbaum, L. (2013). Designing Studies to Evaluate Parent-Mediated Interventions for Toddlers With Autism Spectrum Disorder. *Journal of Early Intervention*, *35*(4), 355-377. https://doi.org/10.1177/1053815114542507

Singh, S. B., Madan, J., Coker, M., Hoen, A., Baker, E. R., Karagas, M. R., & Mueller, N. T. (2020). Does birth mode modify associations of maternal pre-pregnancy BMI and gestational weight gain with the infant gut microbiome? *International Journal of Obesity*, *44*(1), 23-32. https://doi.org/10.1038/s41366-018-0273-0

Slusher, T., Hampton, R., Bode-Thomas, F., Pam, S., Akor, F., & Meier, P. (2003). Promoting the exclusive feeding of own mother's milk through the use of hindmilk and increased

maternal milk volume for hospitalized, low birth weight infants (< 1800 grams) in Nigeria: a feasibility study. *J Hum Lact*, *19*(2), 191-198. https://doi.org/10.1177/0890334403252490

Sordillo, J. E., Korrick, S., Laranjo, N., Carey, V., Weinstock, G. M., Gold, D. R., O'Connor, G., Sandel, M., Bacharier, L. B., Beigelman, A., Zeiger, R., Litonjua, A. A., & Weiss, S. T. (2019). Association of the Infant Gut Microbiome With Early Childhood Neurodevelopmental Outcomes: An Ancillary Study to the VDAART Randomized Clinical Trial. *JAMA Netw Open*, *2*(3), e190905. https://doi.org/10.1001/jamanetworkopen.2019.0905

Squires J, T. E., Bricker D, Potter L. (2009). *ASQ-3 User's Guide*. http://archive.brookespublishing.com/documents/asq-3-technical-report.pdf

Stanislawski, M. A., Dabelea, D., Wagner, B. D., Sontag, M. K., Lozupone, C. A., & Eggesbø, M. (2017). Pre-pregnancy weight, gestational weight gain, and the gut microbiota of mothers and their infants. *Microbiome*, *5*(1), 113. https://doi.org/10.1186/s40168-017-0332-0

Stephens, B. E., Walden, R. V., Gargus, R. A., Tucker, R., McKinley, L., Mance, M., Nye, J., & Vohr, B. R. (2009). First-Week Protein and Energy Intakes Are Associated With 18-Month Developmental Outcomes in Extremely Low Birth Weight Infants. *Pediatrics*, *123*(5), 1337-1343. https://doi.org/10.1542/peds.2008-0211

Stevens, E. E., Patrick, T. E., & Pickler, R. (2009). A history of infant feeding. *J Perinat Educ*, *18*(2), 32-39. https://doi.org/10.1624/105812409x426314

Stewart, C. J., Ajami, N. J., O'Brien, J. L., Hutchinson, D. S., Smith, D. P., Wong, M. C., Ross, M. C., Lloyd, R. E., Doddapaneni, H., Metcalf, G. A., Muzny, D., Gibbs, R. A., Vatanen, T., Huttenhower, C., Xavier, R. J., Rewers, M., Hagopian, W., Toppari, J., Ziegler, A. G., . . . Petrosino, J. F. (2018). Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature*, *562*(7728), 583-588. https://doi.org/10.1038/s41586-018-0617-x

Stuivenberg, G. A., Burton, J. P., Bron, P. A., & Reid, G. (2022). Why Are Bifidobacteria Important for Infants? *Microorganisms*, *10*(2). https://doi.org/10.3390/microorganisms10020278

Sudo, N., Sawamura, S., Tanaka, K., Aiba, Y., Kubo, C., & Koga, Y. (1997). The requirement of intestinal bacterial flora for the development of an IgE production system fully susceptible to oral tolerance induction. *J Immunol*, *159*(4), 1739-1745.

Sugino, K. Y., Ma, T., Kerver, J. M., Paneth, N., & Comstock, S. S. (2021). Human Milk Feeding Patterns at 6 Months of Age are a Major Determinant of Fecal Bacterial

Diversity in Infants. *J Hum Lact*, *37*(4), 703-713. https://doi.org/10.1177/0890334420957571

Sugino, K. Y., Ma, T., Paneth, N., & Comstock, S. S. (2021). Effect of Environmental Exposures on the Gut Microbiota from Early Infancy to Two Years of Age. *Microorganisms*, *9*(10). https://doi.org/10.3390/microorganisms9102140

Sugino, K. Y., Paneth, N., & Comstock, S. S. (2019). Michigan cohorts to determine associations of maternal pre-pregnancy body mass index with pregnancy and infant gastrointestinal microbial communities: Late pregnancy and early infancy. *PLoS One*, *14*(3), e0213733. https://doi.org/10.1371/journal.pone.0213733

Tamana, S. K., Tun, H. M., Konya, T., Chari, R. S., Field, C. J., Guttman, D. S., Becker, A. B., Moraes, T. J., Turvey, S. E., Subbarao, P., Sears, M. R., Pei, J., Scott, J. A., Mandhane, P. J., & Kozyrskyj, A. L. (2021). Bacteroides-dominant gut microbiome of late infancy is associated with enhanced neurodevelopment. *Gut Microbes*, *13*(1), 1-17. https://doi.org/10.1080/19490976.2021.1930875

Temple Newhook, J., Newhook, L. A., Midodzi, W. K., Murphy Goodridge, J., Burrage, L., Gill, N., Halfyard, B., & Twells, L. (2017). Poverty and Breastfeeding: Comparing Determinants of Early Breastfeeding Cessation Incidence in Socioeconomically Marginalized and Privileged Populations in the FiNaL Study. *Health Equity*, *1*(1), 96-102. https://doi.org/10.1089/heq.2016.0028

Thigpen, M. C., Whitney, C. G., Messonnier, N. E., Zell, E. R., Lynfield, R., Hadler, J. L., Harrison, L. H., Farley, M. M., Reingold, A., Bennett, N. M., Craig, A. S., Schaffner, W., Thomas, A., Lewis, M. M., Scallan, E., & Schuchat, A. (2011). Bacterial meningitis in the United States, 1998-2007. *N Engl J Med*, *364*(21), 2016-2025. https://doi.org/10.1056/NEJMoa1005384

Thursby, E., & Juge, N. (2017). Introduction to the human gut microbiota. *Biochemical Journal*, *474*(11), 1823-1836. https://doi.org/10.1042/bcj20160510

Timby, N., Domellöf, E., Hernell, O., Lönnerdal, B., & Domellöf, M. (2014). Neurodevelopment, nutrition, and growth until 12 mo of age in infants fed a low-energy, low-protein formula supplemented with bovine milk fat globule membranes: a randomized controlled trial. *The American Journal of Clinical Nutrition*, *99*(4), 860-868. https://doi.org/10.3945/ajcn.113.064295

Urbaniak, C., Angelini, M., Gloor, G. B., & Reid, G. (2016). Human milk microbiota profiles in relation to birthing method, gestation and infant gender. *Microbiome*, *4*(1), 1. https://doi.org/10.1186/s40168-015-0145-y

Vaher, K., Bogaert, D., Richardson, H., & Boardman, J. P. (2022). Microbiome-gut-brain axis in brain development, cognition and behavior during infancy and early childhood. *Developmental Review*, *66*, 101038. https://doi.org/https://doi.org/10.1016/j.dr.2022.101038

Veena, S. R., Krishnaveni, G. V., Srinivasan, K., Wills, A. K., Hill, J. C., Kurpad, A. V., Muthayya, S., Karat, S. C., Nalinakshi, M., & Fall, C. H. (2010). Infant feeding practice and childhood cognitive performance in South India. *Arch Dis Child*, *95*(5), 347-354. https://doi.org/10.1136/adc.2009.165159

Veereman-Wauters, G., Staelens, S., Van de Broek, H., Plaskie, K., Wesling, F., Roger, L. C., McCartney, A. L., & Assam, P. (2011). Physiological and bifidogenic effects of prebiotic supplements in infant formulae. *J Pediatr Gastroenterol Nutr*, *52*(6), 763-771. https://doi.org/10.1097/MPG.0b013e3182139f39

Victora, C. G., Morris, S. S., Barros, F. C., Horta, B. L., Weiderpass, E., & Tomasi, E. (1998). Breast-feeding and growth in Brazilian infants. *Am J Clin Nutr*, *67*(3), 452-458. https://doi.org/10.1093/ajcn/67.3.452

Warner, B. B., Deych, E., Zhou, Y., Hall-Moore, C., Weinstock, G. M., Sodergren, E., Shaikh, N., Hoffmann, J. A., Linneman, L. A., Hamvas, A., Khanna, G., Rouggly-Nickless, L. C., Ndao, I. M., Shands, B. A., Escobedo, M., Sullivan, J. E., Radmacher, P. G., Shannon, W. D., & Tarr, P. I. (2016). Gut bacteria dysbiosis and necrotising enterocolitis in very low birthweight infants: a prospective case-control study. *Lancet*, *387*(10031), 1928-1936. https://doi.org/10.1016/s0140-6736(16)00081-7

Weiss, P. P. (2005). The storage of breast milk. International Children Medical Research Association. https://alzahrahosp.tbzmed.ac.ir/Uploads/User/26/container%20of%20BM.pdf.

Wu, W., Zhao, A., Liu, B., Ye, W.-H., Su, H.-W., Li, J., & Zhang, Y.-M. (2021). Neurodevelopmental Outcomes and Gut Bifidobacteria in Term Infants Fed an Infant Formula Containing High sn-2 Palmitate: A Cluster Randomized Clinical Trial. *Nutrients*, *13*(2), 693. https://www.mdpi.com/2072-6643/13/2/693

Yates, N. J., Tesic, D., Feindel, K. W., Smith, J. T., Clarke, M. W., Wale, C., Crew, R. C., Wharfe, M. D., Whitehouse, A. J. O., & Wyrwoll, C. S. (2018). Vitamin D is crucial for maternal care and offspring social behaviour in rats. *J Endocrinol*, *237*(2), 73-85. https://doi.org/10.1530/joe-18-0008

Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R. N., Anokhin, A. P., Heath, A. C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J. G., Lozupone, C. A., Lauber, C., Clemente, J. C.,

Knights, D., . . . Gordon, J. I. (2012). Human gut microbiome viewed across age and geography. *Nature*, *486*(7402), 222-227. https://doi.org/10.1038/nature11053

Zhang, C., Li, L., Jin, B., Xu, X., Zuo, X., Li, Y., & Li, Z. (2021). The Effects of Delivery Mode on the Gut Microbiota and Health: State of Art. *Front Microbiol*, *12*, 724449. https://doi.org/10.3389/fmicb.2021.724449

Zhang, H., Chen, Y., Wang, Z., Xie, G., Liu, M., Yuan, B., Chai, H., Wang, W., & Cheng, P. (2022). Implications of Gut Microbiota in Neurodegenerative Diseases. *Front Immunol*, *13*, 785644. https://doi.org/10.3389/fimmu.2022.785644

Zhang, Z., Li, D., Shi, X., Zhai, Y., Guo, Y., Zheng, Y., Zhao, L., He, Y., Chen, Y., Wang, Z., Su, J., Kang, Y., & Gao, Z. (2020). Genomic characterization of an emerging Enterobacteriaceae species: the first case of co-infection with a typical pathogen in a human patient. *BMC Genomics*, *21*(1), 297. https://doi.org/10.1186/s12864-020-6720-z

Zven, S. E., Susi, A., Mitre, E., & Nylund, C. M. (2020). Association Between Use of Multiple Classes of Antibiotic in Infancy and Allergic Disease in Childhood. *JAMA Pediatr*, *174*(2), 199-200. https://doi.org/10.1001/jamapediatrics.2019.4794

APPENDIX A: IRB APPROVAL LETTER

MICHIGAN STATE
UNIVERSITY

Modification and Continuing Review APPROVAL
Pre-2018 Common Rule

March 27, 2023

To: Nigel S Paneth

Re: **MSU Study ID:** LEGACY16-1429M
IRB: Biomedical and Health Inst. Review Board (BIRB)
Principal Investigator: Nigel S Paneth
Category: Expedited 2(b), 3, 5, 7
Submission: Modification and Continuing Review MODCR00001186
Submission Approval Date: 3/24/2023
Effective Date: 3/24/2023
Study Expiration Date: 3/23/2024

Title: Prenatal Exposures and Child Health Outcomes: A Statewide Study (CGA# 149003, 151506)

This submission has been approved by the Michigan State University (MSU) Biomedical and Health Inst. Review Board (BIRB). The submission was reviewed by the Institutional Review Board (IRB) through the Non-Committee Review procedure. The IRB has found that this study protects the rights and welfare of human subjects and meets the requirements of MSU's Federal Wide Assurance (FWA00004556) and the federal regulations for the protection of human subjects in research (e.g., pre-2018 45 CFR 46, 28 CFR 46, 21 CFR 50, 56, other applicable regulations).

This letter notes that the study is closed to new accrual and this approval is for patient follow-up reporting only. Any further new recruitment or contact with new subjects will require IRB review and approval via a modification before implementation.

This letter notes approval for the social media cards, CHARM communications and thank you card, Prenatal 3 survey, Toenail Questionnaire, instructions, collection protocol, and communication scripts, and Data Abstraction Form.

How to Access Final Documents
To access the study's final materials, including those approved by the IRB such as consent forms, recruitment materials, and the approved protocol, if applicable, please log into the Click™ Research Compliance System, open the study's workspace, and view the "Documents" tab. To obtain consent form(s) stamped with the IRB watermark, select the "Final" PDF version of your consent form(s) as applicable in the "Documents" tab. Please note that the consent form(s) stamped with the IRB watermark must typically be used.

Office of Regulatory Affairs
Human Research Protection Program

4000 Collins Road
Suite 136
Lansing, MI 48910

517-355-2180
Fax: 517-432-4503
Email: irb@msu.edu
www.hrpp.msu.edu

Continuing Review: IRB approval is valid until the expiration date listed above. If the research continues to involve human subjects, you must submit a Continuing Review request at least one month before expiration.

Modifications: Any proposed change or modification with certain limited exceptions discussed below must be reviewed and approved by the IRB prior to implementation of the change. Please submit a Modification request to have the changes reviewed. If changes are made at the time of continuing review, please submit a Modification and Continuing Review request.

New Funding: If new external funding is obtained to support this study, a Modification request must be submitted for IRB review and approval before new funds can be spent on human research activities, as the new funding source may have additional or different requirements.

Immediate Change to Eliminate a Hazard: When an immediate change in a research protocol is necessary to eliminate a hazard to subjects, the proposed change need not be reviewed by the IRB prior to its implementation. In such situations, however, investigators must report the change in protocol to the IRB immediately thereafter.

Reportable Events: Certain events require reporting to the IRB. These include:

- Potential unanticipated problems that may involve risks to subjects or others
- Potential noncompliance
- Subject complaints
- Protocol deviations or violations
- Unapproved change in protocol to eliminate a hazard to subjects
- Premature suspension or termination of research
- Audit or inspection by a federal or state agency
- New potential conflict of interest of a study team member
- Written reports of study monitors
- Emergency use of investigational drugs or devices
- Any activities or circumstances that affect the rights and welfare of research subjects
- Any information that could increase the risk to subjects

Please report new information through the study's workspace and contact the IRB office with any urgent events. Please visit the Human Research Protection Program (HRPP) website to obtain more information, including reporting timelines.

Personnel Changes: Key study personnel must be listed on the MSU IRB application for expedited and full board studies and any changes to key study personnel must to be submitted as modifications. Although only key study personnel need to be listed on a non-exempt application, all other individuals engaged in human subject research activities must receive and maintain current human subject training, must disclose conflict of interest, and are subject to MSU HRPP requirements. It is the responsibility of the Principal Investigator (PI) to

maintain oversight over all study personnel and to assure and to maintain appropriate tracking that these requirements are met (e.g. documentation of training completion, conflict of interest). When non-MSU personnel are engaged in human research, there are additional requirements. See HRPP Manual Section 4-10, Designation as Key Project Personnel on Non-Exempt IRB Projects for more information.

Prisoner Research: If a human subject involved in ongoing research becomes a prisoner during the course of the study and the relevant research proposal was not reviewed and approved by the IRB in accordance with the requirements for research involving prisoners under subpart C of 45 CFR part 46, the investigator must promptly notify the IRB.

Site Visits: The MSU HRPP Compliance office conducts post approval site visits for certain IRB approved studies. If the study is selected for a site visit, you will be contacted by the HRPP Compliance office to schedule the site visit.

For Studies that Involve Consent, Parental Permission, or Assent Form(s):

Use of IRB Approved Form: Investigators must use the form(s) approved by the IRB and must typically use the form with the IRB watermark.

Copy Provided to Subjects: A copy of the form(s) must be provided to the individual signing the form. In some instances, that individual must be provided with a copy of the signed form (e.g. studies following ICH-GCP E6 requirements). Assent forms should be provided as required by the IRB.

Record Retention: All records relating to the research must be appropriately managed and retained. This includes records under the investigator's control, such as the informed consent document. Investigators must retain copies of signed forms or oral consent records (e.g., logs). Investigators must retain all pages of the form, not just the signature page. Investigators may not attempt to de-identify the form; it must be retained with all original information. The PI must maintain these records for a minimum of three years after the IRB has closed the research and a longer retention period may be required by law, contract, funding agency, university requirement or other requirements for certain studies, such as those that are sponsored or FDA regulated research. See HRPP Manual Section 4-7-A, Recordkeeping for Investigators, for more information.

Closure: If the research activities no longer involve human subjects, please submit a Continuing Review request, through which study closure may be requested. Human subject research activities are complete if there is no further interactions or interventions with human subjects and/or no further analysis of identifiable private information.

For More Information: See the HRPP Manual (available at hrpp.msu.edu).

Contact Information: If we can be of further assistance or if you have questions, please contact us at 517-355-2180 or via email at IRB@msu.edu. Please visit hrpp.msu.edu to access the HRPP Manual, templates, etc.

Expedited Category. Please see the appropriate research category below for the full regulatory text.

Expedited 1. Clinical studies of drugs and medical devices only when condition (a) or (b) is met.
(a) Research on drugs for which an investigational new drug application (21 CFR Part 312) is not required. (Note: Research on marketed drugs that significantly increases the risks or decreases the acceptability of the risks associated with the use of the product is not eligible for expedited review.)
(b) Research on medical devices for which (i) an investigational device exemption application (21 CFR Part 812) is not required; or (ii) the medical device is cleared/approved for marketing and the medical device is being used in accordance with its cleared/approved labeling.

Expedited 2. Collection of blood samples by finger stick, heel stick, ear stick, or venipuncture as follows:
(a) from healthy, nonpregnant adults who weigh at least 110 pounds. For these subjects, the amounts drawn may not exceed 550 ml in an 8 week period and collection may not occur more frequently than 2 times per week; or
(b) from other adults and children, considering the age, weight, and health of the subjects, the collection procedure, the amount of blood to be collected, and the frequency with which it will be collected. For these subjects, the amount drawn may not exceed the lesser of 50 ml or 3 ml per kg in an 8 week period and collection may not occur more frequently than 2 times per week.

Expedited 3. Prospective collection of biological specimens for research purposes by noninvasive means.
Examples: (a) hair and nail clippings in a nondisfiguring manner; (b) deciduous teeth at time of exfoliation or if routine patient care indicates a need for extraction; (c) permanent teeth if routine patient care indicates a need for extraction; (d) excreta and external secretions (including sweat); (e) uncannulated saliva collected either in an unstimulated fashion or stimulated by chewing gumbase or wax or by applying a dilute citric solution to the tongue; (f) placenta removed at delivery; (g) amniotic fluid obtained at the time of rupture of the membrane prior to or during labor; (h) supra- and subgingival dental plaque and calculus, provided the collection procedure is not more invasive than routine prophylactic scaling of the teeth and the process is accomplished in accordance with accepted prophylactic techniques; (i) mucosal and skin cells collected by buccal scraping or swab, skin swab, or mouth washings; (j) sputum collected after saline mist nebulization.

Expedited 4. Collection of data through noninvasive procedures (not involving general anesthesia or sedation) routinely employed in clinical practice, excluding procedures involving x-rays or microwaves. Where medical devices are employed, they must be cleared/approved for marketing. (Studies intended to evaluate the safety and effectiveness of the medical device are not generally eligible for

expedited review, including studies of cleared medical devices for new indications.) Examples: (a) physical sensors that are applied either to the surface of the body or at a distance and do not involve input of significant amounts of energy into the subject or an invasion of the subject's privacy; (b) weighing or testing sensory acuity; (c) magnetic resonance imaging; (d) electrocardiography, electroencephalography, thermography, detection of naturally occurring radioactivity, electroretinography, ultrasound, diagnostic infrared imaging, doppler blood flow, and echocardiography; (e) moderate exercise, muscular strength testing, body composition assessment, and flexibility testing where appropriate given the age, weight, and health of the individual.

Expedited 5. Research involving materials (data, documents, records, or specimens) that have been collected, or will be collected solely for nonresearch purposes (such as medical treatment or diagnosis). (NOTE: Some research in this category may be exempt from the HHS regulations for the protection of human subjects. 45 CFR 46.101(b)(4). This listing refers only to research that is not exempt.)

Expedited 6. Collection of data from voice, video, digital, or image recordings made for research purposes.

Expedited 7. Research on individual or group characteristics or behavior (including, but not limited to, research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies. (NOTE: Some research in this category may be exempt from the HHS regulations for the protection of human subjects. 45 CFR 46.101(b)(2) and (b)(3). This listing refers only to research that is not exempt.)

Expedited 8. Continuing review of research previously approved by the convened IRB as follows:
(a) where (i) the research is permanently closed to the enrollment of new subjects; (ii) all subjects have completed all research-related interventions; and (iii) the research remains active only for long-term follow-up of subjects; or
(b) where no subjects have been enrolled and no additional risks have been identified; or
(c) where the remaining research activities are limited to data analysis.

Expedited 9. Continuing review of research, not conducted under an investigational new drug application or investigational device exemption where categories two (2) through eight (8) do not apply but the IRB has determined and documented at a convened meeting that the research involves no greater than minimal risk and no additional risks have been identified.

APPENDIX B: CONSENT FORMS

MICHIGAN ARCHIVE FOR RESEARCH IN CHILD HEALTH

RECORD OF CONSENT FOR PARTICIPATION

Participant's Name:

_____

Study Name: Michigan Archive for Research in Child Health

Investigator's Name: Nigel Paneth, MD MPH

Investigator's Phone Number: 517-844-3961 or 1-833-242-7687

Investigator Address: 909 Wilson Rd. Rm 218, East Lansing, MI 48824

Funding Sources: National Institutes of Health (NIH) & Michigan Health Endowment Fund (MHEF)

You are being asked whether you and your child will participate in a research study taking place across Michigan called M-ARCH (Michigan Archive for Research in Child Health). This study is led by a group called CHARM (Child Health Advances from Research with Mothers) which involves researchers from Michigan State University, the University of Michigan, Wayne State University, Henry Ford Health System, and the Michigan Department of Health and Human Services (MDHHS). MARCH is part of a nationwide research study, the ECHO (Environmental influences on Child Health Outcomes) program, which aims to understand the earliest causes of childhood diseases, including causes that may start before children are born.

We are asking you to join the ECHO Program to help understand how things that happen early in children's lives – even before they are born – affect their development, health, and wellbeing. This research program includes about 200 locations in the US. The ECHO Program will combine information from about 50,000 children and their families. With so many participants from many parts of the US, researchers can answer questions that the MARCH study cannot answer alone. The MARCH study hopes to enroll at least 1,100 participants.

Why is this study being done?

We know that some factors in the environment during pregnancy and early childhood, such as lead, can affect a child's health and development. But there is much we do not know. By getting information now, while you are pregnant, we can find out whether factors such as diet, genes, environmental chemicals, infections, hormones, and more might lead to illnesses in children such as asthma, obesity, or problems in physical, intellectual, or social development. The goal of MARCH is to identify these factors, so that we can prevent them from causing illness in children. The mission of ECHO is also to improve the health of children for generations to come. At the same time, we want to learn about the problems and concerns of pregnant women in our state and prevent illness in women too.

What does this study involve?

We will go over each component of the MARCH study with you, but briefly, you will participate in the MARCH and ECHO studies for at least 6 years. We will interview you both during and after your pregnancy. If in the future we cannot get in contact with you, we may use social media and/or other public records to help us keep your contact up to date. We will reserve portions of the samples routinely collected throughout your prenatal care and use them for this research study. Additionally, we will collect samples from you and your child, such as toenails, hair, and shed teeth, as well as information from the MDHHS and your medical records. We would like to share specimens and information that you give MARCH with the other scientists in the ECHO program. Your information could be very helpful to scientists who are trying to solve important health problems facing women

and children. This information may include variables such as your child's development and behavior, medical history and family history, social interactions, and diet. It may also include information about you such as your health and diet during pregnancy, or things that may cause stress in your life. Any study information that is shared with other researchers outside of our research team, including biological samples, will be stripped of most identifying information by giving it a code to protect your privacy. In doing so, we will assign a code that allows us to identify the material but would not allow the scientists who receive this material to do so. The only identifying information we will share will be your addresses, your and your baby's dates of birth and other information including race, sex, gender, language, dates of procedures, collections, and health information. This information is required to answer some research questions, such as linking information about your child's samples and health to information about air or water quality where your child lives or goes to school, but we will take many precautions to safeguard your privacy. All ECHO researchers are protected by a Certificate of Confidentiality in which investigators shall not disclose the name or other identifying information about a participant to any federal, State, or local civil, criminal, administrative, legislative, or other proceeding, without the specific consent of the individual to who the information pertains. This certificate is described below. We will provide financial compensation in recognition of the time and effort it takes you to participate in the study.

What will my child and I be asked to do?

In order for you to participate, we will need to you to provide us with your name, contact information, and the hospital where you plan to deliver. Your participation is voluntary, and for that reason you may refuse to be in the study or stop taking part in this study at any time without penalty.

The section below describes all other components of the study in detail and then ask you to sign to consent to participate. You have the option to refuse participation in any of these collections or questions.

- Urine from the samples you give to your doctor during your prenatal visits to be collected and stored.
- Extra blood (6-8 teaspoons or 30-40ml) will be collected when you have your blood drawn for your prenatal labs.
- Post-delivery, your placenta will be collected, examined, and stored once no longer needed by the delivery hospital. We will let the hospital know that you are a part of the MARCH study.
- Collect samples of your and your child's toenails, hair, and urine.
- Your social security number will be collected to see your baby's birth certificate. This will allow us to make sure we have the right baby's certificate.
- Access to Michigan Department of Health and Human Services registries and program data. These registries and program may include the Michigan Care Improvement Registry for your and your child's vaccination status, the Michigan Birth Defects Registry, the Michigan Newborn Screening program, and the Early Hearing Detection Intervention program, as well as other programs and registries housed in the Michigan Department of Health and Human Services.
- A signed HIPAA form will be filled out to review all portions of your and your baby's hospital records related to this pregnancy, birth and postpartum period.
- When your child is around 3 months old we would like a sample of your baby's poop. This can be done from the privacy of your own home.
- We would like you to send us some of your child's baby teeth as they naturally lose them. This usually happens between the ages of 5-10 years old.
- MARCH will contact you for at least two prenatal surveys, including the one you will complete today or over the phone at a better time for you. After your child is born, we will contact you at

least once a year and ask you to complete phone and online surveys or to set up appointments to meet with you in your home. Topics of the surveys include items such as you and your child's health, home environment, diet, and sleep.

- Some researchers in the ECHO program would like to look at environmental factors by neighborhood. To do this they would need your address. To understand when these factors could have impacted you, they would need your and your child's date of birth.

How long will the MARCH and ECHO research programs last?

The MARCH and ECHO programs will last until 2023, and may continue after that. MARCH and ECHO will store your and your child's information and samples for an unlimited period of time, so researchers can use them in future health research.

What if I decide not to be a part of this study?

You have the right to refuse to be in the study, to refuse to do any part of the study, or to stop at any time without penalty or loss of benefits to which you would otherwise be entitled and without affecting your present or future medical care. You can also decide to withdraw any of your specimens or information that have not been used. Information and biospecimens that have already been distributed for research will not be retrieved. If you decide to do any of these things, please contact the Principal Investigator, Dr. Nigel Paneth, in writing, by phone, or by email. You can send a letter to Dr. Nigel Paneth, Michigan State University, Department of Epidemiology, 909 West Fee Hall, East Lansing MI 48824. You can call him at 517-844-3961 or contact him by email at paneth@msu.edu.

What about my confidentiality?

To avoid having any information about you or your child being used in ways that might discriminate against you or stigmatize you or your family, all of the information collected in the MARCH study is strictly confidential. Your confidentiality and that of your child will be protected to the maximum extent allowed by law, and we will protect it in the ways we will explain. There is no way, however, to make it impossible for unauthorized people to identify you. All the researchers and research staff working with your specimens or information who do not have valid access to your identity have promised not to try and identify you, and will be removed from the investigative team and barred from participating in this research if they try to do so. To protect your confidentiality only the MARCH research staff will see your real name. We will store your sensitive information (for example, your social security number) separately from the rest of the information you provide us and it will be kept in a secured, locked computer servers that only our MARCH research team can access. The MARCH study and the ECHO Data Analysis Center at John Hopkins University and RTI International will maintain ECHO research information. Data and samples that are shared with other researchers will be labeled with a code. The key that links your name to this code will be kept securely by us, and not provided to other researchers. You may have provided information about illegal drug use, and it is also possible that the biological specimens you provide could be tested for illegal drugs. We promise you that we will strictly limit the way this information is used so that researchers who want to study the effects of drug use don't have access to any information that identifies you, will not analyze this information with any identifiers you provide such as your name, address or date of birth. Our local MARCH research team follows this rule, and no one outside our local study can analyze the data we collect without signing an agreement that they will follow this rule as well.

There are also staff members at MSU who oversee research (Human Research Protection

Program) and individuals who fund this research who may see your name and identifiers to be sure that they correctly identify your/your child's blood spot and to ensure that the MARCH project is properly conducting research. Laws help protect your and your baby's genetic information and, in most cases, make it illegal to use genetic information to discriminate against you and your child for health insurance coverage and employment. These laws do not apply to other types of insurance (such as life, disability, or long-term care).

This research is covered by a Certificate of Confidentiality from the National Institutes of Health. The researchers with this Certificate may not disclose or use information, documents, or biospecimens that may identify you in any federal, state, or local civil, criminal, administrative, legislative, or other action, suit or proceeding, or be used as evidence, for example, if there is a court subpoena, unless you have consented for this use. Information, documents, or biospecimens protected by this Certificate cannot be disclosed to anyone else who is not connected with the research, except if there is a federal, state or local law that requires disclosure (such as to report child abuse or communicable diseases, but not for federal, state, or local civil, criminal, administrative, legislative, or other proceedings, see below); if you have consented to the disclosure, including for your medical treatment; or if it is used for other scientific research, as allowed by federal regulation protecting research subjects.

The Certificate of Confidentiality will not be used to prevent disclosure as required by federal, state, or local law of child abuse and neglect or harm to self or others.

You should understand that a Certificate of Confidentiality does not prevent you from voluntarily releasing information about yourself or your involvement in this research. If you want your research information released to an insurer, medical care provider, or any other person not connected with the research, you must provide consent to allow the researchers to release it. Researchers will share summaries of ECHO analyses through scientific articles or other public scientific resources, such as NIH or ECHO databases. We will not publicly share any participant's individual information.

What are the risks or costs to my child and me?

Because of the nature of genomic data, the risks of loss of confidentiality may extend beyond the individual participant to their families, and subgroups of people or populations and general.

There is only a very small risk to your confidentiality because of the measures we have taken to protect your data that we have explained, and participation in this study is free. If there is a breach in confidentiality, information about you and your child may be used to discriminate against you.

Will my child and I benefit from this study?

By being a part of this study, you will help answer questions about how to improve the health of children and mothers. You and your child will not receive medical care or other direct benefits from being in this study. Taking part in ECHO will not improve you or your child's health right now, nor will it change anything about your current medical care. You likely will not directly benefit from this study, however your participation may help scientists and doctors all over the United States learn if there are ways to prevent pregnancy and childhood health issues.

Will I receive any compensation?

You will be compensated for your time for participation in the study. Compensation will come in the form of a check made out in your name or a gift card mailed to your current address. If you consent to the collection and storage of portions of the samples collected by your doctor as a part of your normal standard care, you will receive $10. If you participate in other parts of the study,

you will receive more compensation over the course of your participation. If you participate in all parts of this project, you will receive at least $600 worth of compensation over the next 6 years. For Michigan State University to process and mail a check, the accounts payable department will need your name and address information. After the check is mailed to you, your name will be removed on all further documentation in accounts payable.

Will I have access to the information in my MARCH study record?

MARCH and ECHO will store your and your child's information and samples for an unlimited period of time, so researchers can use them in future health research. From time to time, we will make study results available to all ECHO participants through the ECHO website, newsletters, community presentations, and scientific papers. These results will not be specific to any individual person in ECHO, including you and your child. If the researchers see results they believe are very important to your or your child's health or medical care, we will give you a report with the information and an explanation of what each result means. We will also let you know if we think you should share the results with a doctor or other health professional.

If important new findings come up during the course of the study that might change your decision to be in this study, we will give you information about those findings as soon as possible. MARCH and ECHO are research studies and therefore do not provide medical care. You should always talk to your doctor if you have questions or concerns about you, your pregnancy and/or your child's health. If you would like access to any of your own MARCH study information or have questions about how it is being used, contact the Principal Investigator, Dr. Nigel Paneth, at (517)-844-3961.

Who can I contact about my rights/roles within this study?

If you have questions or concerns about your role and rights as a research participant, or would like to obtain information or offer input, or would like to register a complaint about this study, you may contact, anonymously if you wish, the Michigan State University's Human Research Protection Program at (517)-355-2180, Fax (517)-432-4503, or email irb@msu.edu or regular mail at 4000 Collins Rd, Suite 136, Lansing, MI 48910.

Statement of Consent

By signing below, you will indicate your voluntary agreement to participate in this research and to have your child participate in this research. Upon signature you will receive a copy of the consent form.

I voluntarily agree to participate in the study.

_____

(Signature of Participant)

_____

(Printed Name)

Date: Time:

_____

(Signature of Person Obtaining Consent)

Date: Time:

_____

(Printed Name)

Now that you have agreed to participate in the MARCH and ECHO study, I will now ask you a series of questions about your willingness to participate in specific parts of the study that we would like to describe in more detail.

Six drops of blood are collected from a baby's heel shortly after birth to diagnose disorders that need early treatment. After coding to protect your privacy, blood spots left over after newborn screening can be used for research through the Michigan BioTrust for Health program. When your child is born, you will be asked if you will allow your child's leftover spots to be available for research through the BioTrust. This consent is for use of blood spots that are not identified, where the researcher does not know whose blood spot is being used.

We now ask permission to gain access to both your and your child's identified leftover blood spots. We need the spots to be identified so that we can connect information from the spots to other information you may provide us with during M-ARCH. We will use the smallest amount we can from the blood spots, but we may have to use all of your and your child's leftover blood spots that have been reserved for research. We will ***not use*** the one blood spot reserved by MDHHS in case your family needs access to it for personal use.

Blood spots will only be used for research on mother and child health such as we described above consent document. There are many different types of laboratory methods that we might use in the future that can study factors such as genes, environmental chemicals, and more. Once these spots are provided to the CHARM research team, they will be coded with a unique identification number so that researchers doing specific projects will not see you or your child's name. For extra protection, ***each*** blood spot project must be approved by MDHHS to make sure your privacy is

protected, and that the scientific work is appropriate. In order to access these blood spots we will ask you to provide the hospital at which you were born and your mother's name when she gave birth to you.

1. Will you allow us to gain access to your and your child's identified leftover blood spots?

    a. Yes

    b. No

Some scientists, both inside and outside of the ECHO program, might want to study your genes or the genes of your child. We can get this genetic information from the specimens you provide to us. We know genes and DNA can affect health and illness, so the ECHO researchers are very interested in how they might affect mothers and children. In the future other researchers might use this genetic information to study different scientific and medical questions than the ones ECHO is trying to answer. We don't know now what those future questions might be. Genetic studies will need to access not just to genetic information, but also to the other information you give us in MARCH.

2. Do you give us permission to share de-identified genetic and other information about you and your baby with these other scientists? However, any identifiable information such as address and dates of birth will not be a part of that data set.

    a. Yes

    b. No

You are currently enrolled in the MARCH Study. We would like permission to contact you for future possible studies related to this one. Your contact information will be maintained by MARCH staff and stored in a password protected computer database, separately from your collected information. It will only be available to the investigators and research staff of the study. You may choose to withdraw your permission at any time. Agreeing to allow us to contact you for future studies does NOT mean you agree to participate in future studies.

3. Will you allow other MARCH researchers to contact you about future studies related to MARCH?

    a. Yes

    b. No

APPENDIX C: SAMPLE INFORMATION FORM

Study ID _____ Date rcv'd in lab: _____ # bead tubes: _____ # Eppendorf tubes: _____

Infant 3 Month Sample Info Form

1) What is today's date? (Month/Day/Year)

__ __ / __ __ / __ __ __ __

2) Diaper used to collect sample:

☐ Provided
☐ Personal disposable (Brand: _____)
☐ Cloth

3) What does baby weigh *now*?

__ __ lbs. __ __ ozs.

4) How long is baby *now*?

__ __ __ __ __ inches

5) Is baby taking medicine(s) now?

☐ yes ☐ no

a. If yes, what is the name of the medicine(s)?

b. What is the reason for the medicine(s)?

6) If baby is not taking medicine now, has baby taken any medicine(s) since birth?

☐ yes ☐ no

a. If yes, what was the name of the medicine(s)?

b. What was the reason for the medicine(s)?

7) Is baby sick?

☐ yes ☐ no

If yes, describe.

8) Has baby had any antibiotics since birth?

☐ yes ☐ no

Study ID _____

Infant 3 Month Sample Info Form (continued)

9) Are YOU taking medicine(s) now?

☐ yes ☐ no

a. If yes, what is the name of the medicine(s)?

b. What is the reason for the medicine(s)?

10) Did baby have breast milk from the breast in the past day?

☐ yes ☐ no

11) Did baby have breast milk from a bottle in the past day?

☐ yes ☐ no

12) Did baby have infant formula in the past day?

☐ yes ☐ no

a. If the baby had formula, what is the name and brand of the infant formula?

b. What type of water was used to prepare the formula?
- ☐ Tap water
- ☐ Well water
- ☐ Bottled water (Brand: _____)
- ☐ Formula was purchased as a ready-to-use liquid

13) What else did baby eat and drink in the past day (example: water, sugar water, other milks, purees, vitamin D drops, pedialyte, liquid supplements, etc.)?

14) During this past week, my baby ate: (circle one)

a. 100% breast milk

b. 80% breast milk, 20% formula or other foods

c. 50-80% breast milk, the rest formula or other foods
(2-5 of every 10 feedings were not breast milk)

d. 50% breast milk, 50% formula or other foods

e. 20-50% breast milk, the rest formula or other foods
(5-8 of every 10 feedings are not breast milk)

f. 20% breast milk, 80% formula or other foods

g. 100% formula or other foods

Study ID _____

Infant 3 Month Sample Info Form (continued)

Place an "x" in the box next to anything that baby ate in the past 24 hours:	
☐	Cow's milk (not infant formula)
☐	Other milk: soy, rice, goat, etc. (not infant formula)
☐	Other dairy foods: yogurt, cheese, ice cream, pudding, etc.
☐	Other soy foods: tofu, frozen soy desserts, etc.
☐	100% fruit juice
☐	100% vegetable juice
☐	Sweet drinks: soda, sweet tea, Kool-Aid, Gatorade, etc.
☐	Baby Cereal
☐	Other grains: Cheerios, other breakfast cereals, teething biscuits, crackers, breads, pasta, rice, etc.
☐	Carrots, sweet potatoes, mangos, apricots, bell peppers
☐	Spinach, kale, Swiss chard, romaine lettuce
☐	Other fruit or fruit purees
☐	Other vegetables or vegetable purees
☐	French fries
☐	Meat: chicken, beef, ham, combination dinners
☐	Fish or shellfish
☐	Peanut butter, other peanut foods (Bamba), other nuts
☐	Liver or other organ meats
☐	Eggs
☐	Beans, lentils, peas
☐	Sweet foods: candy, chocolate, cookies, cakes, etc.
☐	Vitamin D supplement
☐	Multi-vitamin supplement
☐	Fluoride drops
☐	Fish oil, DHA or EPA supplement
☐	Prebiotic supplement (Gos, Fos, inulin, beta-glucan, etc.)
☐	Probiotic supplement, kefir, kimchi
☐	Any other dietary supplement *Name:*_____

APPENDIX D: QUESTIONS USE FOR COVARIATES

Table 20. Questions use for covariates

Questionnaires	Time	Variable names	Questions	Aim
MARCH Prenatal 1 Survey questionnaire	During pregnancy	Maternal education level	Looking at page 16, what is the highest grade or level of school you have completed or the highest degree you have received?	1,2,3
		Maternal height	Feet & Inches	1,2,3
		Pre-pregnancy weight	Just before you got pregnant with this baby, how much did you weigh?	1,2,3
		Maternal age	_____	1,2,3
Birth certificate information	Infants were born	Infant sex	Sex	1,2,3
		Estimated weeks of gestation	Estimated weeks of gestational age	1,2,3
		Mode of delivery	Final route and method of delivery	1,2,3
MARCH 3-month survey	Infants were 3 months of age	Infant race	Baby race	1,2,3
Sample collection form	Fecal collection at 3 months	Infant feeding method in the past 24 hours	Did baby have breast milk from the breast in the past day?	1,2
			Did baby have breast milk from the bottle in the past day?	
			Did baby have infant formula in the past day?	
			What else did baby eat and drink in the past day?	
		Infant feeding method in the past week	During this past week, my baby ate	1,3
		Infant breastfeeding patterns in the past 24 hours	Did baby have breast milk from the breast in the past day?	3
			Did baby have breast milk from the bottle in the past day?	
			Did baby have infant formula in the past day?	
		Antibiotics intake since birth	Has baby had any antibiotics since birth?	1,2,3

Table 20 (cont'd)

MARCH 9-month survey dictionary	Infants were 9 months of age	Breastfeeding duration	Was this child EVER breastfed or fed breast milk?	2
			If yes, how old was this child when he/she completely stopped breastfeeding or being fed breast milk?	
		Any breastfeeding duration	How old was this child when he/she was first **fed formula?**	
			How old was this child when he or she was FIRST fed anything other than breast milk or formula? Include juice, cow's milk, sugar water, baby food, or anything else that your child might have been given, even water.	
		ASQ	———	1,2

APPENDIX E: ORIGINAL R CODES

Chapter 2

Data preparation

```
library(vegan)
library(lubridate)
library(tidyr)
library(MASS)
library(car)
library(dunn.test)
library(ggplot2)
library(openxlsx)
library(Hmisc)
library(DirichletMultinomial)
library(microbiome)
library(reshape2)
library(magrittr)
library(dplyr)
library(Maaslin2)
library(ggpubr)
library(funrar)
require(fifer)
library(clusterSim)
library(forcats)

setwd("/Users/busihan/Desktop/2023Mar20_Aim1_double_check")
Data.Subsample.genus_37wks<-read.csv("Data.Subsample.genus_37wks.csv",
header = T,stringsAsFactors = T,row.names = 1)
metadata<-read.csv("metadata_updated_Jan.csv",na="",header = T)
cols<-c("antibiotics_since_birth","FED_PRAC_NEW","SEX","FED_PRAC_LIGHT
_NEW","MD_FINAL_ROUTE","Race_new","EDU_LVL")
summary(metadata)
metadata[cols]<-lapply(metadata[cols], factor)
sapply(metadata,class)

Data.Subsample.genus_37wks$Group
metadata$Group
temp<-merge(Data.Subsample.genus_37wks, metadata,by.x="Group")

Data.Subsample.genus_37wks<-temp[,c(1:(ncol(Data.Subsample.genus_37wks
)))]
metadata<-temp[,c(1,254:294)]

Data.Subsample.genus_37wks$Group
metadata$Group
```

```
Alpha<-function(OTU,Names="Sample",Groups="Sample"){
  Chao<-t(estimateR(OTU))
  Chao<-Chao[,2]
  Shannon<-diversity(OTU,index="shannon")
  Invsimpson<-diversity(OTU,index="invsimpson")
  OTU.Subsample.Alpha<-data.frame(Names,Groups,Chao,Shannon,Invsimpson
)
  return(OTU.Subsample.Alpha)
}

Sor.bray.pcoa<-function(OTUS,Dim=2,Color=1,binary,pch=16,Title="PCoA")
{
  Data.df<-vegdist(OTUS,method="bray", binary)
  Data.df.PCoA<-cmdscale(Data.df, k = Dim, eig = FALSE)
  Data.df.PCoA.eig<-cmdscale(Data.df, k = Dim, eig = TRUE)
  eig.Data.df.PCoA<-Data.df.PCoA.eig$eig
  eig.Data.df.PCoA.sum<-sum(eig.Data.df.PCoA)
  a<-(eig.Data.df.PCoA/eig.Data.df.PCoA.sum)*100
  xlab<-paste("PC1","(",round(a[1],1),"%",")",sep="")
  ylab<-paste("PC2","(",round(a[2],1),"%",")",sep="")
  if(binary==TRUE){
    main<-"Sorensen PCoA"
  }else(main<-"Bray-Curtis PCoA")
  plot(Data.df.PCoA, col=Color,
       main=Title,xlab=xlab,ylab=ylab,pch=c(pch))
  return(Data.df.PCoA)
}


PERMANOVA<-function(OTUS,Group,binary,iters=9999){
  Data.Dist<-vegdist(OTUS,method="bray", binary=binary)
  adonis2(Data.Dist~Group,permutations=iters)
}

PERMDISP<-function(OTUS,Group,binary,iters=9999){
  Data.Dist<-vegdist(OTUS,method="bray", binary=binary)
  Data.betadisper<-betadisper(Data.Dist, group=Group)
  permutest(Data.betadisper, group=Group, permutations=iters)
}
```

Table 1. Population characteristics and scores on the five ASQ scales

```
shapiro.test(metadata$asq_9_total_grossmotor)  #p-value = 8.718e-05
shapiro.test(metadata$asq_9_total_finemotor) #p-value = 3.08e-08
shapiro.test(metadata$asq_9_total_communication.total.) #p-value = 0.0
02063
```

```
shapiro.test(metadata$asq_9_total_personal_social) #p-value = 0.006921
shapiro.test(metadata$asq_9_total_problemsolving) #p-value = 4.254e-07
summary(metadata)
###Categorical variable###
###Baby Sex###
31/64*100
33/64*100
#gross motor
sex<-aggregate(metadata$asq_9_total_grossmotor~metadata$SEX,FUN = leng
th)
sex_median<-aggregate(metadata$asq_9_total_grossmotor~metadata$SEX,FUN
 = median)
sex_median
sex_min<-aggregate(metadata$asq_9_total_grossmotor~metadata$SEX,FUN =
min)
sex_min
sex_max<-aggregate(metadata$asq_9_total_grossmotor~metadata$SEX,FUN =
max)
sex_max
wilcox.test(asq_9_total_grossmotor~SEX, data = metadata,exact = FALSE)


#fine motor
sex_median<-aggregate(metadata$asq_9_total_finemotor~metadata$SEX,FUN
= median)
sex_median
sex_min<-aggregate(metadata$asq_9_total_finemotor~metadata$SEX,FUN = m
in)
sex_min
sex_max<-aggregate(metadata$asq_9_total_finemotor~metadata$SEX,FUN = m
ax)
sex_max
wilcox.test(asq_9_total_finemotor~SEX, data = metadata,exact = FALSE)

#communication
sex_median<-aggregate(metadata$asq_9_total_communication.total.~metada
ta$SEX,FUN = median)
sex_median
sex_min<-aggregate(metadata$asq_9_total_communication.total.~metadata$
SEX,FUN = min)
sex_min
sex_max<-aggregate(metadata$asq_9_total_communication.total.~metadata$
SEX,FUN = max)
sex_max
wilcox.test(asq_9_total_communication.total.~SEX, data = metadata,exac
t = FALSE)
```

```
#personal social
sex_median<-aggregate(metadata$asq_9_total_personal_social~metadata$SE
X,FUN = median)
sex_median
sex_min<-aggregate(metadata$asq_9_total_personal_social~metadata$SEX,F
UN = min)
sex_min
sex_max<-aggregate(metadata$asq_9_total_personal_social~metadata$SEX,F
UN = max)
sex_max
wilcox.test(asq_9_total_personal_social~SEX, data = metadata,exact = F
ALSE)

#problem solving
sex_median<-aggregate(metadata$asq_9_total_problemsolving~metadata$SEX
,FUN = median)
sex_median
sex_min<-aggregate(metadata$asq_9_total_problemsolving~metadata$SEX,FU
N = min)
sex_min
sex_max<-aggregate(metadata$asq_9_total_problemsolving~metadata$SEX,FU
N = max)
sex_max
wilcox.test(asq_9_total_problemsolving~SEX, data = metadata,exact = FA
LSE)

###Baby Race###
44/64*100
20/64*100
#gross motor
Race_new<-aggregate(metadata$asq_9_total_grossmotor~metadata$Race_new,
FUN = length)
Race_new_median<-aggregate(metadata$asq_9_total_grossmotor~metadata$Ra
ce_new,FUN = median)
Race_new_median
Race_new_min<-aggregate(metadata$asq_9_total_grossmotor~metadata$Race_
new,FUN = min)
Race_new_min
Race_new_max<-aggregate(metadata$asq_9_total_grossmotor~metadata$Race_
new,FUN = max)
Race_new_max
wilcox.test(asq_9_total_grossmotor~Race_new, data = metadata,exact = F
ALSE)

#fine motor
```

```
Race_new_median<-aggregate(metadata$asq_9_total_finemotor~metadata$Rac
e_new,FUN = median)
Race_new_median
Race_new_min<-aggregate(metadata$asq_9_total_finemotor~metadata$Race_n
ew,FUN = min)
Race_new_min
Race_new_max<-aggregate(metadata$asq_9_total_finemotor~metadata$Race_n
ew,FUN = max)
Race_new_max
wilcox.test(asq_9_total_finemotor~Race_new, data = metadata,exact = FA
LSE)

#communication
Race_new_median<-aggregate(metadata$asq_9_total_communication.total.~m
etadata$Race_new,FUN = median)
Race_new_median
Race_new_min<-aggregate(metadata$asq_9_total_communication.total.~meta
data$Race_new,FUN = min)
Race_new_min
Race_new_max<-aggregate(metadata$asq_9_total_communication.total.~meta
data$Race_new,FUN = max)
Race_new_max
wilcox.test(asq_9_total_communication.total.~Race_new, data = metadata
,exact = FALSE)

#personal social
Race_new_median<-aggregate(metadata$asq_9_total_personal_social~metada
ta$Race_new,FUN = median)
Race_new_median
Race_new_min<-aggregate(metadata$asq_9_total_personal_social~metadata$
Race_new,FUN = min)
Race_new_min
Race_new_max<-aggregate(metadata$asq_9_total_personal_social~metadata$
Race_new,FUN = max)
Race_new_max
wilcox.test(asq_9_total_personal_social~Race_new, data = metadata,exac
t = FALSE)

#problem solving
Race_new_median<-aggregate(metadata$asq_9_total_problemsolving~metadat
a$Race_new,FUN = median)
Race_new_median
Race_new_min<-aggregate(metadata$asq_9_total_problemsolving~metadata$R
ace_new,FUN = min)
Race_new_min
Race_new_max<-aggregate(metadata$asq_9_total_problemsolving~metadata$R
```

```
ace_new,FUN = max)
Race_new_max
wilcox.test(asq_9_total_problemsolving~Race_new, data = metadata,exact
 = FALSE)

###Maternal education level###
3/64*100
11/64*100
13/64*100
37/64*100
#gross motor
EDU_LVL_median<-aggregate(metadata$asq_9_total_grossmotor~metadata$EDU
_LVL,FUN = median)
EDU_LVL_median
EDU_LVL_min<-aggregate(metadata$asq_9_total_grossmotor~metadata$EDU_LV
L,FUN = min)
EDU_LVL_min
EDU_LVL_max<-aggregate(metadata$asq_9_total_grossmotor~metadata$EDU_LV
L,FUN = max)
EDU_LVL_max
kruskal.test(asq_9_total_grossmotor~EDU_LVL, data = metadata)

#fine motor
EDU_LVL_median<-aggregate(metadata$asq_9_total_finemotor~metadata$EDU_
LVL,FUN = median)
EDU_LVL_median
EDU_LVL_min<-aggregate(metadata$asq_9_total_finemotor~metadata$EDU_LVL
,FUN = min)
EDU_LVL_min
EDU_LVL_max<-aggregate(metadata$asq_9_total_finemotor~metadata$EDU_LVL
,FUN = max)
EDU_LVL_max
kruskal.test(asq_9_total_finemotor~EDU_LVL, data = metadata)
dunn.test(metadata$asq_9_total_finemotor,metadata$EDU_LVL,altp = TRUE,
 method="bh")

#communication
EDU_LVL_median<-aggregate(metadata$asq_9_total_communication.total.~me
tadata$EDU_LVL,FUN = median)
EDU_LVL_median
EDU_LVL_min<-aggregate(metadata$asq_9_total_communication.total.~metad
ata$EDU_LVL,FUN = min)
EDU_LVL_min
EDU_LVL_max<-aggregate(metadata$asq_9_total_communication.total.~metad
ata$EDU_LVL,FUN = max)
EDU_LVL_max
```

```
kruskal.test(asq_9_total_communication.total.~EDU_LVL, data = metadata
)

#personal social
EDU_LVL_median<-aggregate(metadata$asq_9_total_personal_social~metadat
a$EDU_LVL,FUN = median)
EDU_LVL_median
EDU_LVL_min<-aggregate(metadata$asq_9_total_personal_social~metadata$E
DU_LVL,FUN = min)
EDU_LVL_min
EDU_LVL_max<-aggregate(metadata$asq_9_total_personal_social~metadata$E
DU_LVL,FUN = max)
EDU_LVL_max
kruskal.test(asq_9_total_personal_social~EDU_LVL, data = metadata)

#problem solving
EDU_LVL_median<-aggregate(metadata$asq_9_total_problemsolving~metadata
$EDU_LVL,FUN = median)
EDU_LVL_median
EDU_LVL_min<-aggregate(metadata$asq_9_total_problemsolving~metadata$ED
U_LVL,FUN = min)
EDU_LVL_min
EDU_LVL_max<-aggregate(metadata$asq_9_total_problemsolving~metadata$ED
U_LVL,FUN = max)
EDU_LVL_max
kruskal.test(asq_9_total_problemsolving~EDU_LVL, data = metadata)
dunn.test(metadata$asq_9_total_problemsolving,metadata$EDU_LVL,altp =
TRUE, method="bh")

###MD_FINAL_ROUTE###
39/64*100
25/64*100
#gross motor
MD_FINAL_ROUTE<-aggregate(metadata$asq_9_total_grossmotor~metadata$MD_
FINAL_ROUTE,FUN = length)
MD_FINAL_ROUTE_median<-aggregate(metadata$asq_9_total_grossmotor~metad
ata$MD_FINAL_ROUTE,FUN = median)
MD_FINAL_ROUTE_median
MD_FINAL_ROUTE_min<-aggregate(metadata$asq_9_total_grossmotor~metadata
$MD_FINAL_ROUTE,FUN = min)
MD_FINAL_ROUTE_min
MD_FINAL_ROUTE_max<-aggregate(metadata$asq_9_total_grossmotor~metadata
$MD_FINAL_ROUTE,FUN = max)
MD_FINAL_ROUTE_max
wilcox.test(asq_9_total_grossmotor~MD_FINAL_ROUTE,data=metadata,exact
= FALSE)
```

```
#fine motor
MD_FINAL_ROUTE_median<-aggregate(metadata$asq_9_total_finemotor~metada
ta$MD_FINAL_ROUTE,FUN = median)
MD_FINAL_ROUTE_median
MD_FINAL_ROUTE_min<-aggregate(metadata$asq_9_total_finemotor~metadata$
MD_FINAL_ROUTE,FUN = min)
MD_FINAL_ROUTE_min
MD_FINAL_ROUTE_max<-aggregate(metadata$asq_9_total_finemotor~metadata$
MD_FINAL_ROUTE,FUN = max)
MD_FINAL_ROUTE_max
wilcox.test(asq_9_total_finemotor~MD_FINAL_ROUTE,data=metadata,exact =
 FALSE)

#communication
MD_FINAL_ROUTE_median<-aggregate(metadata$asq_9_total_communication.to
tal.~metadata$MD_FINAL_ROUTE,FUN = median)
MD_FINAL_ROUTE_median
MD_FINAL_ROUTE_min<-aggregate(metadata$asq_9_total_communication.total
.~metadata$MD_FINAL_ROUTE,FUN = min)
MD_FINAL_ROUTE_min
MD_FINAL_ROUTE_max<-aggregate(metadata$asq_9_total_communication.total
.~metadata$MD_FINAL_ROUTE,FUN = max)
MD_FINAL_ROUTE_max
wilcox.test(asq_9_total_communication.total.~MD_FINAL_ROUTE,data=metad
ata,exact = FALSE)

#personal social
MD_FINAL_ROUTE_median<-aggregate(metadata$asq_9_total_personal_social~
metadata$MD_FINAL_ROUTE,FUN = median)
MD_FINAL_ROUTE_median
MD_FINAL_ROUTE_min<-aggregate(metadata$asq_9_total_personal_social~met
adata$MD_FINAL_ROUTE,FUN = min)
MD_FINAL_ROUTE_min
MD_FINAL_ROUTE_max<-aggregate(metadata$asq_9_total_personal_social~met
adata$MD_FINAL_ROUTE,FUN = max)
MD_FINAL_ROUTE_max
wilcox.test(asq_9_total_personal_social~MD_FINAL_ROUTE,data=metadata,e
xact = FALSE)

#problem solving
MD_FINAL_ROUTE_median<-aggregate(metadata$asq_9_total_problemsolving~m
etadata$MD_FINAL_ROUTE,FUN = median)
MD_FINAL_ROUTE_median
MD_FINAL_ROUTE_min<-aggregate(metadata$asq_9_total_problemsolving~meta
data$MD_FINAL_ROUTE,FUN = min)
```

```
MD_FINAL_ROUTE_min
MD_FINAL_ROUTE_max<-aggregate(metadata$asq_9_total_problemsolving~meta
data$MD_FINAL_ROUTE,FUN = max)
MD_FINAL_ROUTE_max
wilcox.test(asq_9_total_problemsolving~MD_FINAL_ROUTE,data=metadata,ex
act = FALSE)

###FED_PRAC_LIGHT###
9/64*100
17/64*100
16/64*100
22/64*100
#Gross motor
summary(metadata)
FED_PRAC_LIGHT_median<-aggregate(metadata$asq_9_total_grossmotor~metad
ata$FED_PRAC_LIGHT_NEW,FUN = median)
FED_PRAC_LIGHT_median
FED_PRAC_LIGHT_min<-aggregate(metadata$asq_9_total_grossmotor~metadata
$FED_PRAC_LIGHT_NEW,FUN = min)
FED_PRAC_LIGHT_min
FED_PRAC_LIGHT_max<-aggregate(metadata$asq_9_total_grossmotor~metadata
$FED_PRAC_LIGHT_NEW,FUN = max)
FED_PRAC_LIGHT_max
kruskal.test(asq_9_total_grossmotor~FED_PRAC_LIGHT_NEW, data =metadata
)

#Fine motor
FED_PRAC_LIGHT_median<-aggregate(metadata$asq_9_total_finemotor~metada
ta$FED_PRAC_LIGHT_NEW,FUN = median)
FED_PRAC_LIGHT_median
FED_PRAC_LIGHT_min<-aggregate(metadata$asq_9_total_finemotor~metadata$
FED_PRAC_LIGHT_NEW,FUN = min)
FED_PRAC_LIGHT_min
FED_PRAC_LIGHT_max<-aggregate(metadata$asq_9_total_finemotor~metadata$
FED_PRAC_LIGHT_NEW,FUN = max)
FED_PRAC_LIGHT_max
kruskal.test(asq_9_total_finemotor~FED_PRAC_LIGHT_NEW, data =metadata)

dunn.test(metadata$asq_9_total_finemotor,metadata$FED_PRAC_LIGHT_NEW,a
ltp = TRUE, method="bh")

#Communication
FED_PRAC_LIGHT_median<-aggregate(metadata$asq_9_total_communication.to
tal.~metadata$FED_PRAC_LIGHT_NEW,FUN = median)
FED_PRAC_LIGHT_median
FED_PRAC_LIGHT_min<-aggregate(metadata$asq_9_total_communication.total
```

```
.~metadata$FED_PRAC_LIGHT_NEW,FUN = min)
FED_PRAC_LIGHT_min
FED_PRAC_LIGHT_max<-aggregate(metadata$asq_9_total_communication.total
.~metadata$FED_PRAC_LIGHT_NEW,FUN = max)
FED_PRAC_LIGHT_max
kruskal.test(asq_9_total_communication.total.~FED_PRAC_LIGHT_NEW, data
 =metadata)

#Personal and social
FED_PRAC_LIGHT_median<-aggregate(metadata$asq_9_total_personal_social~
metadata$FED_PRAC_LIGHT_NEW,FUN = median)
FED_PRAC_LIGHT_median
FED_PRAC_LIGHT_min<-aggregate(metadata$asq_9_total_personal_social~met
adata$FED_PRAC_LIGHT_NEW,FUN = min)
FED_PRAC_LIGHT_min
FED_PRAC_LIGHT_max<-aggregate(metadata$asq_9_total_personal_social~met
adata$FED_PRAC_LIGHT_NEW,FUN = max)
FED_PRAC_LIGHT_max
kruskal.test(asq_9_total_personal_social~FED_PRAC_LIGHT_NEW, data =met
adata)

#Problem solving
FED_PRAC_LIGHT_median<-aggregate(metadata$asq_9_total_problemsolving~m
etadata$FED_PRAC_LIGHT_NEW,FUN = median)
FED_PRAC_LIGHT_median
FED_PRAC_LIGHT_min<-aggregate(metadata$asq_9_total_problemsolving~meta
data$FED_PRAC_LIGHT_NEW,FUN = min)
FED_PRAC_LIGHT_min
FED_PRAC_LIGHT_max<-aggregate(metadata$asq_9_total_problemsolving~meta
data$FED_PRAC_LIGHT_NEW,FUN = max)
FED_PRAC_LIGHT_max
kruskal.test(asq_9_total_problemsolving~FED_PRAC_LIGHT_NEW, data =meta
data)

###Continuous variable###
### pre BMI###
mean(metadata$PRE_BMI)
sd(metadata$PRE_BMI)
#Gross motor
preBMI<-lm(asq_9_total_grossmotor~PRE_BMI,data=metadata)
summary(preBMI)
confint(preBMI,'PRE_BMI',level=0.95)

#Fine motor
preBMI<-lm(asq_9_total_finemotor~PRE_BMI,data=metadata)
summary(preBMI)
```

```
confint(preBMI,'PRE_BMI',level=0.95)

#Communication
preBMI<-lm(asq_9_total_communication.total.~PRE_BMI ,data=metadata)
summary(preBMI)
confint(preBMI,'PRE_BMI',level=0.95)

#Personal and social
preBMI<-lm(asq_9_total_personal_social~PRE_BMI,data=metadata)
summary(preBMI)
confint(preBMI,'PRE_BMI',level=0.95)

#Problem solving
preBMI<-lm(asq_9_total_problemsolving~PRE_BMI,data=metadata)
summary(preBMI)
confint(preBMI,'PRE_BMI',level=0.95)

###maternal age###
mean(metadata$maternal_age)
sd(metadata$maternal_age)
#Gross motor
MATERALAGE<-lm(asq_9_total_grossmotor~maternal_age,data=metadata)
summary(MATERALAGE)
confint(MATERALAGE,'maternal_age', level=0.95)

#Fine motor
MATERALAGE <-lm(asq_9_total_finemotor~maternal_age ,data=metadata)
summary(MATERALAGE)
confint(MATERALAGE,'maternal_age',level=0.95)

#Communication
MATERALAGE<-lm(asq_9_total_communication.total.~maternal_age ,data=met
adata)
summary(MATERALAGE)
confint(MATERALAGE,'maternal_age',level=0.95)

#Personal and social
MATERALAGE<-lm(asq_9_total_personal_social~ maternal_age,data=metadata
)
summary(MATERALAGE)
confint(MATERALAGE,'maternal_age',level=0.95)

#Problem solving
MATERALAGE<-lm(asq_9_total_problemsolving~ maternal_age,data=metadata)
summary(MATERALAGE)
confint(MATERALAGE,'maternal_age',level=0.95)
```

```
###gestational age at birth###
mean(metadata$ESTWKSGEST)
sd(metadata$ESTWKSGEST)
#Gross motor
ESTWKSGEST<-lm(asq_9_total_grossmotor~ESTWKSGEST,data=metadata)
summary(ESTWKSGEST)
confint(ESTWKSGEST,'ESTWKSGEST',level=0.95)

#Fine motor
ESTWKSGEST<-lm(asq_9_total_finemotor~ESTWKSGEST,data=metadata)
summary(ESTWKSGEST)
confint(ESTWKSGEST,'ESTWKSGEST',level=0.95)

#Communication
ESTWKSGEST<-lm(asq_9_total_communication.total.~ESTWKSGEST,data=metada
ta)
summary(ESTWKSGEST)
confint(ESTWKSGEST,'ESTWKSGEST',level=0.95)

#Personal and social
ESTWKSGEST<-lm(asq_9_total_personal_social~ESTWKSGEST,data=metadata)
summary(ESTWKSGEST)
confint(ESTWKSGEST,'ESTWKSGEST',level=0.95)

#Problem solving
ESTWKSGEST<-lm(asq_9_total_problemsolving~ESTWKSGEST,data=metadata)
summary(ESTWKSGEST)
confint(ESTWKSGEST,'ESTWKSGEST',level=0.95)
```

Table 2. The associations between alpha diversity of gut microbiota at 3 months and each of the five ASQ scale measurements at 9 months

```
options(scipen = 999)
Data.Subsample.final.Alpha<-read.csv("/Users/busihan/Desktop/MARCH\ B3
m_ASQ_updated/Data.Subsample.final.Alpha_Final.csv", header = T)
chao<-Data.Subsample.final.Alpha$Chao
shan<-Data.Subsample.final.Alpha$Shannon
invismp<-Data.Subsample.final.Alpha$Invsimpson

###gross motor###
grossmotor_chao<-lm(asq_9_total_grossmotor~chao+FED_PRAC_LIGHT_NEW+ant
ibiotics_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+ESTWKSGEST+PR
E_BMI+maternal_age,data=metadata)
summary(grossmotor_chao)
confint(grossmotor_chao,"chao")
```

```
grossmotor_shan<-lm(asq_9_total_grossmotor~shan+FED_PRAC_LIGHT_NEW+ant
ibiotics_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+ESTWKSGEST+PR
E_BMI+maternal_age,data=metadata)
summary(grossmotor_shan)
confint(grossmotor_shan,"shan")

grossmotor_invismp<-lm(asq_9_total_grossmotor~invismp+FED_PRAC_LIGHT_N
EW+antibiotics_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+ESTWKSG
EST+PRE_BMI+maternal_age,data=metadata)
summary(grossmotor_invismp)
confint(grossmotor_invismp,"invismp")

#fine motor
finemotor_chao<-lm(asq_9_total_finemotor~chao+FED_PRAC_LIGHT_NEW+antib
iotics_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+ESTWKSGEST+PRE_
BMI+maternal_age,data=metadata)
summary(finemotor_chao)
confint(finemotor_chao,"chao")

finemotor_shan<-lm(asq_9_total_finemotor~shan+FED_PRAC_LIGHT_NEW+antib
iotics_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+ESTWKSGEST+PRE_
BMI+maternal_age,data=metadata)
summary(finemotor_shan)
confint(finemotor_shan,"shan")

finemotor_invismp<-lm(asq_9_total_finemotor~invismp+FED_PRAC_LIGHT_NEW
+antibiotics_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+ESTWKSGES
T+PRE_BMI+maternal_age,data=metadata)
summary(finemotor_invismp)
confint(grossmotor_invismp,"invismp")

##communication###
communication_chao<-lm(asq_9_total_communication.total.~chao+FED_PRAC_
LIGHT_NEW+antibiotics_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+
ESTWKSGEST+PRE_BMI+maternal_age,data=metadata)
summary(communication_chao)
confint(communication_chao,"chao")

communication_shan<-lm(asq_9_total_communication.total.~shan+FED_PRAC_
LIGHT_NEW+antibiotics_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+
ESTWKSGEST+PRE_BMI+maternal_age,data=metadata)
summary(communication_shan)
confint(communication_shan,"shan")

communication_invismp<-lm(asq_9_total_communication.total.~invismp+FED
_PRAC_LIGHT_NEW+antibiotics_since_birth+SEX+MD_FINAL_ROUTE+Race_new+ED
```

```
U_LVL+ESTWKSGEST+PRE_BMI+maternal_age,data=metadata)
summary(communication_invismp)
confint(communication_invismp,"invismp")

###personal and social###
social_chao<-lm(asq_9_total_personal_social~chao+FED_PRAC_LIGHT_NEW+an
tibiotics_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+ESTWKSGEST+P
RE_BMI+maternal_age,data=metadata)
summary(social_chao)
confint(social_chao,"chao")

social_shan<-lm(asq_9_total_personal_social~shan+FED_PRAC_LIGHT_NEW+an
tibiotics_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+ESTWKSGEST+P
RE_BMI+maternal_age,data=metadata)
summary(social_shan)
confint(social_shan,"shan")

social_invismp<-lm(asq_9_total_personal_social~invismp+FED_PRAC_LIGHT_
NEW+antibiotics_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+ESTWKS
GEST+PRE_BMI+maternal_age,data=metadata)
summary(social_invismp)
confint(social_invismp,"invismp")

###problem solving###
problem_chao<-lm(asq_9_total_problemsolving~chao+FED_PRAC_LIGHT_NEW+an
tibiotics_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+ESTWKSGEST+P
RE_BMI+maternal_age,data=metadata)
summary(problem_chao)
confint(problem_chao,"chao")

problem_shan<-lm(asq_9_total_problemsolving~shan+FED_PRAC_LIGHT_NEW+an
tibiotics_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+ESTWKSGEST+P
RE_BMI+maternal_age,data=metadata)
summary(problem_shan)
confint(problem_shan,"shan")

problem_invismp<-lm(asq_9_total_problemsolving~invismp+FED_PRAC_LIGHT_
NEW+antibiotics_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+ESTWKS
GEST+PRE_BMI+maternal_age,data=metadata)
summary(problem_invismp)
confint(problem_invismp,"invismp")
```

Figure 1. The associations between Chao 1 index and ASQ by different feeding methods at 3months

```
#gross motor
cor.test(Data.Subsample.final.Alpha$Chao,metadata$asq_9_total_grossmot
```

```
or,method="spearman",exact=F)
p1<-ggplot(metadata, aes(x = asq_9_total_grossmotor, y = chao))+
  geom_point(aes(color = FED_PRAC_LIGHT_NEW, shape = FED_PRAC_LIGHT_NE
W))+
  geom_smooth(aes(color = FED_PRAC_LIGHT_NEW), method ="lm",
              se = FALSE, fullrange = TRUE)+
  scale_color_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),
                     values = c("#65A8D3","#67C5AB","#E69D67","#847AB7
"))+
  scale_fill_manual(name="Feeding method",labels=c("Breastmilk","Breas
tmilk_vitaminD","Partial breastmilk","Formula"),values = c("#65A8D3","
#67C5AB","#E69D67","#847AB7"))+
  scale_shape_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),values=c(15, 16, 17,
18))+
  stat_cor(method="spearman",aes(color = FED_PRAC_LIGHT_NEW),show.lege
nd = FALSE,r.accuracy=0.01,p.accuracy=0.01)+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, size=1)
)+
  labs(x="Gross motor",y="Chao1 Index")+
  ggtitle("Gross motor_Chao1")+
  theme(plot.title=element_text(hjust=0.5))+
  geom_text(x=50, y=77, label="R = -0.14, p = 0.27")

p1

## `geom_smooth()` using formula = 'y ~ x'

#fine motor
cor.test(Data.Subsample.final.Alpha$Chao,metadata$asq_9_total_finemoto
r,method="spearman", exact=F)
p2<-ggplot(metadata, aes(x = asq_9_total_finemotor, y = chao))+
  geom_point(aes(color = FED_PRAC_LIGHT_NEW, shape = FED_PRAC_LIGHT_NE
W))+
  geom_smooth(aes(color = FED_PRAC_LIGHT_NEW), method ="lm", se = FALS
E, fullrange = TRUE)+
  scale_color_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),
                     values = c("#65A8D3","#67C5AB","#E69D67","#847AB7
"))+
  scale_fill_manual(name="Feeding method",labels=c("Breastmilk","Breas
tmilk_vitaminD","Partial breastmilk","Formula"),values = c("#65A8D3","
```

```
#67C5AB","#E69D67","#847AB7"))+
  scale_shape_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),values=c(15, 16, 17,
18))+
  stat_cor(method="spearman",aes(color = FED_PRAC_LIGHT_NEW),show.lege
nd = FALSE,r.accuracy=0.01,p.accuracy=0.01)+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, size=1)
)+
  labs(x="Fine motor",y="Chao 1 Index")+
  ggtitle("Fine motor_Chao1")+
  theme(plot.title=element_text(hjust=0.5))+
  geom_text(x=55, y=76, label="R = 0.03, p = 0.81")
p2

## `geom_smooth()` using formula = 'y ~ x'

#Communication
cor.test(Data.Subsample.final.Alpha$Chao,metadata$asq_9_total_communic
ation.total.,method="spearman", exact=F)
p3<-ggplot(metadata, aes(x = asq_9_total_communication.total., y = cha
o))+
  geom_point(aes(color = FED_PRAC_LIGHT_NEW, shape = FED_PRAC_LIGHT_NE
W))+
  geom_smooth(aes(color = FED_PRAC_LIGHT_NEW), method ="lm",
              se = FALSE, fullrange = TRUE)+
  scale_color_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),
                     values = c("#65A8D3","#67C5AB","#E69D67","#847AB7
"))+
  scale_fill_manual(name="Feeding method",labels=c("Breastmilk","Breas
tmilk_vitaminD","Partial breastmilk","Formula"),values = c("#65A8D3","
#67C5AB","#E69D67","#847AB7"))+
  scale_shape_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),values=c(15, 16, 17,
18))+
  stat_cor(method="spearman",aes(color = FED_PRAC_LIGHT_NEW),show.lege
nd = FALSE,r.accuracy=0.01,p.accuracy=0.01)+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, size=1)
)+
  labs(x="Communication",y="Chao 1 Index")+
```

```
  ggtitle("Communication_Chao1")+
  theme(plot.title=element_text(hjust=0.5))+
  geom_text( x=50, y=76, label="R = 0.15, p = 0.25")
p3

## `geom_smooth()` using formula = 'y ~ x'

#personal social
cor.test(Data.Subsample.final.Alpha$Chao,metadata$asq_9_total_personal
_social,method="spearman", exact=F)
p4<-ggplot(metadata, aes(x = asq_9_total_personal_social, y = chao))+
  geom_point(aes(color = FED_PRAC_LIGHT_NEW, shape = FED_PRAC_LIGHT_NE
W))+
  geom_smooth(aes(color = FED_PRAC_LIGHT_NEW), method ="lm",
              se = FALSE, fullrange = TRUE)+
  scale_color_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),
                     values = c("#65A8D3","#67C5AB","#E69D67","#847AB7
"))+
  scale_fill_manual(name="Feeding method",labels=c("Breastmilk","Breas
tmilk_vitaminD","Partial breastmilk","Formula"),values = c("#65A8D3","
#67C5AB","#E69D67","#847AB7"))+
  scale_shape_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),values=c(15, 16, 17,
18))+
  stat_cor(method="spearman",aes(color = FED_PRAC_LIGHT_NEW),show.lege
nd = FALSE,r.accuracy=0.01,p.accuracy=0.01)+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, size=1)
)+
  labs(x="Personal social",y="Chao1 Index")+
  ggtitle("Personal social_Chao1")+
  theme(plot.title=element_text(hjust=0.5))+
  geom_text(x=50, y=76, label="R = 0.29, p = 0.02")
p4

## `geom_smooth()` using formula = 'y ~ x'

#problem solving
cor.test(Data.Subsample.final.Alpha$Chao,metadata$asq_9_total_problems
olving,method="spearman", exact=F)
p5<-ggplot(metadata, aes(x = asq_9_total_problemsolving, y = chao))+
  geom_point(aes(color = FED_PRAC_LIGHT_NEW, shape = FED_PRAC_LIGHT_NE
W))+
  geom_smooth(aes(color = FED_PRAC_LIGHT_NEW), method ="lm",
```

```
              se = FALSE, fullrange = TRUE)+
  scale_color_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),
                     values = c("#65A8D3","#67C5AB","#E69D67","#847AB7
"))+
  scale_fill_manual(name="Feeding method",labels=c("Breastmilk","Breas
tmilk_vitaminD","Partial breastmilk","Formula"),values = c("#65A8D3","
#67C5AB","#E69D67","#847AB7"))+
  scale_shape_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),values=c(15, 16, 17,
18))+
  stat_cor(method="spearman",aes(color = FED_PRAC_LIGHT_NEW),show.lege
nd = FALSE,r.accuracy=0.01,p.accuracy=0.01)+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, size=1)
)+
  labs(x="Problem solving",y="Chao1 Index")+
  ggtitle("Problem solving_Chao1")+
  theme(plot.title=element_text(hjust=0.5))+
  geom_text(x=50, y=75, label="R = 0.13, p = 0.31")
p5

## `geom_smooth()` using formula = 'y ~ x'

png("Chao1_ASQ_5panels_correct_spearman_final_spearman_overall.png", r
es=300, height=9, width=13,units="in")
ggarrange(
  p1, p2,p3,p4,p5, labels = c("A", "B","C","D","E"),
  common.legend = TRUE, legend = "bottom"
)

## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'

while (!is.null(dev.list()))  dev.off()
```

Figure 2. The associations between Shannon index and ASQ by different feeding methods at 3 months

```
#gross motor
cor.test(Data.Subsample.final.Alpha$Shannon,metadata$asq_9_total_gross
motor,method="spearman",exact=F)
```

```
p6<-ggplot(metadata, aes(x = asq_9_total_grossmotor, y = shan))+
  geom_point(aes(color = FED_PRAC_LIGHT_NEW, shape = FED_PRAC_LIGHT_NE
W))+
  geom_smooth(aes(color = FED_PRAC_LIGHT_NEW), method ="lm",
              se = FALSE, fullrange = TRUE)+
  scale_color_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),
                     values = c("#65A8D3","#67C5AB","#E69D67","#847AB7
"))+
  scale_fill_manual(name="Feeding method",labels=c("Breastmilk","Breas
tmilk_vitaminD","Partial breastmilk","Formula"),values = c("#65A8D3","
#67C5AB","#E69D67","#847AB7"))+
  scale_shape_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),values=c(15, 16, 17,
18))+
  stat_cor(method="spearman",aes(color = FED_PRAC_LIGHT_NEW),show.lege
nd = FALSE,r.accuracy=0.01,p.accuracy=0.01)+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, size=1)
)+
  labs(x="Gross motor",y="Shannon Index")+
  ggtitle("Gross motor_Shannon")+
  theme(plot.title=element_text(hjust=0.5))+
  geom_text(x=50, y=2.65, label="R = 0.08, p = 0.54")
p6

## `geom_smooth()` using formula = 'y ~ x'

#fine motor
cor.test(Data.Subsample.final.Alpha$Shannon,metadata$asq_9_total_finem
otor,method="spearman", exact=F)
p7<-ggplot(metadata, aes(x = asq_9_total_finemotor, y = shan))+
  geom_point(aes(color = FED_PRAC_LIGHT_NEW, shape = FED_PRAC_LIGHT_NE
W))+
  geom_smooth(aes(color = FED_PRAC_LIGHT_NEW), method ="lm",
              se = FALSE, fullrange = TRUE)+
  scale_color_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),
                     values = c("#65A8D3","#67C5AB","#E69D67","#847AB7
"))+
  scale_fill_manual(name="Feeding method",labels=c("Breastmilk","Breas
tmilk_vitaminD","Partial breastmilk","Formula"),values = c("#65A8D3","
#67C5AB","#E69D67","#847AB7"))+
  scale_shape_manual(name="Feeding method",labels=c("Breastmilk","Brea
```

```
stmilk_vitaminD","Partial breastmilk","Formula"),values=c(15, 16, 17,
18))+
  stat_cor(method="spearman",aes(color = FED_PRAC_LIGHT_NEW),show.lege
nd = FALSE,r.accuracy=0.01,p.accuracy=0.01)+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, size=1)
)+
  labs(x="Fine motor",y="Shannon Index")+
  ggtitle("Fine motor_Shannon")+
  theme(plot.title=element_text(hjust=0.5))+
  geom_text(x=55, y=2.7, label="R = 0.37, p < 0.01")
p7

## `geom_smooth()` using formula = 'y ~ x'

#Communication
cor.test(Data.Subsample.final.Alpha$Shannon,metadata$asq_9_total_commu
nication.total.,method="spearman", exact=F)
p8<-ggplot(metadata, aes(x = asq_9_total_communication.total., y = sha
n))+
  geom_point(aes(color = FED_PRAC_LIGHT_NEW, shape = FED_PRAC_LIGHT_NE
W))+
  geom_smooth(aes(color = FED_PRAC_LIGHT_NEW), method ="lm",
              se = FALSE, fullrange = TRUE)+
  scale_color_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),
                     values = c("#65A8D3","#67C5AB","#E69D67","#847AB7
"))+
  scale_fill_manual(name="Feeding method",labels=c("Breastmilk","Breas
tmilk_vitaminD","Partial breastmilk","Formula"),values = c("#65A8D3","
#67C5AB","#E69D67","#847AB7"))+
  scale_shape_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),values=c(15, 16, 17,
18))+
  stat_cor(method="spearman",aes(color = FED_PRAC_LIGHT_NEW),show.lege
nd = FALSE,r.accuracy=0.01,p.accuracy=0.01)+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, size=1)
)+
  labs(x="Communication",y="Shannon Index")+
  ggtitle("Communication_Shannon")+
  theme(plot.title=element_text(hjust=0.5))+
```

```
  geom_text(x=50, y=2.7, label="R = 0.37, p < 0.01")
p8

## `geom_smooth()` using formula = 'y ~ x'

#Personal social
cor.test(Data.Subsample.final.Alpha$Shannon,metadata$asq_9_total_perso
nal_social,method="spearman", exact=F)
p9<-ggplot(metadata, aes(x = asq_9_total_personal_social, y = shan))+
  geom_point(aes(color = FED_PRAC_LIGHT_NEW, shape = FED_PRAC_LIGHT_NE
W))+
  geom_smooth(aes(color = FED_PRAC_LIGHT_NEW), method ="lm",
              se = FALSE, fullrange = TRUE)+
  scale_color_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),
                     values = c("#65A8D3","#67C5AB","#E69D67","#847AB7
"))+
  scale_fill_manual(name="Feeding method",labels=c("Breastmilk","Breas
tmilk_vitaminD","Partial breastmilk","Formula"),values = c("#65A8D3","
#67C5AB","#E69D67","#847AB7"))+
  scale_shape_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),values=c(15, 16, 17,
18))+
  stat_cor(method="spearman",aes(color = FED_PRAC_LIGHT_NEW),show.lege
nd = FALSE,r.accuracy=0.01,p.accuracy=0.01)+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, size=1)
)+
  labs(x="Personal social",y="Shannon Index")+
  ggtitle("Personal social_Shannon")+
  theme(plot.title=element_text(hjust=0.5))+
  geom_text(x=50, y=2.7, label="R = 0.33, p < 0.01")
p9

## `geom_smooth()` using formula = 'y ~ x'

#Problem solving
cor.test(Data.Subsample.final.Alpha$Shannon,metadata$asq_9_total_probl
emsolving,method="spearman", exact=F)
p10<-ggplot(metadata, aes(x = asq_9_total_problemsolving, y = shan))+
  geom_point(aes(color = FED_PRAC_LIGHT_NEW, shape = FED_PRAC_LIGHT_NE
W))+
  geom_smooth(aes(color = FED_PRAC_LIGHT_NEW), method ="lm",
              se = FALSE, fullrange = TRUE)+
  scale_color_manual(name="Feeding method",labels=c("Breastmilk","Brea
```

```
stmilk_vitaminD","Partial breastmilk","Formula"),
                      values = c("#65A8D3","#67C5AB","#E69D67","#847AB7
"))+
  scale_fill_manual(name="Feeding method",labels=c("Breastmilk","Breas
tmilk_vitaminD","Partial breastmilk","Formula"),values = c("#65A8D3","
#67C5AB","#E69D67","#847AB7"))+
  scale_shape_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),values=c(15, 16, 17,
18))+
  stat_cor(method="spearman",aes(color = FED_PRAC_LIGHT_NEW),show.lege
nd = FALSE,r.accuracy=0.01,p.accuracy=0.01)+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, size=1)
)+
  labs(x="Problem solving",y="Shannon Index")+
  ggtitle("Problem solving_Shannon")+
  theme(plot.title=element_text(hjust=0.5))+
  geom_text(x=49, y=2.7, label="R = 0.33, p < 0.01")
p10

## `geom_smooth()` using formula = 'y ~ x'

png("Shannon_ASQ_5panels_correct_spearman_final_spearman_overall.png",
 res=300, height=9, width=13,units="in")
ggarrange(
  p6, p7,p8,p9,p10, labels = c("A","B","C","D","E"),
  common.legend = TRUE, legend = "bottom"
)

## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'

while (!is.null(dev.list()))  dev.off()
```

Figure 3. The associations between inverse Simpson index and ASQ by different feeding methods at 3 months

```
#gross motor
cor.test(Data.Subsample.final.Alpha$Invsimpson,metadata$asq_9_total_gr
ossmotor,method="spearman",exact=F)
p11<-ggplot(metadata, aes(x = asq_9_total_grossmotor, y = invismp))+
  geom_point(aes(color = FED_PRAC_LIGHT_NEW, shape = FED_PRAC_LIGHT_NE
```

```
W))+
  geom_smooth(aes(color = FED_PRAC_LIGHT_NEW), method ="lm",
              se = FALSE, fullrange = TRUE)+
  scale_color_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),
                     values = c("#65A8D3","#67C5AB","#E69D67","#847AB7
"))+
  scale_fill_manual(name="Feeding method",labels=c("Breastmilk","Breas
tmilk_vitaminD","Partial breastmilk","Formula"),values = c("#65A8D3","
#67C5AB","#E69D67","#847AB7"))+
  scale_shape_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),values=c(15, 16, 17,
18))+
  stat_cor(method="spearman",aes(color = FED_PRAC_LIGHT_NEW),show.lege
nd = FALSE,r.accuracy=0.01,p.accuracy=0.01)+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, size=1)
)+
  labs(x="Gross motor",y="Inverse Simpson Index")+
  ggtitle("Gross motor_Inverse Simpson")+
  theme(plot.title=element_text(hjust=0.5))+
  geom_text(x=50, y=12.5, label="R = 0.05, p = 0.67")
p11

## `geom_smooth()` using formula = 'y ~ x'

# fine motor
cor.test(Data.Subsample.final.Alpha$Invsimpson,metadata$asq_9_total_fi
nemotor,method="spearman", exact=F)
p12<-ggplot(metadata, aes(x = asq_9_total_finemotor, y = invismp))+
  geom_point(aes(color = FED_PRAC_LIGHT_NEW, shape = FED_PRAC_LIGHT_NE
W))+
  geom_smooth(aes(color = FED_PRAC_LIGHT_NEW), method ="lm",
              se = FALSE, fullrange = TRUE)+
  scale_color_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),
                     values = c("#65A8D3","#67C5AB","#E69D67","#847AB7
"))+
  scale_fill_manual(name="Feeding method",labels=c("Breastmilk","Breas
tmilk_vitaminD","Partial breastmilk","Formula"),values = c("#65A8D3","
#67C5AB","#E69D67","#847AB7"))+
  scale_shape_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),values=c(15, 16, 17,
18))+
```

```
  stat_cor(method="spearman",aes(color = FED_PRAC_LIGHT_NEW),show.lege
nd = FALSE,r.accuracy=0.01,p.accuracy=0.01)+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, size=1)
)+
  labs(x="Fine motor",y="Inverse Simpson Index")+
  ggtitle("Fine motor_Inverse Simpson")+
  theme(plot.title=element_text(hjust=0.5))+
  geom_text(x=53, y=12.5, label="R = 0.32, p = 0.01")
p12

## `geom_smooth()` using formula = 'y ~ x'

# communication
cor.test(Data.Subsample.final.Alpha$Invsimpson,metadata$asq_9_total_co
mmunication.total.,method="spearman", exact=F)
p13<-ggplot(metadata, aes(x = asq_9_total_communication.total., y = in
vismp))+
  geom_point(aes(color = FED_PRAC_LIGHT_NEW, shape = FED_PRAC_LIGHT_NE
W))+
  geom_smooth(aes(color = FED_PRAC_LIGHT_NEW), method ="lm",
              se = FALSE, fullrange = TRUE)+
  scale_color_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),
                     values = c("#65A8D3","#67C5AB","#E69D67","#847AB7
"))+
  scale_fill_manual(name="Feeding method",labels=c("Breastmilk","Breas
tmilk_vitaminD","Partial breastmilk","Formula"),values = c("#65A8D3","
#67C5AB","#E69D67","#847AB7"))+
  scale_shape_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),values=c(15, 16, 17,
18))+
  stat_cor(method="spearman",aes(color = FED_PRAC_LIGHT_NEW),show.lege
nd = FALSE,r.accuracy=0.01,p.accuracy=0.01)+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, size=1)
)+
  labs(x="Communication",y="Inverse Simspon Index")+
  ggtitle("Communication_Inverse Simspon")+
  theme(plot.title=element_text(hjust=0.5))+
  geom_text(x=50, y=12.5, label="R = 0.33, p < 0.01")
p13
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
#personal social
cor.test(Data.Subsample.final.Alpha$Invsimpson,metadata$asq_9_total_pe
rsonal_social,method="spearman", exact=F)
p14<-ggplot(metadata, aes(x = asq_9_total_personal_social, y =invismp)
)+
  geom_point(aes(color = FED_PRAC_LIGHT_NEW, shape = FED_PRAC_LIGHT_NE
W))+
  geom_smooth(aes(color = FED_PRAC_LIGHT_NEW), method ="lm",
              se = FALSE, fullrange = TRUE)+
  scale_color_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),
                     values = c("#65A8D3","#67C5AB","#E69D67","#847AB7
"))+
  scale_fill_manual(name="Feeding method",labels=c("Breastmilk","Breas
tmilk_vitaminD","Partial breastmilk","Formula"),values = c("#65A8D3","
#67C5AB","#E69D67","#847AB7"))+
  scale_shape_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),values=c(15, 16, 17,
18))+
  stat_cor(method="spearman",aes(color = FED_PRAC_LIGHT_NEW),show.lege
nd = FALSE,r.accuracy=0.01,p.accuracy=0.01)+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, size=1)
)+
  labs(x="Personal social",y="Inverse Simpson Index")+
  ggtitle("Personal social_Inverse Simpson")+
  theme(plot.title=element_text(hjust=0.5))+
  geom_text(x=50, y=12.3, label="R = 0.28, p = 0.03")
p14
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
#problem solving
cor.test(Data.Subsample.final.Alpha$Invsimpson,metadata$asq_9_total_pr
oblemsolving,method="spearman", exact=F)
p15<-ggplot(metadata, aes(x = asq_9_total_problemsolving, y = invismp)
)+
  geom_point(aes(color = FED_PRAC_LIGHT_NEW, shape = FED_PRAC_LIGHT_NE
W))+
  geom_smooth(aes(color = FED_PRAC_LIGHT_NEW), method ="lm",
              se = FALSE, fullrange = TRUE)+
  scale_color_manual(name="Feeding method",labels=c("Breastmilk","Brea
```

```
stmilk_vitaminD","Partial breastmilk","Formula"),
                      values = c("#65A8D3","#67C5AB","#E69D67","#847AB7
"))+
  scale_fill_manual(name="Feeding method",labels=c("Breastmilk","Breas
tmilk_vitaminD","Partial breastmilk","Formula"),values = c("#65A8D3","
#67C5AB","#E69D67","#847AB7"))+
  scale_shape_manual(name="Feeding method",labels=c("Breastmilk","Brea
stmilk_vitaminD","Partial breastmilk","Formula"),values=c(15, 16, 17,
18))+
  stat_cor(method="spearman",aes(color = FED_PRAC_LIGHT_NEW),show.lege
nd = FALSE,r.accuracy=0.01,p.accuracy=0.01)+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, size=1)
)+
  labs(x="Problem solving",y="Inverse Simpson Index")+
  ggtitle("Problem solving_Inverse Simpson")+
  theme(plot.title=element_text(hjust=0.5))+
  geom_text(x=50, y=12.5, label="R = 0.27, p = 0.03")
p15

## `geom_smooth()` using formula = 'y ~ x'

png("Inv Simpson_ASQ_5panels_correct_spearman_final_spearman_overall.p
ng", res=300, height=9, width=13,units="in")
ggarrange(
  p11, p12,p13,p14,p15, labels = c("A","B","C","D","E"),
  common.legend = TRUE, legend = "bottom"
)

while (!is.null(dev.list()))  dev.off()
```

Table 3. The associations between beta diversity of the infant gut microbiota and each of the five ASQ scales

```
#gross motor
#Sorensen#
a<-metadata$asq_9_total_grossmotor
PERMANOVA(Data.Subsample.genus_37wks[,-c(1:3,254)],a,TRUE,9999)
#Bray-Curtis
PERMANOVA(Data.Subsample.genus_37wks[,-c(1:3,254)],a,FALSE,9999)

#fine motor
#Sorensen#
a<-metadata$asq_9_total_finemotor
PERMANOVA(Data.Subsample.genus_37wks[,-c(1:3,254)],a,TRUE,9999)
#Bray-Curtis
```

```
PERMANOVA(Data.Subsample.genus_37wks[,-c(1:3,254)],a,FALSE,9999)

#Communication#
a<-metadata$asq_9_total_communication.total.
#Sorenson
PERMANOVA(Data.Subsample.genus_37wks[,-c(1:3,254)],a,TRUE,9999)
#Bray Curtis
PERMANOVA(Data.Subsample.genus_37wks[,-c(1:3,254)],a,FALSE,9999)

#Personal Social#
a<-metadata$asq_9_total_personal_social
#Sorenson
PERMANOVA(Data.Subsample.genus_37wks[,-c(1:3,254)],a,TRUE,9999)
#Bray Curtis
PERMANOVA(Data.Subsample.genus_37wks[,-c(1:3,254)],a,FALSE,9999)

#Problem Solving#
a<-metadata$asq_9_total_problemsolving
#Sorenson
PERMANOVA(Data.Subsample.genus_37wks[,-c(1:3,254)],a,TRUE,9999)
#Bray Curtis
PERMANOVA(Data.Subsample.genus_37wks[,-c(1:3,254)],a,FALSE,9999)

###Multivariate analysis###
OTUS<-Data.Subsample.genus_37wks[,-c(1:3,254)]
#Gross motor
#Sorensen
Data.Dist<-vegdist(OTUS,method="bray", binary=TRUE)
adonis2(Data.Dist~asq_9_total_grossmotor+FED_PRAC_LIGHT_NEW+antibiotic
s_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+ESTWKSGEST+PRE_BMI+m
aternal_age,data=metadata,permutations=9999)
#Bray-Curtis
Data.Dist<-vegdist(OTUS,method="bray", binary=FALSE)
adonis2(Data.Dist~asq_9_total_grossmotor+FED_PRAC_LIGHT_NEW+antibiotic
s_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+ESTWKSGEST+PRE_BMI+m
aternal_age,data=metadata,permutations=9999)

#Fine motor
#Sorensen
Data.Dist<-vegdist(OTUS,method="bray", binary=TRUE)
adonis2(Data.Dist~asq_9_total_finemotor+FED_PRAC_LIGHT_NEW+antibiotics
_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+ESTWKSGEST+PRE_BMI+ma
ternal_age,data=metadata,permutations=9999)
#Bray-Curtis
Data.Dist<-vegdist(OTUS,method="bray", binary=FALSE)
adonis2(Data.Dist~asq_9_total_finemotor+FED_PRAC_LIGHT_NEW+antibiotics
```

```
_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+ESTWKSGEST+PRE_BMI+ma
ternal_age,data=metadata,permutations=9999)

#Communication#
#Sorenson
Data.Dist<-vegdist(OTUS,method="bray", binary=TRUE)
adonis2(Data.Dist~asq_9_total_communication.total.+FED_PRAC_LIGHT_NEW+
antibiotics_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+ESTWKSGEST
+PRE_BMI+maternal_age,data=metadata,permutations=9999)
#Bray Curtis
Data.Dist<-vegdist(OTUS,method="bray", binary=FALSE)
adonis2(Data.Dist~asq_9_total_communication.total.+FED_PRAC_LIGHT_NEW+
antibiotics_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+ESTWKSGEST
+PRE_BMI+maternal_age,data=metadata,permutations=9999)

#Personal Social#
#Sorenson
Data.Dist<-vegdist(OTUS,method="bray", binary=TRUE)
adonis2(Data.Dist~asq_9_total_personal_social+FED_PRAC_LIGHT_NEW+antib
iotics_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+ESTWKSGEST+PRE_
BMI+maternal_age,data=metadata,permutations=9999)
#Bray Curtis
Data.Dist<-vegdist(OTUS,method="bray", binary=FALSE)
adonis2(Data.Dist~asq_9_total_personal_social+FED_PRAC_LIGHT_NEW+antib
iotics_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+ESTWKSGEST+PRE_
BMI+maternal_age,data=metadata,permutations=9999)

#Problem Solving#
#Sorenson
Data.Dist<-vegdist(OTUS,method="bray", binary=TRUE)
adonis2(Data.Dist~asq_9_total_problemsolving+FED_PRAC_LIGHT_NEW+antibi
otics_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+ESTWKSGEST+PRE_B
MI+maternal_age,data=metadata,permutations=9999)
#Bray Curtis
Data.Dist<-vegdist(OTUS,method="bray", binary=FALSE)
adonis2(Data.Dist~asq_9_total_problemsolving+FED_PRAC_LIGHT_NEW+antibi
otics_since_birth+SEX+MD_FINAL_ROUTE+Race_new+EDU_LVL+ESTWKSGEST+PRE_B
MI+maternal_age,data=metadata,permutations=9999)
```

Figure 4. The significant associations between Bray-Curtis dissimilarity matrix and ASQ scales

```
legend.col <- function(col, lev){
  opar <- par
  n <- length(col)
  bx <- par("usr")
  box.cx <- c(bx[2] + (bx[2] - bx[1]) / 1000,
```

```
              bx[2] + (bx[2] - bx[1]) / 1000 + (bx[2] - bx[1]) / 50)
  box.cy <- c(bx[3], bx[3])
  box.sy <- (bx[4] - bx[3]) / n
  xx <- rep(box.cx, each = 2)
  par(xpd = TRUE)
  for(i in 1:n){
    yy <- c(box.cy[1] + (box.sy * (i - 1)),
            box.cy[1] + (box.sy * (i)),
            box.cy[1] + (box.sy * (i)),
            box.cy[1] + (box.sy * (i - 1)))
    polygon(xx, yy, col = col[i], border = col[i])
  }
  par(new = TRUE)
  plot(0, 0, type = "n",
       ylim = c(min(lev), max(lev)),
       yaxt = "n", ylab = "",
       xaxt = "n", xlab = "",
       frame.plot = FALSE)
  axis(side = 4, las = 2, tick = FALSE, line = .25)
  par <- opar
}

#Fine motor
a<-metadata$asq_9_total_finemotor
#Bray-Curtis
PERMANOVA(Data.Subsample.genus_37wks[,-c(1:3,254)],a,FALSE,9999)

#Communication#
a<-metadata$asq_9_total_communication.total.
#Bray Curtis

df.Genus.Sor_2<-Sor.bray.pcoa(Data.Subsample.genus_37wks[,-c(1:3,254)]
,binary=FALSE)

shapes<-c(21,22,23,24)
shapes<-shapes[as.numeric(metadata$FED_PRAC_LIGHT_NEW)]
shapes

#png("Beta_diversity_fine_comm.png", res=300, height=5, width=10,units
="in")
par(mfrow= c(1,2),mar=c(4,4.1,4,4.1))
rbPal <- colorRampPalette(c('white','black'))
a<-metadata$asq_9_total_finemotor
b<-rank(a)
Col <- rbPal(20)[as.numeric(cut(b,breaks = 20))]
{plot(df.Genus.Sor_2,bg=Col,xlab="PC1(25.1%)",ylab="PC2(13.6%)",main =
```

```
"A.Bray-Curtis and fine motor",,pch=shapes,cex.axis=1.5,cex.lab=1.5,ce
x.main=1.5)
text(-0.3,0.4,"p-value<0.01",cex = 1)
legend(0.06,0.53,c("Breastmilk","Breastmilk_vitaminD","Partial breastm
ilk","Formula"), pch=c(21,22,23,24),cex = 0.7)
legend.col(col = rbPal(10), lev = a)}


a<-metadata$asq_9_total_communication.total
b<-rank(a)
Col <- rbPal(20)[as.numeric(cut(b,breaks = 20))]
{plot(df.Genus.Sor_2,bg=Col,xlab="PC1(25.1%)",ylab="PC2(13.6%)",main =
"B.Bray-Curtis and communication",pch=shapes,cex.axis=1.5,cex.lab=1.5,
cex.main=1.5)
text(-0.3,0.4,"p-value=0.01",cex = 1)
legend(0.06,0.53,c("Breastmilk","Breastmilk_vitaminD","Partial breastm
ilk","Formula"), pch=c(21,22,23,24),cex = 0.7)
legend.col(col = rbPal(10), lev = a)}

#while (!is.null(dev.list()))  dev.off()
```

Figure 5. The gut microbiota composition of infant stool samples organized by cluster

```
TaxName<-read.table("/Users/busihan/Desktop/MARCH\ B3m_ASQ_updated/sta
bility.trim.contigs.good.unique.good.filter.unique.precluster.pick.pds
.wang.pick.tx.1.cons.taxonomy",header=TRUE, fill=TRUE,row.names=NULL)
head(TaxName)
Edit.Taxname<-function(n,level){
  if(level=="Genus"|level==1){
    n<-as.matrix(n)
    for (i in 1:4){
      n<-gsub('^.*?;', '', n)
    }
    n<-gsub(';',' ',n)
    n<-gsub('\\(100)','',n)
    n<-data.frame(n)
    n<-separate(n, col=1,into=c("Family","Genus"), sep=" ")
    x<-ifelse(n$Genus%in%c("unclassified","uncultured"), paste(n$Genus
, n$Family), paste(n$Genus,n$Other1,n$Other2))
    n<-as.matrix(x)
    return(n)
  }else if(level=="Family"|level==2){
    n<-as.matrix(n)
    for (i in 1:3){
      n<-gsub('^.*?;', '', n)
    }
    n<-gsub(';',' ',n)
```

```
    n<-gsub('\\(100)','',n)
    n<-data.frame(n)
    n<-separate(n,col=1, into=c("Order","Family","Genus"), sep=" ")
    x<-ifelse(n$Family%in%c("unclassified","uncultured"), paste(n$Orde
r, n$Family), paste(n$Family))
    n<-as.matrix(x)
    return(n)
  }else if(level=="Order"|level==3){
    n<-as.matrix(n)
    for (i in 1:2){
      n<-gsub('^.*?;', '', n)
    }
    n<-gsub(';',' ',n)
    n<-gsub('\\(100)','',n)
    n<-data.frame(n)
    n<-separate(n,col=1, into=c("Class","Order","Family","Genus"), sep
=" ")
    x<-ifelse(n$Order%in%c("unclassified","uncultured"), paste(n$Class
, n$Order), paste(n$Order))
    n<-as.matrix(x)
    return(n)
  }else if(level=="Class"|level==4){
    n<-as.matrix(n)
    for (i in 1){
      n<-gsub('^.*?;', '', n)
    }
    n<-gsub(';',' ',n)
    n<-gsub('\\(100)','',n)
    n<-data.frame(n)
    n<-separate(n,col=1, into=c("Phylum","Class","Order","Family","Gen
us"), sep=" ")
    x<-ifelse(n$Class%in%c("unclassified","uncultured"), paste(n$Phylu
m, n$Class), paste(n$Class))
    n<-as.matrix(x)
    return(n)
  }else if(level=="Phylum"|level==5){
    n<-as.matrix(n)
    n<-gsub('[(0-9);""]{1,}', '_', n)
    n<-gsub('^.*?_', '', n)
    n<-gsub('_.*', '', n)
  }
}

TaxName<-Edit.Taxname(TaxName$Taxonomy,level=1)
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 250 rows
 [1, 2, 3, 4, 5, 6,
## 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].

OTU<-Data.Subsample.genus_37wks[,c(4:253)]
colnames(OTU) <- TaxName
mat = as.matrix(OTU)
rel_mat = make_relative(mat)
rel_otu<-as.data.frame(t(rel_mat))
colnames(rel_otu)<-Data.Subsample.genus_37wks$Group

dist.JSD <- function(inMatrix, pseudocount=0.000001, ...) {
  KLD <- function(x,y) sum(x *log(x/y))
  JSD<- function(x,y) sqrt(0.5 * KLD(x, (x+y)/2) + 0.5 * KLD(y, (x+y)/
2))
  matrixColSize <- length(colnames(inMatrix))
  matrixRowSize <- length(rownames(inMatrix))
  colnames <- colnames(inMatrix)
  resultsMatrix <- matrix(0, matrixColSize, matrixColSize)

  inMatrix = apply(inMatrix,1:2,function(x) ifelse (x==0,pseudocount,x
))

  for(i in 1:matrixColSize) {
    for(j in 1:matrixColSize) {
      resultsMatrix[i,j]=JSD(as.vector(inMatrix[,i]),
                             as.vector(inMatrix[,j]))
    }
  }
  colnames -> colnames(resultsMatrix) -> rownames(resultsMatrix)
  as.dist(resultsMatrix)->resultsMatrix
  attr(resultsMatrix, "method") <- "dist"
  return(resultsMatrix)
}

data.dist=dist.JSD(rel_otu)

pam.clustering=function(x,k) {
  require(cluster)
  cluster = as.vector(pam(as.dist(x), k, diss=TRUE)$clustering)
  return(cluster)
}

data.cluster=pam.clustering(data.dist, k=3)
data<-rel_otu
nclusters = index.G1(t(data), data.cluster, d = data.dist, centrotypes
```

```
 = "medoids")
nclusters=NULL

for (k in 1:10) {
  if (k==1) {
    nclusters[k]=NA
  } else {
    data.cluster_temp=pam.clustering(data.dist, k)
    nclusters[k]=index.G1(t(data),data.cluster_temp,  d = data.dist,
                          centrotypes = "medoids")
  }
}
plot(nclusters, type="h", xlab="k clusters", ylab="CH index",main="Opt
imal number of clusters")   #k=3

#k=3
cluster=data.frame(row.names = colnames(data),Cluster=data.cluster)
OTU_new<-Data.Subsample.genus_37wks[,c(1,4:253)]
OTU_new$Group
cluster['Group'] <- Data.Subsample.genus_37wks$Group
OTU_new<-merge(OTU_new,cluster,by="Group")
metadata<-merge(metadata,cluster,by="Group")
cluster$Cluster<-as.factor(cluster$Cluster)
summary(cluster)

# rank the stacked bars #
StackedBarPlot<-function(OTU,Group="Samples",TaxName,N=19,Title="Stack
ed Bar Chart"){
  Rowsum<-as.matrix(rowSums(OTU))
  abun<-matrix(0,nrow=nrow(OTU),ncol=ncol(OTU))
  for (i in 1:nrow(OTU)){
    for (j in 1:ncol(OTU)){
      abun[i,j]=(OTU[i,j])/(Rowsum[i])*100
    }
  }
  colnames(abun)<-TaxName
  abun<-abun[,order(-colSums(abun))]
  taxa_list<-colnames(abun)[1:N]
  taxa_list<-taxa_list[!grepl("unclassified unclassified",taxa_list)]
  N<-length(taxa_list)
  new_x<-data.frame(abun[,colnames(abun) %in% taxa_list],Others=rowSum
s(abun[,!colnames(abun) %in% taxa_list]))
  if (ncol(new_x)>(N+1)){
    Other<-rowSums(new_x[,c((N+1):ncol(new_x))])
    new_x<-new_x[,c(1:N)]
    new_x$Other<-Other
```

```
  }
  abun_groups<-cbind(Group,new_x)
  new_x <- abun_groups
  grouping_info<-new_x$Group
  new_x2<-new_x[,-1]
  tempname<-c(taxa_list,"Other")
  colnames(new_x2)<-tempname
  df<-NULL
  for (i in 1:dim(new_x2)[2]){
    tmp<-data.frame(row.names=NULL,Sample=rownames(new_x2),Taxa=rep(co
lnames(new_x2)[i],dim(new_x2)[1]),Value=new_x2[,i],Type=grouping_info)
    if(i==1){df<-tmp} else {df<-rbind(df,tmp)}
  }
  colours <- c("#F0A3FF", "#0075DC", "#993F00","#4C005C","#2BCE48","#F
FCC99","#808080","#94FFB5","#8F7C00","#9DCC00","#C20088","#003380","#F
FA405","#FFA8BB","#426600","#FF0010","#5EF1F2","#00998F","#740AFF","#9
90000","#FFFF00");
  p<-ggplot(df,aes(Sample,Value,fill=fct_reorder(Taxa,Value)))+geom_ba
r(stat="identity")+facet_grid(. ~ Type, drop=TRUE,scale="free",space="
free_x")
  p<-p+scale_fill_manual(values=colours[1:(N+1)])
  p<-p+theme_bw(base_size = 24)+ylab("Relative Abundance %")+ggtitle("
Top 19 taxa in 3 clusters")+xlab("Clusters")
  p<-p+guides(fill=guide_legend(title="Taxa"))
  p<-p+scale_y_continuous(expand = c(0,0))+theme(strip.background = el
ement_rect(fill="gray85"))+theme(panel.spacing = unit(0, "lines"))
  p<-p+theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5))+the
me(axis.title.x=element_blank(),axis.text.x=element_blank(),axis.ticks
.x=element_blank())
  print(p)
  return(df)
}

#png("Barchart_cluster", res=300, height=7, width=11,units="in")
a<-StackedBarPlot(OTU=OTU_new[,c(2:251)],TaxName=TaxName,Group = OTU_n
ew$Cluster)

#while (!is.null(dev.list()))  dev.off()
```

Figure 6. The composition of the top five overall most abundant taxa presented by cluster

```
## chose average(relative) abundance > 1%
Subset.Taxa<-function(OTUS,TaxName,CutOff=1){
  colnames(OTUS)<-TaxName
  row<-rowSums(OTUS)
  row<-sum(row)
  col<-colSums(OTUS)
```

```
  ratio<-as.matrix(col/row*100)
  ratio<-cbind(TaxName,ratio)
  subset<-data.frame(ratio[ratio[,2]>=CutOff,])
  subset<-data.frame(subset[!subset$X1=="unclassified unclassified",])
  newOTUS<-data.frame(OTUS[,colnames(OTUS) %in% subset$X1])
  colname<-colnames(newOTUS)
  colnames(newOTUS)<-gsub("\\."," ",colname)
  return(newOTUS)
}

rownames(OTU_new)<-OTU_new$Group
newOTUS<-Subset.Taxa(OTU_new[,c(2:251)],TaxName=TaxName,CutOff=1)
newOTUS<-as.matrix(newOTUS)
rel_mat_1<-make_relative(newOTUS)
rel_mat_1<-rel_mat_1 * 100
rank(colSums(rel_mat_1))

# choose the following 5 taxa
TaxName<-as.data.frame(TaxName)
colnames(OTU_new)[2:251]<- TaxName$V1
OTU_new$`Lachnospiraceae_unclassified  `   # column 4, 3rd
OTU_new$`Bifidobacterium  `  # column 3, 1st
OTU_new$`Bacteroides  `  # column 6, 4th
OTU_new$`Veillonella  ` # column 2, 2nd
OTU_new$`Escherichia/Shigella  ` # column 7, 5th

# calculate the rel abun in whole otu table
OTU_new<-as.data.frame(OTU_new)
OTU_new_rel<-OTU_new[,c(2:251)]
OTU_new_rel<-as.matrix(OTU_new_rel)
OTU_new_rel <- make_relative(OTU_new_rel)
OTU_new_rel<-OTU_new_rel * 100
OTU_new_rel<-as.data.frame(cbind(OTU_new_rel,OTU_new$Cluster))

cluster<-OTU_new_rel[,c(3,2,5,1,6,251)]
names(cluster)[5]<-"Escherichia Shigella"
cluster<-melt(cluster, id = "V251")
cluster$V251[cluster$V251 == "1"]<-"Cluster1"
cluster$V251[cluster$V251 == "2"]<-"Cluster2"
cluster$V251[cluster$V251 == "3"]<-"Cluster3"

png("Cluster_top5_update_color", res=300, height=5, width=15,units="in
")
ggplot(cluster, aes(x=V251, y=value, fill=variable)) +
  labs(title=NULL,x=NULL, y = "Relative abundance %")+
  scale_color_manual(values=c("#173F5F", "#20639B", "#3CAEA3","#F6D55C
```

```
", "#ED553B"), name=NULL)+
  scale_fill_manual(values=c("#173F5F", "#20639B", "#3CAEA3","#F6D55C"
, "#ED553B"))+
  theme_classic()+
  theme(legend.position="top")+
  theme(legend.title=element_blank())+
  theme(axis.text.x = element_text(size=18, color="black", face="bold"
,angle=0))+
  theme(axis.text.y = element_text(size=18, color="black", face="bold"
,angle=0))+
  theme(axis.title.y = element_text(size=18, color="black", face="bold
",angle=90))+
  theme(legend.text = element_text(size=18, color="black", face="bold"
,angle=0))+
  geom_boxplot()
while (!is.null(dev.list()))  dev.off()
```

Figure 7. Shannon and inverse Simpson indices of gut microbial alpha diversity differs across the three clusters

```
shapiro.test(Data.Subsample.final.Alpha$Chao)    ## p-value =0.3152
shapiro.test(Data.Subsample.final.Alpha$Shannon)  ## p-value =0.5289
shapiro.test(Data.Subsample.final.Alpha$Invsimpson)  ## p-value = 0.00
8523
chao<-Data.Subsample.final.Alpha$Chao
shan<-Data.Subsample.final.Alpha$Shannon
invismp<-Data.Subsample.final.Alpha$Invsimpson
a<-as.factor(OTU_new$Cluster)

#Chao1
summary(aov(chao~a))  #p=0.113

#Shannon
summary(aov(shan~a))  #p=1.12e-07
TukeyHSD(aov(shan~a))

#Inverse Simp
kruskal.test(invismp~a) #p-value = 4.837e-07
dunn.test(invismp,a,altp = TRUE, method="bh")

labels<-c("Cluster1","Cluster2","Cluster3")
png("Alpha_diversity_3_clusters_updated.png", res=300, height=6, width
=16.7,units="in")
par(mfrow= c(1,3),mar=c(5, 5, 3, 1) + 0.1)
{boxplot(chao~a,main="A. Chao1 index of 3 clusters",xlab=NA,ylab="Chao
1 Index",cex.axis=2.5,cex.lab=2.5,cex.main=2.5, names=labels)
text(x=1.5,y=75,labels= "p-value=0.11", cex=2)}
```

```
{boxplot(shan~a,main="B. Shannon index of 3 clusters",ylab="Shannon In
dex",xlab=NA,cex.axis=2.5,cex.lab=2.5,cex.main=2.5,names=labels )
text(x=2,y=2.8,labels= "p-value<0.01", cex=2)
text(x=1,y=1.47,labels= "a", cex=2.3)
text(x=2,y=2.23,labels= "b", cex=2.3)
text(x=3,y=2.62,labels= "a", cex=2.3)}
{boxplot(invismp~a,main="C. Inverse Simpson index of 3 clusters",ylab=
"Inverse Simpson Index",xlab=NA,cex.axis=2.5,cex.lab=2.5,cex.main=2.5,
names=labels)
  text(x=1.7,y=12,labels= "p-value<0.01", cex=2)
  text(x=1,y=10.97,labels= "a", cex=2.3)
  text(x=2,y=7.2,labels= "b", cex=2.3)
  text(x=3,y=8,labels= "a", cex=2.3)}
while (!is.null(dev.list()))  dev.off()
```

Figure 8. The gut microbiota beta diversity is differed by cluster

```
#Sorensen
a<-as.factor(OTU_new$Cluster)
PERMANOVA(OTU_new[,c(2:251)],a,TRUE,9999)   #p=1e-04
Sor_cluster<-Sor.bray.pcoa(OTU_new[,c(2:252)],Dim=2,Color=OTU_new$Clus
ter,binary=TRUE)

Color<-ifelse(grepl("1", OTU_new$Cluster),"#000000", ifelse(grepl("2",
 OTU_new$Cluster),"#E79F00","#0072B2"))

#Bray-curtis
PERMANOVA(OTU_new[,c(2:252)],a,FALSE,9999)  #1e-04
Bray_cluster<-Sor.bray.pcoa(OTU_new[,c(2:252)],Dim=2,Color=OTU_new$Clu
ster,binary=FALSE)

shapes<-c(21,22,23,24)
shapes<-shapes[as.numeric(metadata$FED_PRAC_LIGHT_NEW)]
shapes

#png("Beta_diversity_3_clusters_beauty_shapes.png", res=300, height=5,
 width=10,units="in")
par(mfrow= c(1,2),mar=c(5, 5, 3, 1) + 0.1)
plot(Sor_cluster,cex.axis=1.5,cex.lab=1.5,cex.main=2,cex=2,col=1,
     pch=shapes,xlim=c(-.45,.5),ylim=c(-.3,.35),xlab="PC1 (22.5%)",yla
b="PC2 (11.8%)",bg=Color,main="A. Sorensen")
ordiellipse(Sor_cluster,OTU_new$Cluster,col=Color,lwd=2)
legend(0.2,-0.17,c("Cluster1","Cluster2","Cluster3"),
       pch=21,col=1,pt.bg=c("#000000","#E79F00","#0072B2"),cex = 0.8,y
.intersp = 0.72)
legend(0.1,0.35,c("Breastmilk","Breastmilk_vitaminD","Partial breastmi
lk","Formula"), pch=c(21,22,23,24),cex =0.8,y.intersp = 0.72)
text(-0.3,0.25, labels= "p-value<0.01",cex=1)
```

```
plot(Bray_cluster,cex.axis=1.5,cex.lab=1.5,cex.main=2,cex=2,col=1,
     pch=shapes,xlim=c(-.5,.65),ylim=c(-.35,.35),xlab="PC1 (25.1%)",yl
ab="PC2 (13.6%)",bg=Color,main="B. Bray-Curtis")
ordiellipse(Bray_cluster,OTU_new$Cluster,col=Color,lwd=2)
legend(0.34,-0.2,c("Cluster1","Cluster2","Cluster3"),
       pch=21,col=1,pt.bg=c("#000000","#E79F00","#0072B2"),cex=0.8,y.i
ntersp = 0.72)
legend(0.215,0.35,c("Breastmilk","Breastmilk_vitaminD","Partial breast
milk","Formula"), pch=c(21,22,23,24),cex=0.75,y.intersp = 0.72)
text(0.4,0.16, labels= "p-value<0.01",cex=1)

#while (!is.null(dev.list()))  dev.off()
```

Table 4. The associations between three clusters and ASQ scales

```
summary(metadata)
metadata$Cluster<-as.factor(metadata$Cluster)
###univariate regression###
# gross motor
grossmotor<-lm(asq_9_total_grossmotor~Cluster,data=metadata)
summary(grossmotor )
confint(grossmotor)

#Fine motor
finemotor<-lm(asq_9_total_finemotor~Cluster, data=metadata)
summary(finemotor)
confint(finemotor)

#Communication
Communication<-lm(asq_9_total_communication.total.~Cluster,data=metada
ta)
summary(Communication)
confint(Communication)

#Personal and social
personal<-lm(asq_9_total_personal_social~ Cluster,data=metadata)
summary(personal)
confint(personal)

#Problem solving
problem<-lm(asq_9_total_problemsolving~Cluster,data=metadata)
summary(problem)
confint(problem)
###multivariate regression###
#Gross motor
grossmotor<-lm(asq_9_total_grossmotor~Cluster+FED_PRAC_LIGHT_NEW+antib
```

```
iotics_since_birth+SEX+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+EDU_LVL+PRE_
BMI+maternal_age,data=metadata)
summary(grossmotor)
confint(grossmotor)
#Fine motor
finemotor<-lm(asq_9_total_finemotor~Cluster+FED_PRAC_LIGHT_NEW+antibio
tics_since_birth+SEX+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+EDU_LVL+PRE_BM
I+maternal_age, data=metadata)
summary(finemotor)
confint(finemotor)

#Communication
Communication<-lm(asq_9_total_communication.total.~Cluster+FED_PRAC_LI
GHT_NEW+antibiotics_since_birth+SEX++ESTWKSGEST+MD_FINAL_ROUTE+Race_ne
w+EDU_LVL+PRE_BMI+maternal_age,data=metadata)
summary(Communication)
confint(Communication)

#Personal and social
personal<-lm(asq_9_total_personal_social~Cluster+FED_PRAC_LIGHT_NEW+an
tibiotics_since_birth+SEX+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+EDU_LVL+P
RE_BMI+maternal_age,data=metadata)
summary(personal)
confint(personal)

#Problem solving
problem<-lm(asq_9_total_problemsolving~Cluster+FED_PRAC_LIGHT_NEW+anti
biotics_since_birth+SEX+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+EDU_LVL+PRE
_BMI+maternal_age,data=metadata)
summary(problem)
confint(problem)
```

Figure 9. The relationships between ASQ and relative abundance of specific taxa

```
OTU<-Data.Subsample.genus_37wks[,c(4:253)]
colnames(OTU) <- TaxName$V1
mat = as.matrix(OTU)
rel_mat = make_relative(mat)
rel_otu<-as.data.frame(t(rel_mat))
colnames(rel_otu)<-Data.Subsample.genus_37wks$Group
rel_otu<-t(rel_otu)

# based on cluster analysis in table 4
#Fine motor is negatively associated with cluster 2(bifidobacterium) c
ompared to cluster 1(Lachnospiraceae_unclassified)
metadata$Bifi<-rel_otu[,2]
shapiro.test(metadata$Bifi)  # p-value = 0.0001565
```

```
metadata$Bifi_percent<-metadata$Bifi*100

c<-ggplot(metadata,aes(x=asq_9_total_finemotor, y= Bifi_percent)) +
  geom_smooth(method='lm',se=FALSE, color='darkblue')+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, linewid
th=1))+
  geom_point()+
  labs(x='Fine motor', y='Bifidobacterium %', title='Fine motor_Bifido
bacterium') +
  theme(plot.title = element_text(hjust=0.5, size=15, face='bold'))+
  theme(axis.text.x=element_text(size=12),axis.text.y=element_text(siz
e=12),axis.title=element_text(size=12,face="bold"))+
  stat_cor(method="pearson",show.legend = FALSE,r.accuracy=0.01,p.accu
racy=0.01)
c

## `geom_smooth()` using formula = 'y ~ x'

metadata$Lach<-rel_otu[,3]
metadata$Lach_percent<-metadata$Lach*100

d<-ggplot(metadata,aes(x=asq_9_total_finemotor, y= Lach_percent)) +
  geom_smooth(method='lm',se=FALSE, color='darkblue')+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, linewid
th=1))+
  geom_point()+
  labs(x='Fine motor', y='Lachnospiraceae_unclassified %', title='Fine
 motor_Lachnospiraceae unclassified') +
  theme(plot.title = element_text(hjust=0.5, size=15, face='bold'))+
  stat_cor(method="pearson",show.legend = FALSE,r.accuracy=0.01,p.accu
racy=0.01)+
  theme(axis.text.x=element_text(size=12),axis.text.y=element_text(siz
e=12),axis.title=element_text(size=12,face="bold"))
d

#problem solving is negatively associated with cluster3(bacteriodes) c
ompared to cluster 1(lach)
metadata$Bacter<-rel_otu[,5]
metadata$Bacter_percent<-metadata$Bacter*100

f<-ggplot(metadata,aes(x=asq_9_total_problemsolving, y= Bacter_percent
```

```
)) +
  geom_smooth(method='lm',se=FALSE, color='darkblue')+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, linewid
th=1))+
  geom_point()+
  labs(x='Problem-solving', y='Bacteroides %', title='Problem-solving_
Bacteroides') +
  theme(plot.title = element_text(hjust=0.5, size=15, face='bold'))+
  stat_cor(method="pearson",show.legend = FALSE,r.accuracy=0.01,p.accu
racy=0.01)+
  theme(axis.text.x=element_text(size=12),axis.text.y=element_text(siz
e=12),axis.title=element_text(size=12,face="bold"))
f
## `geom_smooth()` using formula = 'y ~ x'

h<-ggplot(metadata,aes(x=asq_9_total_problemsolving, y=Lach_percent))
+
  geom_smooth(method='lm',se=FALSE, color='darkblue')+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, linewid
th=1))+
  geom_point()+
  labs(x='Problem-solving', y='Lachnospiraceae_unclassified %', title=
'Problem-solving_Lachnospiraceae unclassified') +
  theme(plot.title = element_text(hjust=0.5, size=15, face='bold'))+
  stat_cor(method="pearson",show.legend = FALSE,r.accuracy=0.01,p.accu
racy=0.01)+
  theme(axis.text.x=element_text(size=12),axis.text.y=element_text(siz
e=12),axis.title=element_text(size=12,face="bold"))
h

## `geom_smooth()` using formula = 'y ~ x'

png("problemsolving_bacter_Lach_finemotor_bifi_lach.png", res=300, hei
ght=10, width=12,units="in")
ggarrange(c,d,f,h, labels = c("A", "B","C","D"))

## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```

```
while (!is.null(dev.list()))  dev.off()
```

Figure 10. The frequency of feeding methods in the past 24 hours and past week at 3 months of age in each cluster

```
percent<-read.xlsx("Figure10_percent.xlsx")
summary(percent)

chisq.test(table(metadata$Cluster,metadata$FED_PRAC_LIGHT_NEW),simulat
e.p.value = TRUE)
p<-as.data.frame(chisq.post.hoc(table(metadata$Cluster,metadata$FED_PR
AC_LIGHT_NEW)))
p

chisq.test(table(metadata$Cluster,metadata$During.the.past.week..my.ba
by.ate.),simulate.p.value = TRUE)
p_1<-as.data.frame(chisq.post.hoc(table(metadata$Cluster,metadata$Duri
ng.the.past.week..my.baby.ate.)))
p_1

p1<-ggplot(percent, aes(x = Cluster_pastday, y = percent_pastday, fill
 =factor(fed_pastday), label = percent_pastday,color=factor(fed_pastda
y)))+
    geom_bar(stat = "identity")+
    labs(x= "Cluster", y = "Percentage %")+
   ggtitle("A. Feeding method in the past day") +
    geom_text(size = 5, position = position_stack(vjust = 0.5),color="
black")+
    theme_classic()+
    theme(legend.title=element_blank())+
    scale_fill_discrete(labels=c('Breastmilk', 'Breastmilk_vitaminD',
'Partial breastmilk', 'Formula'))+
  scale_color_discrete(labels=c('Breastmilk', 'Breastmilk_vitaminD', '
Partial breastmilk', 'Formula'))+
  theme(text = element_text(size = 5),axis.text = element_text(size =
17),axis.title = element_text(size = 16),legend.text = element_text(si
ze = 15),
        plot.title = element_text(size = 16))+
  annotate("text", x=1, y=105, label= "a",size=6)+
  annotate("text", x=2, y=105, label= "b",size=6)+
  annotate("text", x=3, y=105, label= "b",size=6)

p2<-ggplot(percent, aes(x = Cluster_pastweek, y = percent_pastweek, fi
ll =factor(fed_pastweek), label = percent_pastweek,color=factor(fed_pa
stweek)))+
    geom_bar(stat = "identity")+
    labs(x= "Cluster", y = "Percentage %")+
```

```
    ggtitle("B.Feeding method in the past week") +
    geom_text(size = 5, position = position_stack(vjust = 0.5),color="
black")+
    theme_classic()+
    theme(legend.title=element_blank())+
    theme(text = element_text(size = 5),axis.text = element_text(size
= 17),axis.title = element_text(size = 16),legend.text = element_text(
size = 15),
        plot.title = element_text(size = 16))+
    annotate("text", x=1, y=105, label= "a",size=6)+
    annotate("text", x=2, y=105, label= "b",size=6)+
    annotate("text", x=3, y=105, label= "a",size=6)

#png("Feeding_2_variables_cluster.png", res=300, height=5, width=10,un
its="in")
ggarrange(p1,p2)
```

```
## Warning: Removed 3 rows containing missing values (`position_stack(
)`).
## Removed 3 rows containing missing values (`position_stack()`).
```

```
#while (!is.null(dev.list()))  dev.off()
```

Chapter 3

Data preparation

```
library(vegan)
library(lubridate)
library(tidyr)
library(MASS)
library(car)
library(dunn.test)
library(ggplot2)
library(openxlsx)
library(Hmisc)
library(pairwiseAdonis)
library(Maaslin2)
library(mediation)
library(MeMoBootR)
library(Rfast)
library(energy)
library(tidyr)
library(phyloseq); packageVersion("phyloseq")
library(energy); packageVersion("energy")
library(LDM)
library(dplyr)


setwd("/Users/busihan/Desktop/2023Mar22_Aim2_double_check")
metadata<-read.csv("metadata_updated_Jan.csv",na="",header = T)
Data.Subsample.genus_37wks<-read.csv("Data.Subsample.genus_37wks.csv",
header = T,stringsAsFactors = T,row.names = 1)
cols<-c("antibiotics_since_birth","FED_PRAC_NEW","SEX","FED_PRAC_LIGHT
_NEW","MD_FINAL_ROUTE","Race_new","EDU_LVL")
summary(metadata)
metadata[cols]<-lapply(metadata[cols], factor)
sapply(metadata,class)

Data.Subsample.genus_37wks$Group
metadata$Group
temp<-merge(Data.Subsample.genus_37wks, metadata,by="Group")

Data.Subsample.genus_37wks<-temp[,c(1:(ncol(Data.Subsample.genus_37wks
)))]
metadata<-temp[,c(1,254:294)]

Data.Subsample.genus_37wks$Group
metadata$Group

Alpha<-function(OTU,Names="Sample",Groups="Sample"){
```

```
  Chao<-t(estimateR(OTU))
  Chao<-Chao[,2]
  Shannon<-diversity(OTU,index="shannon")
  Invsimpson<-diversity(OTU,index="invsimpson")
  OTU.Subsample.Alpha<-data.frame(Names,Groups,Chao,Shannon,Invsimpson
)
  return(OTU.Subsample.Alpha)
}
Sor.bray.pcoa<-function(OTUS,Dim=2,Color=1,binary,pch=16,Title="PCoA")
{
  Data.df<-vegdist(OTUS,method="bray", binary)
  Data.df.PCoA<-cmdscale(Data.df, k = Dim, eig = FALSE)
  Data.df.PCoA.eig<-cmdscale(Data.df, k = Dim, eig = TRUE)
  eig.Data.df.PCoA<-Data.df.PCoA.eig$eig
  eig.Data.df.PCoA.sum<-sum(eig.Data.df.PCoA)
  a<-(eig.Data.df.PCoA/eig.Data.df.PCoA.sum)*100
  xlab<-paste("PC1","(",round(a[1],1),"%",")",sep="")
  ylab<-paste("PC2","(",round(a[2],1),"%",")",sep="")
  if(binary==TRUE){
    main<-"Sorensen PCoA"
  }else(main<-"Bray-Curtis PCoA")
  plot(Data.df.PCoA, col=Color,
       main=Title,xlab=xlab,ylab=ylab,pch=c(pch))
  return(Data.df.PCoA)
}

PERMANOVA<-function(OTUS,Group,binary,iters=9999){
  Data.Dist<-vegdist(OTUS,method="bray", binary=binary)
  adonis2(Data.Dist~Group,permutations=iters)
}

PERMANOVA_pairwise<-function(OTUS,Group,binary,iters=9999){
  Data.Dist<-vegdist(OTUS,method="bray", binary=binary)
  pairwise.adonis(Data.Dist,Group)
}
```

Table 5. The associations between infant feeding methods of infants at 3 months of age and ASQ scores at 9 months of age

```
###FED_PRAC_NEW###
summary(metadata$FED_PRAC_NEW)
26/64*100
16/64*100
22/64*100
#gross motor#
FED_PRAC_NEW_median<-aggregate(metadata$asq_9_total_grossmotor~metadat
a$FED_PRAC_NEW,FUN = median)
```

```
FED_PRAC_NEW_median
FED_PRAC_NEW_min<-aggregate(metadata$asq_9_total_grossmotor~metadata$F
ED_PRAC_NEW,FUN = min)
FED_PRAC_NEW_min
FED_PRAC_NEW_max<-aggregate(metadata$asq_9_total_grossmotor~metadata$F
ED_PRAC_NEW,FUN = max)
FED_PRAC_NEW_max
kruskal.test(asq_9_total_grossmotor~FED_PRAC_NEW, data=metadata)

#Fine motor
FED_PRAC_NEW_median<-aggregate(metadata$asq_9_total_finemotor~metadata
$FED_PRAC_NEW,FUN = median)
FED_PRAC_NEW_median
FED_PRAC_NEW_min<-aggregate(metadata$asq_9_total_finemotor~metadata$FE
D_PRAC_NEW,FUN = min)
FED_PRAC_NEW_min
FED_PRAC_NEW_max<-aggregate(metadata$asq_9_total_finemotor~metadata$FE
D_PRAC_NEW,FUN = max)
FED_PRAC_NEW_max
kruskal.test(asq_9_total_finemotor~FED_PRAC_NEW, data=metadata)

#Communication
FED_PRAC_NEW_median<-aggregate(metadata$asq_9_total_communication.tota
l.~metadata$FED_PRAC_NEW,FUN = median)
FED_PRAC_NEW_median
FED_PRAC_NEW_min<-aggregate(metadata$asq_9_total_communication.total.~
metadata$FED_PRAC_NEW,FUN = min)
FED_PRAC_NEW_min
FED_PRAC_NEW_max<-aggregate(metadata$asq_9_total_communication.total.~
metadata$FED_PRAC_NEW,FUN = max)
FED_PRAC_NEW_max
kruskal.test(asq_9_total_communication.total.~FED_PRAC_NEW, data=metad
ata)

#Personal and social
FED_PRAC_NEW_median<-aggregate(metadata$asq_9_total_personal_social~me
tadata$FED_PRAC_NEW,FUN = median)
FED_PRAC_NEW_median
FED_PRAC_NEW_min<-aggregate(metadata$asq_9_total_personal_social~metad
ata$FED_PRAC_NEW,FUN = min)
FED_PRAC_NEW_min
FED_PRAC_NEW_max<-aggregate(metadata$asq_9_total_personal_social~metad
ata$FED_PRAC_NEW,FUN = max)
FED_PRAC_NEW_max
kruskal.test(asq_9_total_personal_social~FED_PRAC_NEW, data=metadata)
```

```
#Problem solving
FED_PRAC_NEW_median<-aggregate(metadata$asq_9_total_problemsolving~met
adata$FED_PRAC_NEW,FUN = median)
FED_PRAC_NEW_median
FED_PRAC_NEW_min<-aggregate(metadata$asq_9_total_problemsolving~metada
ta$FED_PRAC_NEW,FUN = min)
FED_PRAC_NEW_min
FED_PRAC_NEW_max<-aggregate(metadata$asq_9_total_problemsolving~metada
ta$FED_PRAC_NEW,FUN = max)
FED_PRAC_NEW_max
kruskal.test(asq_9_total_problemsolving~FED_PRAC_NEW, data=metadata)


###FED_PRAC_LIGHT_NEW###
summary(metadata$FED_PRAC_LIGHT_NEW)
9/64*100
17/64*100
16/64*100
22/64*100
#Gross motor
FED_PRAC_LIGHT_NEW_median<-aggregate(metadata$asq_9_total_grossmotor~m
etadata$FED_PRAC_LIGHT_NEW,FUN = median)
FED_PRAC_LIGHT_NEW_median
FED_PRAC_LIGHT_NEW_min<-aggregate(metadata$asq_9_total_grossmotor~meta
data$FED_PRAC_LIGHT_NEW,FUN = min)
FED_PRAC_LIGHT_NEW_min
FED_PRAC_LIGHT_NEW_max<-aggregate(metadata$asq_9_total_grossmotor~meta
data$FED_PRAC_LIGHT_NEW,FUN = max)
FED_PRAC_LIGHT_NEW_max
kruskal.test(asq_9_total_grossmotor~FED_PRAC_LIGHT_NEW,data=metadata)

#Fine motor
FED_PRAC_LIGHT_NEW_median<-aggregate(metadata$asq_9_total_finemotor~me
tadata$FED_PRAC_LIGHT_NEW,FUN = median)
FED_PRAC_LIGHT_NEW_median
FED_PRAC_LIGHT_NEW_min<-aggregate(metadata$asq_9_total_finemotor~metad
ata$FED_PRAC_LIGHT_NEW,FUN = min)
FED_PRAC_LIGHT_NEW_min
FED_PRAC_LIGHT_NEW_max<-aggregate(metadata$asq_9_total_finemotor~metad
ata$FED_PRAC_LIGHT_NEW,FUN = max)
FED_PRAC_LIGHT_NEW_max
kruskal.test(asq_9_total_finemotor~FED_PRAC_LIGHT_NEW, data=metadata)

dunn.test(metadata$asq_9_total_finemotor,metadata$FED_PRAC_LIGHT_NEW,a
ltp = TRUE, method="bh")
```

```
#Communication
FED_PRAC_LIGHT_NEW_median<-aggregate(metadata$asq_9_total_communicatio
n.total.~metadata$FED_PRAC_LIGHT_NEW,FUN = median)
FED_PRAC_LIGHT_NEW_median
FED_PRAC_LIGHT_NEW_min<-aggregate(metadata$asq_9_total_communication.t
otal.~metadata$FED_PRAC_LIGHT_NEW,FUN = min)
FED_PRAC_LIGHT_NEW_min
FED_PRAC_LIGHT_NEW_max<-aggregate(metadata$asq_9_total_communication.t
otal.~metadata$FED_PRAC_LIGHT_NEW,FUN = max)
FED_PRAC_LIGHT_NEW_max
kruskal.test(asq_9_total_communication.total.~FED_PRAC_LIGHT_NEW, data
=metadata)

#Personal and social
FED_PRAC_LIGHT_NEW_median<-aggregate(metadata$asq_9_total_personal_soc
ial~metadata$FED_PRAC_LIGHT_NEW,FUN = median)
FED_PRAC_LIGHT_NEW_median
FED_PRAC_LIGHT_NEW_min<-aggregate(metadata$asq_9_total_personal_social
~metadata$FED_PRAC_LIGHT_NEW,FUN = min)
FED_PRAC_LIGHT_NEW_min
FED_PRAC_LIGHT_NEW_max<-aggregate(metadata$asq_9_total_personal_social
~metadata$FED_PRAC_LIGHT_NEW,FUN = max)
FED_PRAC_LIGHT_NEW_max
kruskal.test(asq_9_total_personal_social~FED_PRAC_LIGHT_NEW, data=meta
data)

#Problem solving
FED_PRAC_LIGHT_NEW_median<-aggregate(metadata$asq_9_total_problemsolvi
ng~metadata$FED_PRAC_LIGHT_NEW,FUN = median)
FED_PRAC_LIGHT_NEW_median
FED_PRAC_LIGHT_NEW_min<-aggregate(metadata$asq_9_total_problemsolving~
metadata$FED_PRAC_LIGHT_NEW,FUN = min)
FED_PRAC_LIGHT_NEW_min
FED_PRAC_LIGHT_NEW_max<-aggregate(metadata$asq_9_total_problemsolving~
metadata$FED_PRAC_LIGHT_NEW,FUN = max)
FED_PRAC_LIGHT_NEW_max
kruskal.test(asq_9_total_problemsolving~FED_PRAC_LIGHT_NEW, data=metad
ata)
```

Table 6. Associations between feeding methods in the 24 hours prior to stool sample collection at 3 months and infant ASQ scales at 9 months of age

```
# Gross motor
grossmotor_uni<-lm(asq_9_total_grossmotor~FED_PRAC_NEW,data=metadata)
summary(grossmotor_uni)
confint(grossmotor_uni)
```

```
grossmotor<-lm(asq_9_total_grossmotor~FED_PRAC_NEW+antibiotics_since_b
irth+SEX+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+EDU_LVL+PRE_BMI+maternal_a
ge,data=metadata)
summary(grossmotor)
confint(grossmotor)

#Fine motor
finemotor_uni<-lm(asq_9_total_finemotor~FED_PRAC_NEW, data=metadata)
summary(finemotor_uni)
confint(finemotor_uni)

finemotor<-lm(asq_9_total_finemotor~FED_PRAC_NEW+antibiotics_since_bir
th+SEX+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+EDU_LVL+PRE_BMI+maternal_age
, data=metadata)
summary(finemotor)
confint(finemotor)

#Communication
Communication_uni<-lm(asq_9_total_communication.total.~FED_PRAC_NEW,da
ta=metadata)
summary(Communication_uni)
confint(Communication_uni)

Communication<-lm(asq_9_total_communication.total.~FED_PRAC_NEW+antibi
otics_since_birth+SEX+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+EDU_LVL+PRE_B
MI+maternal_age,data=metadata)
summary(Communication)
confint(Communication)

#Personal and social
personal_uni<-lm(asq_9_total_personal_social~FED_PRAC_NEW,data=metadat
a)
summary(personal_uni)
confint(personal_uni)

personal<-lm(asq_9_total_personal_social~FED_PRAC_NEW+antibiotics_sinc
e_birth+SEX+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+EDU_LVL+PRE_BMI+materna
l_age,data=metadata)
summary(personal)
confint(personal)

#Problem solving
problem_uni<-lm(asq_9_total_problemsolving~FED_PRAC_NEW,data=metadata)

summary(problem_uni)
confint(problem_uni)
```

```
problem<-lm(asq_9_total_problemsolving~FED_PRAC_NEW+antibiotics_since_
birth+SEX+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+EDU_LVL+PRE_BMI+maternal_
age,data=metadata)
summary(problem)
confint(problem)
```

Table 7. Associations between infant feeding in the 24 hours prior to stool sample collection and population characteristics

```
###Baby Sex###
summary(metadata)
31/64*100
33/64*100
#Male
male_breast<-filter(metadata,SEX=="1" & FED_PRAC_LIGHT_NEW=="1") #n=4
count(male_breast)
male_breast_D<-filter(metadata,SEX=="1" & FED_PRAC_LIGHT_NEW=="2") #n=
6
count(male_breast_D)
male_mix<-filter(metadata,SEX=="1" & FED_PRAC_LIGHT_NEW=="3") #n=10
count(male_mix)
male_formula<-filter(metadata,SEX=="1" & FED_PRAC_LIGHT_NEW=="4") #n=1
1
count(male_formula)
4/9*100
6/17*100
10/16*100
11/22*100
#Female
gross_breast<-filter(metadata,SEX=="2" & FED_PRAC_LIGHT_NEW=="1") #n=5
count(gross_breast)
gross_breast_D<-filter(metadata,SEX=="2" & FED_PRAC_LIGHT_NEW=="2") #n
=11
count(gross_breast_D)
gross_mix<-filter(metadata,SEX=="2" & FED_PRAC_LIGHT_NEW=="3") #n=6
count(gross_mix)
gross_formula<-filter(metadata,SEX=="2" & FED_PRAC_LIGHT_NEW=="4") #n=
11
count(gross_formula)
chisq.test(table(metadata$SEX,metadata$FED_PRAC_LIGHT_NEW),simulate.p.
value = TRUE)
5/9*100
11/17*100
6/16*100
11/22*100

###Baby Race###
```

```
44/64*100
20/64*100
#white#
white_breast<-filter(metadata,Race_new=="1" & FED_PRAC_LIGHT_NEW=="1")
 #n=8
count(white_breast)
white_breast_D<-filter(metadata,Race_new=="1" & FED_PRAC_LIGHT_NEW=="2
") #n=12
count(white_breast_D)
white_mix<-filter(metadata,Race_new=="1" & FED_PRAC_LIGHT_NEW=="3") #n
=11
count(white_mix)
white_formula<-filter(metadata,Race_new=="1" & FED_PRAC_LIGHT_NEW=="4"
) #n=13
count(white_formula)
8/9*100
12/17*100
11/16*100
13/22*100
#non white#
nowhite_breast<-filter(metadata,Race_new=="2" & FED_PRAC_LIGHT_NEW=="1
") #n=1
count(nowhite_breast)
nowhite_breast_D<-filter(metadata,Race_new=="2" & FED_PRAC_LIGHT_NEW==
"2") #n=5
count(nowhite_breast_D)
nowhite_mix<-filter(metadata,Race_new=="2" & FED_PRAC_LIGHT_NEW=="3")
#n=5
count(nowhite_mix)
nowhite_formula<-filter(metadata,Race_new=="2" & FED_PRAC_LIGHT_NEW=="
4") #n=9
count(nowhite_formula)
chisq.test(table(metadata$Race_new,metadata$FED_PRAC_LIGHT_NEW),simula
te.p.value = TRUE)
1/9*100
5/17*100
5/16*100
9/22*100

###materanl education level###
3/64*100
11/64*100
13/64*100
37/64*100
# non-high school
nohigh_breast<-filter(metadata,EDU_LVL=="1" & FED_PRAC_LIGHT_NEW=="1")
```

```
#n=0
count(nohigh_breast)
nohigh_breast_D<-filter(metadata,EDU_LVL=="1" & FED_PRAC_LIGHT_NEW=="2
") #n=0
count(nohigh_breast_D)
nohigh_mix<-filter(metadata,EDU_LVL=="1" & FED_PRAC_LIGHT_NEW=="3") #n
=0
count(nohigh_mix)
nohigh_formula<-filter(metadata,EDU_LVL=="1" & FED_PRAC_LIGHT_NEW=="4"
) #n=3
count(nohigh_formula)
3/22*100
#high school
high_breast<-filter(metadata,EDU_LVL=="2" & FED_PRAC_LIGHT_NEW=="1") #
n=1
count(high_breast)
high_breast_D<-filter(metadata,EDU_LVL=="2" & FED_PRAC_LIGHT_NEW=="2")
 #n=0
count(high_breast_D)
high_mix<-filter(metadata,EDU_LVL=="2" & FED_PRAC_LIGHT_NEW=="3") #n=3
count(high_mix)
high_formula<-filter(metadata,EDU_LVL=="2" & FED_PRAC_LIGHT_NEW=="4")
#n=7
count(high_formula)
1/9*100
3/16*100
7/22*100
#some college
socoll_breast<-filter(metadata,EDU_LVL=="3" & FED_PRAC_LIGHT_NEW=="1")
 #n=2
count(socoll_breast)
socoll_breast_D<-filter(metadata,EDU_LVL=="3" & FED_PRAC_LIGHT_NEW=="2
") #n=3
count(socoll_breast_D)
socoll_mix<-filter(metadata,EDU_LVL=="3" & FED_PRAC_LIGHT_NEW=="3") #n
=3
count(socoll_mix)
socoll_formula<-filter(metadata,EDU_LVL=="3" & FED_PRAC_LIGHT_NEW=="4"
) #n=5
count(socoll_formula)
2/9*100
3/17*100
3/16*100
5/22*100
#college
coll_breast<-filter(metadata,EDU_LVL=="4" & FED_PRAC_LIGHT_NEW=="1") #
```

```
n=6
count(coll_breast)
coll_breast_D<-filter(metadata,EDU_LVL=="4" & FED_PRAC_LIGHT_NEW=="2")
 #n=14
count(coll_breast_D)
coll_mix<-filter(metadata,EDU_LVL=="4" & FED_PRAC_LIGHT_NEW=="3") #n=1
0
count(coll_mix)
coll_formula<-filter(metadata,EDU_LVL=="4" & FED_PRAC_LIGHT_NEW=="4")
#n=7
count(coll_formula)
chisq.test(table(metadata$EDU_LVL,metadata$FED_PRAC_LIGHT_NEW),simulat
e.p.value = TRUE)
6/9*100
14/17*100
10/16*100
7/22*100

#Delivery mode
39/64*100
25/64*100
#vaginal
vag_breast<-filter(metadata,MD_FINAL_ROUTE=="1" & FED_PRAC_LIGHT_NEW==
"1") #n=5
count(vag_breast)
vag_breast_D<-filter(metadata,MD_FINAL_ROUTE=="1" & FED_PRAC_LIGHT_NEW
=="2") #n=10
count(vag_breast_D)
vag_mix<-filter(metadata,MD_FINAL_ROUTE=="1" & FED_PRAC_LIGHT_NEW=="3"
) #n=15
count(vag_mix)
vag_formula<-filter(metadata,MD_FINAL_ROUTE=="1" & FED_PRAC_LIGHT_NEW=
="4") #n=9
count(vag_formula)
5/9*100
10/17*100
15/16*100
9/22*100
#c section
C_breast<-filter(metadata,MD_FINAL_ROUTE=="2" & FED_PRAC_LIGHT_NEW=="1
") #n=4
count(C_breast)
C_breast_D<-filter(metadata,MD_FINAL_ROUTE=="2" & FED_PRAC_LIGHT_NEW==
"2") #n=7
count(C_breast_D)
C_mix<-filter(metadata,MD_FINAL_ROUTE=="2" & FED_PRAC_LIGHT_NEW=="3")
```

```
#n=1
count(C_mix)
C_formula<-filter(metadata,MD_FINAL_ROUTE=="2" & FED_PRAC_LIGHT_NEW=="
4") #n=13
count(C_formula)
chisq.test(table(metadata$MD_FINAL_ROUTE,metadata$FED_PRAC_LIGHT_NEW),
simulate.p.value = TRUE)
4/9*100
7/17*100
1/16*100
13/22*100

###continuous variables###
mean(metadata$PRE_BMI)
sd(metadata$PRE_BMI)
#pre_bmi
bmi_breast<-filter(metadata, FED_PRAC_LIGHT_NEW=="1")
mean(bmi_breast$PRE_BMI)
sd(bmi_breast$PRE_BMI)
bmi_breast_D<-filter(metadata, FED_PRAC_LIGHT_NEW=="2")
mean(bmi_breast_D$PRE_BMI)
sd(bmi_breast_D$PRE_BMI)
bmi_mix<-filter(metadata, FED_PRAC_LIGHT_NEW=="3")
mean(bmi_mix$PRE_BMI)
sd(bmi_mix$PRE_BMI)
bmi_formula<-filter(metadata, FED_PRAC_LIGHT_NEW=="4")
mean(bmi_formula$PRE_BMI)
sd(bmi_formula$PRE_BMI)
kruskal.test(PRE_BMI~FED_PRAC_LIGHT_NEW, data=metadata)

#maternal age
mean(metadata$maternal_age)
sd(metadata$maternal_age)

age_breast<-filter(metadata, FED_PRAC_LIGHT_NEW=="1")
mean(age_breast$maternal_age)
sd(age_breast$maternal_age)
age_breast_D<-filter(metadata, FED_PRAC_LIGHT_NEW=="2")
mean(age_breast_D$maternal_age)
sd(age_breast_D$maternal_age)
age_mix<-filter(metadata, FED_PRAC_LIGHT_NEW=="3")
mean(age_mix$maternal_age)
sd(age_mix$maternal_age)
age_formula<-filter(metadata, FED_PRAC_LIGHT_NEW=="4")
mean(age_formula$maternal_age)
sd(age_formula$maternal_age)
```

```
kruskal.test(maternal_age~FED_PRAC_LIGHT_NEW, data=metadata)

#gestational age
mean(metadata$ESTWKSGEST)
sd(metadata$ESTWKSGEST)

gest_breast<-filter(metadata, FED_PRAC_LIGHT_NEW=="1")
mean(gest_breast$ESTWKSGEST)
sd(gest_breast$ESTWKSGEST)
gest_breast_D<-filter(metadata, FED_PRAC_LIGHT_NEW=="2")
mean(gest_breast_D$ESTWKSGEST)
sd(gest_breast_D$ESTWKSGEST)
gest_mix<-filter(metadata, FED_PRAC_LIGHT_NEW=="3")
mean(gest_mix$ESTWKSGEST)
sd(gest_mix$ESTWKSGEST)
gest_formula<-filter(metadata, FED_PRAC_LIGHT_NEW=="4")
mean(gest_formula$ESTWKSGEST)
sd(gest_formula$ESTWKSGEST)

kruskal.test(ESTWKSGEST~FED_PRAC_LIGHT_NEW, data=metadata)
```

Table 8. Associations between feeding methods after stratification by vitamin D supplementation in the 24 hours prior to stool sample collection at 3 months of age and infant ASQ scales at 9 months of age

```
# gross motor
grossmotor_uni1<-lm(asq_9_total_grossmotor~FED_PRAC_LIGHT_NEW,data=met
adata)
summary(grossmotor_uni1)
confint(grossmotor_uni1)

grossmotor<-lm(asq_9_total_grossmotor~FED_PRAC_LIGHT_NEW+antibiotics_s
ince_birth+SEX+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+EDU_LVL+PRE_BMI+mate
rnal_age,data=metadata)
summary(grossmotor)
confint(grossmotor)

#Fine motor
finemotor_uni1<-lm(asq_9_total_finemotor~FED_PRAC_LIGHT_NEW, data=meta
data)
summary(finemotor_uni1)
confint(finemotor_uni1)

finemotor<-lm(asq_9_total_finemotor~FED_PRAC_LIGHT_NEW+antibiotics_sin
ce_birth+SEX+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+EDU_LVL+PRE_BMI+matern
al_age, data=metadata)
summary(finemotor)
```

```
confint(finemotor)

#Communication
Communication_uni1<-lm(asq_9_total_communication.total.~FED_PRAC_LIGHT
_NEW,data=metadata)
summary(Communication_uni1)
confint(Communication_uni1)

Communication<-lm(asq_9_total_communication.total.~FED_PRAC_LIGHT_NEW+
antibiotics_since_birth+SEX+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+EDU_LVL
+PRE_BMI+maternal_age,data=metadata)
summary(Communication)
confint(Communication)

#Personal and social
personal_uni1<-lm(asq_9_total_personal_social~ FED_PRAC_LIGHT_NEW,data
=metadata)
summary(personal_uni1)
confint(personal_uni1)

personal<-lm(asq_9_total_personal_social~FED_PRAC_LIGHT_NEW+antibiotic
s_since_birth+SEX+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+EDU_LVL+PRE_BMI+m
aternal_age,data=metadata)
summary(personal)
confint(personal)

#Problem solving
problem_uni1<-lm(asq_9_total_problemsolving~FED_PRAC_LIGHT_NEW,data=me
tadata)
summary(problem_uni1)
confint(problem_uni1)

problem<-lm(asq_9_total_problemsolving~FED_PRAC_LIGHT_NEW+antibiotics_
since_birth+SEX+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+EDU_LVL+PRE_BMI+mat
ernal_age,data=metadata)
summary(problem)
confint(problem)
```

Table 9. Associations between exclusive breastfeeding duration and infant ASQ scales at 9 months of age

```
options(scipen = 100)
summary(metadata)
#Gross motor
grossmotor_uni2<-lm(asq_9_total_grossmotor~exclusive_feeding_duration_
updated,data=metadata)
summary(grossmotor_uni2)
```

```
confint(grossmotor_uni2)

grossmotor<-lm(asq_9_total_grossmotor~exclusive_feeding_duration_updat
ed+SEX+antibiotics_since_birth+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+EDU_
LVL+PRE_BMI+maternal_age ,data=metadata)
summary(grossmotor)
confint(grossmotor,"exclusive_feeding_duration_updated")

#Fine motor
finemotor_uni2<-lm(asq_9_total_finemotor~exclusive_feeding_duration_up
dated,data=metadata)
summary(finemotor_uni2)
confint(finemotor_uni2)

finemotor<-lm(asq_9_total_finemotor~exclusive_feeding_duration_updated
+SEX+antibiotics_since_birth+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+EDU_LV
L+PRE_BMI+maternal_age,data=metadata)
summary(finemotor )
confint(finemotor,"exclusive_feeding_duration_updated")

#Communication
Communication_uni2<-lm(asq_9_total_communication.total.~exclusive_feed
ing_duration_updated,data=metadata)
summary(Communication_uni2)
confint(Communication_uni2)

Communication<-lm(asq_9_total_communication.total.~exclusive_feeding_d
uration_updated+SEX+antibiotics_since_birth+ESTWKSGEST+MD_FINAL_ROUTE+
Race_new+EDU_LVL+PRE_BMI+maternal_age,data=metadata)
summary(Communication)
confint(Communication,"exclusive_feeding_duration_updated")

#Personal and social
personal_uni2<-lm(asq_9_total_personal_social~exclusive_feeding_durati
on_updated,data=metadata)
summary(personal_uni2)
confint(personal_uni2)

personal<-lm(asq_9_total_personal_social~exclusive_feeding_duration_up
dated+SEX+antibiotics_since_birth+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+E
DU_LVL+PRE_BMI+maternal_age,data=metadata)
summary(personal)
confint(personal,"exclusive_feeding_duration_updated")

#Problem solving
problem_uni2<-lm(asq_9_total_problemsolving~exclusive_feeding_duration
```

```
_updated,data=metadata)
summary(problem_uni2)
confint(problem_uni2)

problem<-lm(asq_9_total_problemsolving~exclusive_feeding_duration_upda
ted+SEX+antibiotics_since_birth+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+EDU
_LVL+PRE_BMI+maternal_age,data=metadata)
summary(problem)
confint(problem,"exclusive_feeding_duration_updated")
```

Table 10. Associations between any breastfeeding duration and infant ASQ scales at 9 months of age

```
options(scipen = 100)
summary(metadata)
#Gross motor
grossmotor_uni2<-lm(asq_9_total_grossmotor~ mix_breastfeeding_duration
_updated,data=metadata)
summary(grossmotor_uni2)
confint(grossmotor_uni2)

grossmotor<-lm(asq_9_total_grossmotor~ mix_breastfeeding_duration_upda
ted +SEX+antibiotics_since_birth+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+ED
U_LVL+PRE_BMI+maternal_age,data=metadata)
summary(grossmotor)
confint(grossmotor,"mix_breastfeeding_duration_updated")

#Fine motor
finemotor_uni2<-lm(asq_9_total_finemotor~ mix_breastfeeding_duration_u
pdated,data=metadata)
summary(finemotor_uni2)
confint(finemotor_uni2)

finemotor<-lm(asq_9_total_finemotor~ mix_breastfeeding_duration_update
d +SEX+antibiotics_since_birth+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+EDU_
LVL+PRE_BMI+maternal_age,data=metadata)
summary(finemotor)
confint(finemotor,"mix_breastfeeding_duration_updated")

#Communication
Communication_uni2<-lm(asq_9_total_communication.total.~ mix_breastfee
ding_duration_updated,data=metadata)
summary(Communication_uni2)
confint(Communication_uni2)

Communication<-lm(asq_9_total_communication.total.~ mix_breastfeeding_
duration_updated +SEX+antibiotics_since_birth+ESTWKSGEST+MD_FINAL_ROUT
```

```
E+Race_new+EDU_LVL+PRE_BMI+maternal_age,data=metadata)
summary(Communication)
confint(Communication,"mix_breastfeeding_duration_updated")

#Personal and social
personal_uni2<-lm(asq_9_total_personal_social~ mix_breastfeeding_durat
ion_updated,data=metadata)
summary(personal_uni2)
confint(personal_uni2)

personal<-lm(asq_9_total_personal_social~ mix_breastfeeding_duration_u
pdated +SEX+antibiotics_since_birth+ESTWKSGEST+MD_FINAL_ROUTE+Race_new
+EDU_LVL+PRE_BMI+maternal_age,data=metadata)
summary(personal)
confint(personal,"mix_breastfeeding_duration_updated")

#Problem solving
problem_uni2<-lm(asq_9_total_problemsolving~ mix_breastfeeding_duratio
n_updated,data=metadata)
summary(problem_uni2)
confint(problem_uni2)

problem<-lm(asq_9_total_problemsolving~ mix_breastfeeding_duration_upd
ated +SEX+antibiotics_since_birth+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+E
DU_LVL+PRE_BMI+maternal_age,data=metadata)
summary(problem)
confint(problem,"mix_breastfeeding_duration_updated")
```

Figure 11. Associations between infant feeding method in the 24 hours prior to stool sample collection and infant gut microbiota alpha diversity at 3 months of age

```
Data.Subsample.final.Alpha<-read.csv("/Users/busihan/Desktop/MARCH\ B3
m_ASQ_updated/Data.Subsample.final.Alpha_Final.csv", header = T)
shapiro.test(Data.Subsample.final.Alpha$Chao)   #p=0.3152
shapiro.test(Data.Subsample.final.Alpha$Shannon)  #p=0.5375
shapiro.test(Data.Subsample.final.Alpha$Invsimpson) #p=0.008523
chao<-Data.Subsample.final.Alpha$Chao
shan<-Data.Subsample.final.Alpha$Shannon
invismp<-Data.Subsample.final.Alpha$Invsimpson
a<-metadata$FED_PRAC_LIGHT_NEW

# Chao
summary(aov(chao~a))
TukeyHSD(aov(chao~a))

#Shannon
summary(aov(shan~a))
```

```
TukeyHSD(aov(shan~a))

#Inverse Simp
kruskal.test(invismp~a)
dunn.test(invismp,a,altp = TRUE, method="bh")

labels<-c("Breastmilk","Breastmilk_VitaminD","Partial breastmilk","For
mula")
png("Fed_PRAC_light_Alpha.png", res=300, height=5, width=13,units="in"
)
par(mfrow= c(1,3),mar=c(7, 5, 3, 1))
boxplot(chao~a,main="A.Feeding method_Chao1",ylab="Chao1 Index",xlab =
 NA,cex.lab=2,cex.main=2,cex.axis=2,xaxt = "n")
axis(side = 2, labels = FALSE)
text(x = 1:4,y = par("usr")[3]-3.3,labels =labels,xpd = NA,srt = 25,ce
x = 1.7,adj = 1)
text(x=2,y=75,labels= "p-value=0.04",cex=1.5)

boxplot(shan~a,main="B.Feeding method_Shannon",ylab="Shannon Index",xl
ab=NA,cex.lab=2,cex.main=2,cex.axis=2,xaxt = "n")
axis(side = 2, labels = FALSE)
text(x = 1:4,y = par("usr")[3] -0.1,labels =labels,xpd = NA,srt = 25,c
ex = 1.7,adj = 1)
text(x=1,y=2.09,labels= "a",cex=1.5)
text(x=2,y=2.285,labels= "a",cex=1.5)
text(x=3,y=2.41,labels= "b",cex=1.5)
text(x=4,y=1.49,labels= "c",cex=1.5)
text(x=2,y=2.8,labels= "p-value<0.01",cex=1.5)

boxplot(invismp~a,main="C.Feeding method_inverse Simpson",ylab="Invers
e Simpson Index",xlab = NA,cex.lab=2,cex.main=2,cex.axis=2,,xaxt = "n"
)
text(x=1,y=6.55,labels= "a",cex=1.5)
text(x=2,y=6.6,labels= "a",cex=1.5)
text(x=3,y=7.95,labels= "ab",cex=1.5)
text(x=4,y=10.9,labels= "b",cex=1.5)
axis(side = 1, labels = FALSE)
text(x = 1:4,y = par("usr")[3] - 0.55,labels =labels,xpd = NA,srt = 25
,cex = 1.7,adj = 1)
text(x=1.5,y=11,labels= "p-value<0.01",cex=1.5)
while (!is.null(dev.list()))  dev.off()
```

Figure 12. Associations between infant feeding methods in the 24 hours prior to stool sample collection and gut microbiota beta diversity at 3 months of age

```
a<-metadata$FED_PRAC_LIGHT_NEW
#Sorenson
```

```
PERMANOVA(Data.Subsample.genus_37wks[,-c(1:3,254)],a,TRUE,9999) #p =1e
-04
Sor<-Sor.bray.pcoa(Data.Subsample.genus_37wks[,-c(1:3,254)],Dim=2,Colo
r=metadata$FED_PRAC_LIGHT_NEW,binary=TRUE)

#Bray-Curtis
PERMANOVA(Data.Subsample.genus_37wks[,-c(1:3,254)],a,FALSE,9999)#p = 1
e-04
Bray<-Sor.bray.pcoa(Data.Subsample.genus_37wks[,-c(1:3,254)],Dim=2,Col
or=metadata$FED_PRAC_LIGHT_NEW,binary=FALSE)

Color<-ifelse(grepl("1", metadata$FED_PRAC_LIGHT_NEW),"#000000", ifels
e(grepl("2", metadata$FED_PRAC_LIGHT_NEW),"#E79F00",ifelse(grepl("3",
metadata$FED_PRAC_LIGHT_NEW),"#652DC1","#0072B2")))

png("Beta_diversity_Feeding.png", res=300, height=7, width=17,units="i
n")
par(mfrow= c(1,2),mar=c(7, 5, 3, 1))
plot(Sor,cex.axis=2,cex.lab=2,cex.main=3,cex=3,col=1,
     pch=21,xlim=c(-.38,.5),ylim=c(-.3,.4),xlab="PC1 (22.1%)",ylab="PC
2 (11.7%)",bg=Color,main="A. Sorensen")
ordiellipse(Sor,groups=metadata$FED_PRAC_LIGHT_NEW,col= c("#000000","#
E79F00","#652DC1","#0072B2"),lwd=2)
legend(0.15,0.41,c("Breastmilk","Breastmilk_vitaminD","Partial breastm
ilk","Formula"), pch=21,cex = 1.5,pt.bg=c("#000000","#E79F00","#652DC1
","#0072B2"),y.intersp = 0.72)
text(0.3,-0.25, labels= "p-value<0.01",cex=1.5)

plot(Bray,cex.axis=2,cex.lab=2,cex.main=3,cex=3,col=1,
     pch=21,xlim=c(-.5,.55),ylim=c(-.35,.58),xlab="PC1 (25.1%)",ylab="
PC2 (13.6%)",bg=Color,main="B. Bray-Curtis")
ordiellipse(Bray,groups=metadata$FED_PRAC_LIGHT_NEW,col= c("#000000","
#E79F00","#652DC1","#0072B2"),lwd=2)
legend(0.1,0.55,c("Breastmilk","Breastmilk_vitaminD","Partial breastmi
lk","Formula"), pch=21,cex =1.5,pt.bg=c("#000000","#E79F00","#652DC1",
"#0072B2"),y.intersp = 0.72)
text(0.28,-0.3, labels= "p-value<0.01",cex=1.5)
while (!is.null(dev.list()))  dev.off()
```

Table 11. Mediation effect of the inverse Simpson index on the associations of feeding method with communication

```
#Exposure : feeding practice light
#Mediator: Inverse Simpson
#Outcome: communication
Data.Subsample.final.Alpha<-read.csv("/Users/busihan/Desktop/MARCH\ B3
m_ASQ_updated/Data.Subsample.final.Alpha_Final.csv", header = T)
metadata<-merge(metadata,Data.Subsample.final.Alpha,by.x="Group", by.y
```

```
 = "Names")

saved = mediation1(y = "asq_9_total_communication.total.",
                   x = "FED_PRAC_LIGHT_NEW",
                   m = "Invsimpson",
                   cvs = c("SEX","MD_FINAL_ROUTE","Race_new","EDU_LVL"
,"ESTWKSGEST","PRE_BMI","maternal_age"),
                   df = metadata,
                   with_out = T,
                   nboot = 1000,
                   conf_level = .95)

####view the analysis####
summary(saved$model1) #c path,total effect
summary(saved$model2) #a path,
summary(saved$model3) #b and c' path

# total effect
summary(saved$model1)
#double check
Communication<-lm(asq_9_total_communication.total.~FED_PRAC_LIGHT_NEW+
SEX+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+EDU_LVL+PRE_BMI+maternal_age,da
ta=metadata)
summary(Communication)
confint(Communication)

# direct effect, # X predicts Y with M as the exposure not outcome
summary(saved$model3) #b and c' path
#double check
Communication<-lm(asq_9_total_communication.total.~FED_PRAC_LIGHT_NEW+
Invsimpson+SEX+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+EDU_LVL+PRE_BMI+mate
rnal_age,data=metadata)
summary(Communication)
confint(Communication)

#total, direct, indirect effects (estimates)
saved$total.effect; saved$direct.effect; saved$indirect.effect

#Sobel test to test the significance of indirect effects(p-value)
saved$z.score; saved$p.value

#bootstrapped indirect effect (95%CI)
saved$boot.results
saved$boot.ci
```

Table 12. Mediation effect of inverse Simpson on the associations of feeding method with

problem-solving

```
#Exposure : feeding practice light
#Mediator: Inverse simpson
#Outcome: problem solving
saved = mediation1(y = "asq_9_total_problemsolving",
                   x = "FED_PRAC_LIGHT_NEW",
                   m = "Invsimpson",
                   cvs = c("SEX","MD_FINAL_ROUTE","Race_new","EDU_LVL"
,"ESTWKSGEST","PRE_BMI","maternal_age"),
                   df = metadata,
                   with_out = T,
                   nboot = 1000,
                   conf_level = .95)

####view the analysis####
summary(saved$model1) #c path,total effect
summary(saved$model2) #a path,
summary(saved$model3) #b and c' path

# total effect
summary(saved$model1)
problem<-lm(asq_9_total_problemsolving~FED_PRAC_LIGHT_NEW+SEX+ESTWKSGE
ST+MD_FINAL_ROUTE+Race_new+EDU_LVL+PRE_BMI+maternal_age,data=metadata)

summary(problem)
confint(problem)

# direct effect, # X predicts Y with M as the exposure not outcome
summary(saved$model3) #b and c' path
problem<-lm(asq_9_total_problemsolving~FED_PRAC_LIGHT_NEW+Invsimpson+S
EX+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+EDU_LVL+PRE_BMI+maternal_age,dat
a=metadata)
summary(problem)
confint(problem)

#total, direct, indirect effects(estimates)
saved$total.effect; saved$direct.effect; saved$indirect.effect

#Sobel test to test the significance of indirect effects(p-value)
saved$z.score; saved$p.value

#bootstrapped indirect (95%CI)
saved$boot.results
saved$boot.ci
```

Table 13. Mediation effect of the Shannon index on the associations of feeding method with

problem-solving

```
#Exposure : feeding practice light
#Mediator: Shannon
#Outcome: problem solving
saved = mediation1(y = "asq_9_total_problemsolving",
                   x = "FED_PRAC_LIGHT_NEW",
                   m = "Shannon",
                   cvs = c("SEX","MD_FINAL_ROUTE","Race_new","EDU_LVL"
,"ESTWKSGEST","PRE_BMI","maternal_age"),
                   df = metadata,
                   with_out = T,
                   nboot = 1000,
                   conf_level = .95)

####view the analysis####
 ####view the analysis####
summary(saved$model1) #c path,total effect
summary(saved$model2) #a path,
summary(saved$model3) #b and c' path

# total effect
summary(saved$model1)
problem<-lm(asq_9_total_problemsolving~FED_PRAC_LIGHT_NEW+SEX+ESTWKSGE
ST+MD_FINAL_ROUTE+Race_new+EDU_LVL+PRE_BMI+maternal_age,data=metadata)

summary(problem)
confint(problem)

# direct effect, # X predicts Y with M as the exposure not outcome
summary(saved$model3) #b and c' path
problem<-lm(asq_9_total_problemsolving~FED_PRAC_LIGHT_NEW+Shannon+SEX+
ESTWKSGEST+MD_FINAL_ROUTE+Race_new+EDU_LVL+PRE_BMI+maternal_age,data=m
etadata)
summary(problem)
confint(problem)

#total, direct, indirect effects (estimates)
saved$total.effect; saved$direct.effect; saved$indirect.effect

#Sobel test to test the significance of indirect effects(p-value)
saved$z.score; saved$p.value

#bootstrapped indirect effect(95%CI)
saved$boot.results
saved$boot.ci
```

Table 14. Mediation effect of the Bray-Curtis dissimilarity matrix on the associations of feeding method with ASQ scales

```
otu_table<-Data.Subsample.genus_37wks[,-c(1:3,254)]

#Exposure : feeding practice light
#Mediator: Bray-Curtis
#Outcome: communication

#univariate
med_uni <- permanovaFL(otu_table ~ FED_PRAC_LIGHT_NEW + asq_9_total_co
mmunication.total.,data=metadata, seed=82955, n.cores=4,test.mediation
=TRUE,dist.method="bray",square.dist=TRUE)
med_uni$med.p.permanova   #p=0.1552

#multivariate analysis
med_multi<- permanovaFL(otu_table|(SEX+antibiotics_since_birth+EDU_LVL
+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+PRE_BMI+maternal_age)~FED_PRAC_LIG
HT_NEW + asq_9_total_communication.total.,data=metadata, seed=82955, n
.cores=4,test.mediation=TRUE, dist.method="bray",
square.dist=TRUE)
med_multi$med.p.permanova   #p=0.5476

#Exposure : feeding practice light
#Mediator: Bray-Curtis
#Outcome: fine motor

#univariate
med_uni <- permanovaFL(otu_table ~ FED_PRAC_LIGHT_NEW + asq_9_total_fi
nemotor,data=metadata, seed=82955, n.cores=4,test.mediation=TRUE,dist.
method="bray",square.dist=TRUE)
med_uni$med.p.permanova   #p=0.037

#multivariate analysis
med_multi<- permanovaFL(otu_table|(SEX+antibiotics_since_birth+EDU_LVL
+ESTWKSGEST+MD_FINAL_ROUTE+Race_new+PRE_BMI+maternal_age)~FED_PRAC_LIG
HT_NEW + asq_9_total_finemotor,data=metadata, seed=82955, n.cores=4,te
st.mediation=TRUE,dist.method="bray",square.dist=TRUE)
med_multi$med.p.permanova   #p=0.283
```

Chapter 4

Data preparation

```
require(vegan)
require(lubridate)
require(tidyr)
require(MASS)
require(car)
require(dunn.test)
require(ggplot2)
require(ggpubr)
require(dplyr)
require(pBrackets)
require(grid)
require(Maaslin2)
require(pairwiseAdonis)

setwd("/Users/busihan/Desktop/2023Mar27_Aim3_double_check/")

Alpha<-function(OTU,Names="Sample",Groups="Sample"){
  Chao<-t(estimateR(OTU))
  Chao<-Chao[,2]
  Shannon<-diversity(OTU,index="shannon")
  Invsimpson<-diversity(OTU,index="invsimpson")
  OTU.Subsample.Alpha<-data.frame(Names,Groups,Chao,Shannon,Invsimpson
)
  return(OTU.Subsample.Alpha)
}

Sor.bray.pcoa<-function(OTUS,Dim=2,Color=1,binary,pch=16,Title="PCoA")
{
  Data.df<-vegdist(OTUS,method="bray", binary)
  Data.df.PCoA<-cmdscale(Data.df, k = Dim, eig = FALSE)
  Data.df.PCoA.eig<-cmdscale(Data.df, k = Dim, eig = TRUE)
  eig.Data.df.PCoA<-Data.df.PCoA.eig$eig
  eig.Data.df.PCoA.sum<-sum(eig.Data.df.PCoA)
  a<-(eig.Data.df.PCoA/eig.Data.df.PCoA.sum)*100
  xlab<-paste("PC1","(",round(a[1],1),"%",")",sep="")
  ylab<-paste("PC2","(",round(a[2],1),"%",")",sep="")
  if(binary==TRUE){
    main<-"Sorensen PCoA"
  }else(main<-"Bray-Curtis PCoA")
  plot(Data.df.PCoA, col=Color,
       main=Title,xlab=xlab,ylab=ylab,pch=c(pch))
  return(Data.df.PCoA)
}
```

```
PERMANOVA<-function(OTUS,Group,binary,iters=9999){
  Data.Dist<-vegdist(OTUS,method="bray", binary=binary)
  adonis2(Data.Dist~Group,permutations=iters,p.adjust.m = "BH")
}

PERMANOVA_pairwise<-function(OTUS,Group,binary,iters=9999){
  Data.Dist<-vegdist(OTUS,method="bray", binary=binary)
  pairwise.adonis(Data.Dist,Group)
}

TaxName<-read.table("stability.trim.contigs.good.unique.good.filter.un
ique.precluster.pick.pds.wang.pick.tx.1.cons.taxonomy",header = T,fill
 = T)
Edit.Taxname<-function(n,level){
  if(level=="Genus"|level==1){
    n<-as.matrix(n)
    for (i in 1:4){
      n<-gsub('^.*?;', '', n)
    }
    n<-gsub(';',' ',n)
    n<-gsub('\\(100)','',n)
    n<-data.frame(n)
    n<-separate(n, col=1,into=c("Family","Genus"), sep=" ")
    x<-ifelse(n$Genus%in%c("unclassified","uncultured"), paste(n$Genus
, n$Family), paste(n$Genus,n$Other1,n$Other2))
    n<-as.matrix(x)
    return(n)
  }else if(level=="Family"|level==2){
    n<-as.matrix(n)
    for (i in 1:3){
      n<-gsub('^.*?;', '', n)
    }
    n<-gsub(';',' ',n)
    n<-gsub('\\(100)','',n)
    n<-data.frame(n)
    n<-separate(n,col=1, into=c("Order","Family","Genus"), sep=" ")
    x<-ifelse(n$Family%in%c("unclassified","uncultured"), paste(n$Orde
r, n$Family), paste(n$Family))
    n<-as.matrix(x)
    return(n)
  }else if(level=="Order"|level==3){
    n<-as.matrix(n)
    for (i in 1:2){
      n<-gsub('^.*?;', '', n)
    }
    n<-gsub(';',' ',n)
```

```
    n<-gsub('\\(100)','',n)
    n<-data.frame(n)
    n<-separate(n,col=1, into=c("Class","Order","Family","Genus"), sep
=" ")
    x<-ifelse(n$Order%in%c("unclassified","uncultured"), paste(n$Class
, n$Order), paste(n$Order))
    n<-as.matrix(x)
    return(n)
  }else if(level=="Class"|level==4){
    n<-as.matrix(n)
    for (i in 1){
      n<-gsub('^.*?;', '', n)
    }
    n<-gsub(';',' ',n)
    n<-gsub('\\(100)','',n)
    n<-data.frame(n)
    n<-separate(n,col=1, into=c("Phylum","Class","Order","Family","Gen
us"), sep=" ")
    x<-ifelse(n$Class%in%c("unclassified","uncultured"), paste(n$Phylu
m, n$Class), paste(n$Class))
    n<-as.matrix(x)
    return(n)
  }else if(level=="Phylum"|level==5){
    n<-as.matrix(n)
    n<-gsub('[(0-9);""]{1,}', '_', n)
    n<-gsub('^.*?_', '', n)
    n<-gsub('_.*', '', n)
  }
}
TaxName<-Edit.Taxname(TaxName$Taxonomy,level=1)

## Warning: Expected 2 pieces. Additional pieces discarded in 347 rows
 [1, 2, 3, 4, 5, 6,
## 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].

Subset.Taxa<-function(OTUS,TaxName,CutOff=1){
  colnames(OTUS)<-TaxName
  row<-rowSums(OTUS)
  row<-sum(row)
  col<-colSums(OTUS)
  ratio<-as.matrix(col/row*100)
  ratio<-cbind(TaxName,ratio)
  subset<-data.frame(ratio[ratio[,2]>=CutOff,])
  subset<-data.frame(subset[!subset$X1=="unclassified unclassified",])
  newOTUS<-data.frame(OTUS[,colnames(OTUS) %in% subset$X1])
  colname<-colnames(newOTUS)
```

```
  colnames(newOTUS)<-gsub("\\."," ",colname)
  return(newOTUS)
}

NB.overall<-function(newOTUS,Group){
  m<-as.matrix(NA)
  n<-as.matrix(NA)
  o<-as.matrix(NA)
  for (i in 1:ncol(newOTUS)){
    l<-glm.nb(newOTUS[,i]~Group)
    m<-anova(l)
    n[i]<-data.frame(m[2,5])
    o[i]<-colnames(newOTUS[i])
  }
  n<-p.adjust(n, method="BH")
  p<-cbind(o,n)
  return(p)
  p[,1]<-as.character(p[,1])
  p[,2]<-as.numeric(as.character(p[,2]))
  par(mar=c(10,4,1,1))
  plot(p[,2],xaxt = "n",ylim=c(0,1),xlab="",pch=16,ylab="p-value",main
="Overall p-values")
  axis(1, at=1:nrow(p), labels=FALSE)
  text(x=c(1:nrow(p)), y=par()$usr[3]-0.1*(par()$usr[4]-par()$usr[3]),
       labels=p[,1], srt=45, adj=1, xpd=TRUE)
  abline(h=0.05)
}

NB.pairwise<-function(newOTUS,Group){
  Group<-as.factor(Group)
  grp<-length(levels(Group))
  otu.name<-colnames(newOTUS)
  p.vals<-data.frame()
  comp<-c()
  for(i in 1:grp){
    if(levels(Group)[1]!=levels(Group)[i]){
      comp<-c(comp,paste(levels(Group)[1],"vs",levels(Group)[i]))
    }
  }
  for (i in 1:ncol(newOTUS)){
    l<-glm.nb(newOTUS[,i]~Group)
    m<-data.frame(coef(summary(l))[,4][2:length(levels(Group))])
    j<-1
    while(j!=grp){
      p.vals[i,j]<-m[j,]
      j<-j+1
```

```
    }
  }
  for(i in 1:(grp-1)){
    p.vals[,i]<-p.adjust(p.vals[,i], method="BH")
  }
  overall<-cbind(otu.name,p.vals)
  colnames(overall)<-c("Taxa",comp)
  return(overall)
}
```

Table 15. Population characteristics and breastfeeding patterns among exclusively breastfed infants

```
## Start analyzing the data ##
metadata<-read.csv("breast_bottle_metadata_UPDATE.csv",header = T, str
ingsAsFactors = T)
summary(metadata)
cols<-c("SEX","MD_FINAL_ROUTE","EDUC_LVL","BABY_RACE","FED_PATTERN")
metadata[cols]<-lapply(metadata[cols], factor)
sapply(metadata,class)

summary(metadata$FED_PATTERN)
#Breast:63
#Bottle:11
#Mix: 62

# BABY SEX
summary(metadata$SEX)
63/136* 100
73/136* 100
Male_breast<-filter(metadata, SEX=="1"& FED_PATTERN=="1") #N=24
nrow(Male_breast)
Male_bottle<-filter(metadata, SEX=="1"& FED_PATTERN=="2") #N=6
nrow(Male_bottle)
Male_mix<-filter(metadata, SEX=="1"& FED_PATTERN=="3") #N=33
nrow(Male_mix)
24/63* 100
6/11* 100
33/62* 100
Female_breast<-filter(metadata, SEX=="2"& FED_PATTERN=="1") #N=39
nrow(Female_breast)
Female_bottle<-filter(metadata, SEX=="2"& FED_PATTERN=="2") #N=5
nrow(Female_bottle)
Female_mix<-filter(metadata, SEX=="2"& FED_PATTERN=="3") #N=29
nrow(Female_mix)
39/63* 100
5/11* 100
```

```
29/62* 100
chisq.test(table(metadata$SEX,metadata$FED_PATTERN))  #p=0.20

# BABY race
summary(metadata$BABY_RACE)
119/136* 100
4/136* 100
13/136* 100
white_breast<-filter(metadata, BABY_RACE=="1"& FED_PATTERN=="1") #N=56
nrow(white_breast)
white_bottle<-filter(metadata, BABY_RACE=="1"& FED_PATTERN=="2") #N=10
nrow(white_bottle)
white_mix<-filter(metadata, BABY_RACE=="1"& FED_PATTERN=="3") #N=53
nrow(white_mix)
56/63* 100
10/11* 100
53/62* 100
black_breast<-filter(metadata, BABY_RACE=="2"& FED_PATTERN=="1") #N=0
nrow(black_breast)
black_bottle<-filter(metadata, BABY_RACE=="2"& FED_PATTERN=="2") #N=0
nrow(black_bottle)
black_mix<-filter(metadata, BABY_RACE=="2"& FED_PATTERN=="3") #N=4
nrow(black_mix)
4/62* 100
other_breast<-filter(metadata, BABY_RACE=="3"& FED_PATTERN=="1") #N=7
nrow(other_breast)
other_bottle<-filter(metadata, BABY_RACE=="3"& FED_PATTERN=="2") #N=1
nrow(other_bottle)
other_mix<-filter(metadata, BABY_RACE=="3"& FED_PATTERN=="3") #N=5
nrow(other_mix)
7/63* 100
1/11* 100
5/62* 100
chisq.test(table(metadata$BABY_RACE,metadata$FED_PATTERN),simulate.p.v
alue = TRUE) #p-value = 0.26

# EDUC_LVL
summary(metadata$EDUC_LVL)
10/136* 100
31/136* 100
43/136* 100
52/136* 100
high_breast<-filter(metadata, EDUC_LVL=="1"& FED_PATTERN=="1") #N=6
nrow(high_breast)
high_bottle<-filter(metadata, EDUC_LVL=="1"& FED_PATTERN=="2") #N=0
nrow(high_bottle)
```

```
high_mix<-filter(metadata, EDUC_LVL=="1"& FED_PATTERN=="3") #N=4
nrow(high_mix)
6/63* 100
0/11* 100
4/62* 100
somecoll_breast<-filter(metadata, EDUC_LVL=="2"& FED_PATTERN=="1") #N=
21
nrow(somecoll_breast)
somecoll_bottle<-filter(metadata, EDUC_LVL=="2"& FED_PATTERN=="2") #N=
1
nrow(somecoll_bottle)
somecoll_mix<-filter(metadata, EDUC_LVL=="2"& FED_PATTERN=="3") #N=9
nrow(somecoll_mix)
21/63* 100
1/11* 100
9/62* 100
Bach_breast<-filter(metadata, EDUC_LVL=="3"& FED_PATTERN=="1") #N=19
nrow(Bach_breast)
Bach_bottle<-filter(metadata, EDUC_LVL=="3"& FED_PATTERN=="2") #N=4
nrow(Bach_bottle)
Bach_mix<-filter(metadata, EDUC_LVL=="3"& FED_PATTERN=="3") #N=20
nrow(Bach_mix)
19/63* 100
4/11* 100
20/62* 100
MasPhD_breast<-filter(metadata, EDUC_LVL=="4"& FED_PATTERN=="1") #N=17
nrow(MasPhD_breast)
MasPhD_bottle<-filter(metadata, EDUC_LVL=="4"& FED_PATTERN=="2") #N=6
nrow(MasPhD_bottle)
MasPhD_mix<-filter(metadata, EDUC_LVL=="4"& FED_PATTERN=="3") #N=29
nrow(MasPhD_mix)
17/63* 100
6/11* 100
29/62* 100
chisq.test(table(metadata$EDUC_LVL,metadata$FED_PATTERN),simulate.p.va
lue = TRUE) #p-value =  0.08

# delivery mode
summary(metadata$MD_FINAL_ROUTE)
99/136* 100
37/136* 100

vag_breast<-filter(metadata, MD_FINAL_ROUTE=="1"& FED_PATTERN=="1") #N
=50
nrow(vag_breast)
vag_bottle<-filter(metadata, MD_FINAL_ROUTE=="1"& FED_PATTERN=="2") #N
```

```
=9
nrow(vag_bottle)
vag_mix<-filter(metadata, MD_FINAL_ROUTE=="1"& FED_PATTERN=="3") #N=40
nrow(vag_mix)
50/63* 100
9/11* 100
40/62* 100
csection_breast<-filter(metadata, MD_FINAL_ROUTE=="2"& FED_PATTERN=="1
") #N=13
nrow(csection_breast)
csection_bottle<-filter(metadata, MD_FINAL_ROUTE=="2"& FED_PATTERN=="2
") #N=2
nrow(csection_bottle)
csection_mix<-filter(metadata, MD_FINAL_ROUTE=="2"& FED_PATTERN=="3")
#N=22
nrow(csection_mix)
13/63* 100
2/11* 100
22/62* 100
chisq.test(table(metadata$MD_FINAL_ROUTE,metadata$FED_PATTERN),simulat
e.p.value = TRUE) #p-value = 0.15

# pre_bmi
mean(metadata$PRE_BMI)
sd(metadata$PRE_BMI)

shapiro.test(metadata$PRE_BMI) #p-value = 1.342e-09
breast<-filter(metadata, FED_PATTERN=="1")
median(breast$PRE_BMI) #24.27
min(breast$PRE_BMI) #17.57
max(breast$PRE_BMI) #47.09

bottle<-filter(metadata, FED_PATTERN=="2")
median(bottle$PRE_BMI) #23.49
min(bottle$PRE_BMI) #19.01
max(bottle$PRE_BMI) #39.46

mix<-filter(metadata, FED_PATTERN=="3")
median(mix$PRE_BMI)  #23.89
min(mix$PRE_BMI)  #17.01
max(mix$PRE_BMI)  #46.46

kruskal.test(PRE_BMI~FED_PATTERN, data =metadata) #p-value = 0.9861

# gestional age at birth
mean(metadata$ESTWKSGEST)
```

```
sd(metadata$ESTWKSGEST)

shapiro.test(metadata$ESTWKSGEST) #p-value = 2.115e-11
breast<-filter(metadata, FED_PATTERN=="1")
median(breast$ESTWKSGEST) #39
min(breast$ESTWKSGEST) #34
max(breast$ESTWKSGEST)  #41

bottle<-filter(metadata, FED_PATTERN=="2")
median(bottle$ESTWKSGEST)  #39
min(bottle$ESTWKSGEST) #37
max(bottle$ESTWKSGEST) #40

mix<-filter(metadata, FED_PATTERN=="3")
median(mix$ESTWKSGEST) #39
min(mix$ESTWKSGEST) #31
max(mix$ESTWKSGEST) #41

kruskal.test(ESTWKSGEST~FED_PATTERN, data =metadata) #p-value = 0.2286

# maternal age
mean(metadata$age_enrollment)
sd(metadata$age_enrollment)

shapiro.test(metadata$age_enrollment) #p-value = 0.0007524
breast<-filter(metadata, FED_PATTERN=="1")
median(breast$age_enrollment) #31
min(breast$age_enrollment) #20
max(breast$age_enrollment)  #51

bottle<-filter(metadata, FED_PATTERN=="2")
median(bottle$age_enrollment)  #32
min(bottle$age_enrollment) #24
max(bottle$age_enrollment) #34

mix<-filter(metadata, FED_PATTERN=="3")
median(mix$age_enrollment) #30.5
min(mix$age_enrollment) #19
max(mix$age_enrollment) #42

kruskal.test(age_enrollment~FED_PATTERN, data =metadata) #p-value =  0
.8885
```

Figure 14. The associations between alpha diversity of the gut microbiota and infant breastfeeding patterns

```
# subsample: rareified to 1383 reads
metadata<-read.csv("breast_bottle_metadata_UPDATE.csv",header = T, str
```

```
ingsAsFactors = T)
Data.Subsample<-read.csv("Data.Subsample.original.csv", header = T)
temp<-merge(Data.Subsample, metadata,by="Group")
Data.Subsample.genus<-temp[,c(1:(ncol(Data.Subsample)))]
metadata<-temp[,c(1,(ncol(Data.Subsample)+1):(ncol(temp)))]
Data.Subsample.genus$Group
metadata$Group

Data.Alpha<-Alpha(Data.Subsample.genus[,-c(1:3)])
shapiro.test(Data.Alpha$Chao) #p-value = 5.556e-10
shapiro.test(Data.Alpha$Shannon) #p-value = 0.02078
shapiro.test(Data.Alpha$Invsimpson) #p-value =0.1702

#Chao 1
kruskal.test(Data.Alpha$Chao~metadata$FED_PATTERN)  #p-value = 0.148
#Shannon
kruskal.test(Data.Alpha$Shannon~metadata$FED_PATTERN)  #p-value =0.385
2

#inverse Simpson
summary(aov(Data.Alpha$Invsimpson~metadata$FED_PATTERN)) #p=0.198

png("Alpha_breastfeeding_pattern_UPDATE.png", res=300, height=4, width
=12,units="in")
par(mfrow= c(1,3),cex.main=1.8,cex.axis=1.8,mar=c(7, 5, 3, 1))
label<-c("Breast","Bottle","Mix")
boxplot(Data.Alpha$Chao~metadata$FED_PATTERN,main="A. Chao1 index",yla
b="Chao1 Index",xlab="Breastfeeding patterns",names=label,cex.lab = 2)
text(labels="p-value=0.15", x=2, y=80,cex=1.8)
boxplot(Data.Alpha$Shannon~metadata$FED_PATTERN,main="B. Shannon index
",ylab="Shannon Index",xlab="Breastfeeding patterns",names=label,cex.l
ab = 2)
text(labels="p-value=0.39", x=2, y=0.5,cex=1.8)
boxplot(Data.Alpha$Invsimpson~metadata$FED_PATTERN,main="C. inverse Si
mpson index",ylab="inverse Simpson Index",xlab="Breastfeeding patterns
",names=label,cex.lab = 2)
text(labels="p-value=0.20", x=2, y=1.6,cex=1.8)
while (!is.null(dev.list()))  dev.off()
```

Figure 15. The associations between beta diversity of the gut microbiota and infant breastfeeding patterns

```
#Sorensen
a<-as.factor(metadata$FED_PATTERN)
PERMANOVA(Data.Subsample.genus[,-c(1:3)],a,TRUE,9999) #P=0.0263
Sor_pattern<-Sor.bray.pcoa(Data.Subsample.genus[,-c(1:3)],Dim=2,Color=
a,binary=TRUE)
```

```
#Bray-Curtis
PERMANOVA(Data.Subsample.genus[,-c(1:3)],a,FALSE,9999) #P=0.4839
Bray_pattern<-Sor.bray.pcoa(Data.Subsample.genus[,-c(1:3)],Dim=2,Color
=a,binary=FALSE)

png("Beta_breastfeeding_pattern_UPDATE.png", res=300, height=5, width=
12,units="in")
par(mfrow= c(1,2))
Color_pattern<-ifelse(grepl("1", a),"#000000", ifelse(grepl("2", a),"#
E79F00","#0072B2"))
plot(Sor_pattern,cex.axis=1.5,cex.lab=1.5,cex.main=2,cex=2,col=1,
     pch=21,xlim=c(-.38,.38),ylim=c(-.3,.35),xlab="PC1 (21.1%)",ylab="
PC2 (9%)",bg=Color_pattern,main="A. Sorensen")
ordiellipse(Sor_pattern,groups=a,col= c("#000000","#E79F00","#0072B2")
,lwd=2)
legend(0,0.35,c("Breast","Bottle","Mix"), pch=21,cex = 1.2,pt.bg=c("#0
00000","#E79F00","#0072B2"),y.intersp = 0.72)
text(0.2,-0.26, labels= "p-value=0.03",cex=1.4)

Color_pattern<-ifelse(grepl("1", a),"#000000", ifelse(grepl("2", a),"#
E79F00","#0072B2"))
plot(Bray_pattern,cex.axis=1.5,cex.lab=1.5,cex.main=2,cex=2,col=1,
     pch=21,xlim=c(-.35,.55),ylim=c(-.55,.5),xlab="PC1 (25.9%)",ylab="
PC2 (14.9%)",bg=Color_pattern,main="B. Bray-Curtis")
ordiellipse(Bray_pattern,groups=a,col= c("#000000","#E79F00","#0072B2"
),lwd=2)
legend(0.15,0.51,c("Breast","Bottle","Mix"), pch=21,cex = 1.2,pt.bg=c(
"#000000","#E79F00","#0072B2"),y.intersp = 0.72)
text(0.44,-0.4, labels= "p-value=0.48",cex=1.4)
while (!is.null(dev.list()))  dev.off()
```

Figure 16. The associations between alpha diversity of the gut microbiota and breastfeeding patterns inthe 24 hours immediately preceding stool sample collection for infants exclusively fed human milk and dietary intake in the past week for infants fed at least some formula

```
metadata<-read.csv("breast_bottle_feed_past_wk_UPDATE.csv",header=T, s
tringsAsFactors = T)
Data.Subsample<-read.csv("Data.Subsample.csv",header = T)
temp<-merge(Data.Subsample, metadata,by="Group")
Data.Subsample.genus<-temp[,c(1:(ncol(Data.Subsample)))]  #N=299
metadata<-temp[,c(1,(ncol(Data.Subsample)+1):(ncol(temp)))] #N=299
Data.Subsample.genus$Group
metadata$Group

Data.Alpha<-Alpha(Data.Subsample.genus[,-c(1:3)],Groups=metadata$FED_P
ATTERN_50CUT)
```

```
shapiro.test(Data.Alpha$Chao) #p-value = 4.441e-13
shapiro.test(Data.Alpha$Shannon) #p-value = 0.1949
shapiro.test(Data.Alpha$Invsimpson) #p-value = 5.107e-11

metadata$FED_PATTERN_50CUT<-as.factor(metadata$FED_PATTERN_50CUT)
levels(metadata$FED_PATTERN_50CUT)

#Chao 1
kruskal.test(Data.Alpha$Chao~metadata$FED_PATTERN_50CUT)  #p-value =1.
903e-06
dunn.test(Data.Alpha$Chao,metadata$FED_PATTERN_50CUT,altp = TRUE, meth
od="bh")

p1<-ggplot(Data.Alpha,aes(x=as.factor(Groups), y=Chao)) +
  stat_boxplot(geom ='errorbar')+
  geom_boxplot(outlier.shape = NA)+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, linewid
th=1))+
  scale_x_discrete(labels=c("Breast","Bottle","Mix","Breastmilk>50","B
reastmilk≤50","Formula"))+
  labs(y= "Chao1 index", x="",title ="Chao1 index and feeding methods"
)+
  theme(text = element_text(size=23),plot.title = element_text(size =
23,hjust = 0.5),axis.text.x=element_text(size=23, angle = 45,hjust = 1
),axis.text.y = element_text(size=23))+
  geom_bracket(xmin = "1", xmax = "3", y.position = 100, label = "Brea
stfeeding patterns \n in the past day", tip.length = c(0.08, 0.08),lab
el.size=7)+
  geom_bracket(xmin = "4", xmax = "6", y.position = 100, label = "Diet
ary intake \n in the past week", tip.length = c(0.08, 0.08),label.size
=7)+
  annotate("text", x=1, y=61.5, label= "ab",size=7)+
  annotate("text", x=2, y=28.5, label= "a",size=7)+
  annotate("text", x=3, y=60.5, label= "a",size=7)+
  annotate("text", x=4, y=88, label= "bc",size=7)+
  annotate("text", x=5, y=63, label= "bc",size=7)+
  annotate("text", x=6, y=74.5, label= "c",size=7)+
  annotate("text", x=2, y=86, label= "p-value<0.01",size=7)+
  scale_y_continuous(limits = c(10, 120))


#Shannon
summary(aov(Data.Alpha$Shannon~metadata$FED_PATTERN_50CUT)) #p=2e-16
```

```
TukeyHSD(aov(Data.Alpha$Shannon~metadata$FED_PATTERN_50CUT))

p2<-ggplot(Data.Alpha,aes(x=as.factor(Groups), y=Shannon)) +
  stat_boxplot(geom ='errorbar')+
  geom_boxplot(outlier.shape = NA)+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, linewid
th=1))+
  scale_x_discrete(labels=c("Breast","Bottle","Mix","Breastmilk>50","B
reastmilk≤50","Formula"))+
  labs(y= "Shannon index", x="",title ="Shannon index and feeding meth
ods")+
  theme(text = element_text(size=23),plot.title = element_text(size =
23,hjust = 0.5),axis.text.x=element_text(size=23, angle = 45,hjust = 1
),axis.text.y = element_text(size=23))+
  geom_bracket(xmin = "1", xmax = "3", y.position = 3, label = "Breast
feeding patterns \n in the past day", tip.length = c(0.08, 0.08),label
.size=7)+
  geom_bracket(xmin = "4", xmax = "6", y.position = 3, label = "Dietar
y intake \n in the past week", tip.length = c(0.08, 0.001),label.size=
7)+
  annotate("text", x=1, y=2.44, label= "ab",size=7)+
  annotate("text", x=2, y=2.12, label= "ab",size=7)+
  annotate("text", x=3, y=2.53, label= "a",size=7)+
  annotate("text", x=4, y=2.69, label= "b",size=7)+
  annotate("text", x=5, y=2.86, label= "c",size=7)+
  annotate("text", x=6, y=3.1, label= "c",size=7)+
  annotate("text", x=5.5, y=1, label= "p-value<0.01",size=7)+
  scale_y_continuous(limits = c(0.7, 3.5))


#inverse Simpson
kruskal.test(Data.Alpha$Invsimpson~metadata$FED_PATTERN_50CUT)  # p-va
lue < 2.2e-16
dunn.test(Data.Alpha$Invsimpson,metadata$FED_PATTERN_50CUT,altp = TRUE
, method="bh")

p3<-ggplot(Data.Alpha,aes(x=as.factor(Groups), y=Invsimpson)) +
  stat_boxplot(geom ='errorbar')+
  geom_boxplot(outlier.shape = NA)+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, linewid
```

```
th=1))+
  scale_x_discrete(labels=c("Breast","Bottle","Mix","Breastmilk>50","B
reastmilk≤50","Formula"))+
  labs(y= "inverse Simpson index", x="",title ="Inverse Simpson index
and feeding methods")+
  theme(text = element_text(size=23),plot.title = element_text(size =
23,hjust = 0.5),axis.text.x=element_text(size=23, angle = 45,hjust = 1
),axis.text.y = element_text(size=23))+
  geom_bracket(xmin = "1", xmax = "3", y.position = 15, label = "Breas
tfeeding patterns \n in the past day", tip.length = c(0.08, 0.08),labe
l.size=7)+
  geom_bracket(xmin = "4", xmax = "6", y.position = 15, label = "Dieta
ry intake \n in the past week", tip.length = c(0.08, 0.08),label.size=
7)+
  annotate("text", x=1, y=7.9, label= "ab", size=7)+
  annotate("text", x=2, y=5.95, label= "ab",size=7)+
  annotate("text", x=3, y=7.3, label= "a",size=7)+
  annotate("text", x=4, y=8.7, label= "b",size=7)+
  annotate("text", x=5, y=11.17, label= "c",size=7)+
  annotate("text", x=6, y=11.5, label= "c",size=7)+
  annotate("text", x=2, y=12, label= "p-value<0.01",size=7)

png("Alpha_diversity_6feedinggroups_50cutoff_UPDATE_vertical.png", res
=300, height=20, width=8,units="in")
ggarrange(p1, p2,p3, labels = c("A", "B","C"),font.label=list(size=28)
, nrow = 3, ncol = 1)

## Warning: Removed 4 rows containing non-finite values (`stat_boxplot
()`).
## Removed 4 rows containing non-finite values (`stat_boxplot()`).

## Warning: Removed 2 rows containing non-finite values (`stat_boxplot
()`).
## Removed 2 rows containing non-finite values (`stat_boxplot()`).

while (!is.null(dev.list()))  dev.off()
```

Figure 17.The associations between beta diversity of the gut microbiota and breastfeeding patterns in the past day for exclusively human milk fed infants and dietary intake in the past week for infants fed at least some formula

Table 16. Significant pairwise comparisons of the relationships between beta diversity of the gut microbiota and breastfeeding patterns in the past day and dietary intake in the past week

```
metadata<-read.csv("breast_bottle_feed_past_wk_UPDATE.csv",header=T, s
tringsAsFactors = T)
Data.Subsample<-read.csv("Data.Subsample.csv",header = T)
temp<-merge(Data.Subsample, metadata,by="Group")
```

```
Data.Subsample.genus<-temp[,c(1:(ncol(Data.Subsample)))]  #N=299
metadata<-temp[,c(1,(ncol(Data.Subsample)+1):(ncol(temp)))] #N=299
Data.Subsample.genus$Group
metadata$Group

a<-metadata$FED_PATTERN_50CUT
#Sorensen
PERMANOVA(Data.Subsample.genus[,-c(1:3)],a,TRUE,9999) #P=1e-04
Sor_pattern<-Sor.bray.pcoa(Data.Subsample.genus[,-c(1:3)],Dim=2,Color=
a,binary=TRUE)

b<-PERMANOVA_pairwise(Data.Subsample.genus[,-c(1:3)],a,TRUE,9999)
#Bray-Curtis
PERMANOVA(Data.Subsample.genus[,-c(1:3)],a,FALSE,9999) #P=1e-04
Bray_pattern<-Sor.bray.pcoa(Data.Subsample.genus[,-c(1:3)],Dim=2,Color
=a,binary=FALSE)

c<-PERMANOVA_pairwise(Data.Subsample.genus[,-c(1:3)],a,FALSE,9999)

Color<-ifelse(grepl("1", a),"#009392", ifelse(grepl("2", a),"#39b1b5",
ifelse(grepl("3", a),"#9ccb86",ifelse(grepl("4", a),"#e9e29c",ifelse(g
repl("5", a),"#eeb479","#e88471")))))

png("Beta_diversity_6feedinggroups_50cutoff_UPDATE.png", res=300, heig
ht=5, width=10,units="in")
par(mfrow= c(1,2))
plot(Sor_pattern,cex.axis=1.5,cex.lab=1.5,cex.main=2,cex=1.6,col=1,
     pch=21,xlim=c(-.4,.41),ylim=c(-.3,.36),xlab="PC1 (21.2%)",ylab="P
C2 (10%)",bg=Color,main="A. Sorensen")
ordiellipse(Sor_pattern,groups=a,col= c("#009392","#39b1b5","#9ccb86",
"#e9e29c","#eeb479","#e88471"),lwd=3)
legend(0.16,0.375,c("Breast","Bottle","Mix","Breastmilk>50","Breastmil
k≤50","Formula"),pch=21,cex = 0.8,y.intersp = 0.72,pt.bg=c("#009392","
#39b1b5","#9ccb86","#e9e29c","#eeb479","#e88471"))
text(-0.25,-0.26, labels= "p-value<0.01",cex=0.95)

plot(Bray_pattern,cex.axis=1.5,cex.lab=1.5,cex.main=2,cex=1.6,col=1,
     pch=21,xlim=c(-.45,.47),ylim=c(-.55,.35),xlab="PC1 (21.6%)",ylab=
"PC2 (13.3%)",bg=Color,main="B. Bray-Curtis")
ordiellipse(Bray_pattern,groups=a,col= c("#009392","#39b1b5","#9ccb86"
,"#e9e29c","#eeb479","#e88471"),lwd=3)
legend(0.195,-0.315,c("Breast","Bottle","Mix","Breastmilk>50","Breastm
ilk≤50","Formula"),pch=21,cex = 0.8,y.intersp = 0.72,pt.bg=c("#009392"
,"#39b1b5","#9ccb86","#e9e29c","#eeb479","#e88471"))
text(0.37,0.3, labels= "p-value<0.01",cex=0.95)
while (!is.null(dev.list()))  dev.off()
```

Figure 18. The comparison of the relative abundance of taxa in three groups of breastfeeding patterns

```
metadata<-read.csv("breast_bottle_metadata_UPDATE.csv",header = T, str
ingsAsFactors = T)
Data.Subsample<-read.csv("Data.Subsample.original.csv", header = T)
temp<-merge(Data.Subsample, metadata,by="Group")
Data.Subsample.genus<-temp[,c(1:(ncol(Data.Subsample)))]
metadata<-temp[,c(1,(ncol(Data.Subsample)+1):(ncol(temp)))]
Data.Subsample.genus$Group
metadata$Group

#chose the taxa with rel abun >1%
newOTUS<-Subset.Taxa(Data.Subsample.genus[,-c(1:3)],TaxName=TaxName,Cu
tOff=1)
#calculate the overall p-value
Group<-as.factor(metadata$FED_PATTERN)
p<-NB.overall(newOTUS,Group)

#negative binomial
temp<-NB.pairwise(newOTUS=newOTUS,Group=Group)

Group<-factor(Group, levels(Group)[c(2,1,3)])
levels(Group)
temp2<-NB.pairwise(newOTUS,Group)

pairwise<-cbind(p,temp,temp2)
#write.csv(pairwise,"Negative_biomial_breastfeeding_pattern_p-values_o
riginal_UPDATE.csv",row.names = F)

NB.pair<-read.csv("Negative_biomial_breastfeeding_pattern_p-values_ori
ginal_UPDATE.csv",header = T)
colnames(NB.pair)<-c("Taxa","Breast vs Bottle","Breast vs Mix","Bottle
 vs Mix")

p.plot<-function(NB.pair,title=""){
  taxa<-NB.pair[,1]
  p<-NB.pair[,-1]
  par(mar=c(11,6,3,4))
  plot(p[,1],xaxt = "n",ylim=c(0,1),xlab="",pch=16,ylab="p-value",main
=paste(title))
  text(x=c(1:length(taxa)), y=par()$usr[3]-0.03*(par()$usr[4]-par()$us
r[3]),
       labels=taxa, srt=45, adj=1, xpd=TRUE)
  legend(12,.4,legend=paste(colnames(p)), pch=16,col=seq(1,ncol(p)),ce
x = 0.7)
  axis(1, at=1:nrow(p), labels=FALSE)
  abline(h=0.1)
```

```
  for(i in 2:ncol(p)-1){
    par(new=TRUE)
    plot(jitter(1:nrow(p)),p[,i+1],ylim=c(0,1),xaxt ="n",pch=16,xlab="
",yaxt = "n",ylab="",col=c(i+1))
  }
}
p.plot(NB.pair)

png("Top15taxa_breastfeeding_pattern_UPDATE.png", res=300, height=5.5,
 width=7,units="in")
p.plot(NB.pair)
while (!is.null(dev.list()))  dev.off()
```

Table 17. The relative abundance of taxa in three groups of breastfeeding patterns

```
NB.table<-function(OTUS,newOTUS,Group){
  total<-rowSums(OTUS)
  rel.otu<-newOTUS/total*100
  overall<-paste(round(colMeans(rel.otu),1),"\u00b1",round(apply(rel.o
tu,2,sd),1))
  taxa.mean<-as.matrix(round(aggregate(rel.otu,list(Group),mean)[,-1],
1))
  taxa.sd<-as.matrix(round(aggregate(rel.otu,list(Group),sd)[,-1],1))
  taxa1<-t(matrix(nrow=3,paste(taxa.mean,"\u00b1",taxa.sd)))
  colnames(taxa1)<-levels(Group)
  tables<-cbind(matrix(colnames(taxa.mean)),overall,taxa1)
  return(tables)
}
test<-NB.table(Data.Subsample.genus[,-c(1:3)],newOTUS,Group)
test
test_p<-cbind(test,p)
write.csv(test_p,"Negative_biomial_breastfeeding_pattern_UPDATE.csv",r
ow.names = F)
```

Table 18. The relative abundance of taxa in six feeding groups, results from NB

```
metadata<-read.csv("breast_bottle_feed_past_wk_UPDATE.csv",header=T, s
tringsAsFactors = T)
Data.Subsample<-read.csv("Data.Subsample.csv",header = T)
temp<-merge(Data.Subsample, metadata,by="Group")
Data.Subsample.genus<-temp[,c(1:(ncol(Data.Subsample)))]  #N=299
metadata<-temp[,c(1,(ncol(Data.Subsample)+1):(ncol(temp)))] #N=299
Data.Subsample.genus$Group
metadata$Group

newOTUS<-Subset.Taxa(Data.Subsample.genus[,-c(1:3)],TaxName=TaxName,Cu
tOff=1)  #N=15
Group<-as.factor(metadata$FED_PATTERN_50CUT)
p<-NB.overall(newOTUS,Group)
```

```
NB_taxa<-as.data.frame(p)

temp<-NB.pairwise(newOTUS,Group)
Group<-factor(Group, levels(Group)[c(2,1,3,4,5,6)])
levels(Group)
temp2<-NB.pairwise(newOTUS,Group)

Group<-factor(Group, levels(Group)[c(3,1,2,4,5,6)])
levels(Group)
temp3<-NB.pairwise(newOTUS,Group)

Group<-factor(Group, levels(Group)[c(4,1,2,3,5,6)])
levels(Group)
temp4<-NB.pairwise(newOTUS,Group)

Group<-factor(Group, levels(Group)[c(5,1,2,3,4,6)])
levels(Group)
temp5<-NB.pairwise(newOTUS,Group)

Group<-factor(Group, levels(Group)[c(6,1,2,3,4,5)])
levels(Group)
temp6<-NB.pairwise(newOTUS,Group)

pairwise<-cbind(p,temp,temp2,temp3,temp4,temp5,temp6)
write.csv(pairwise,"Negative_biomial_all_feeding_groups_p-values_UPDAT
E.csv",row.names = F)

NB.table<-function(OTUS,newOTUS,Group){
  total<-rowSums(OTUS)
  rel.otu<-newOTUS/total*100
  overall<-paste(round(colMeans(rel.otu),1),"\u00b1",round(apply(rel.o
tu,2,sd),1))
  taxa.mean<-as.matrix(round(aggregate(rel.otu,list(Group),mean)[,-1],
1))
  taxa.sd<-as.matrix(round(aggregate(rel.otu,list(Group),sd)[,-1],1))
  taxa1<-t(matrix(nrow=6,paste(taxa.mean,"\u00b1",taxa.sd)))
  colnames(taxa1)<-levels(Group)
  tables<-cbind(matrix(colnames(taxa.mean)),overall,taxa1)
  return(tables)
}
test<-NB.table(Data.Subsample.genus[,-c(1:3)],newOTUS,Group)
test
write.csv(test,"Negative_biomial_6_groups_pattern_rel_abun_UPDATE.csv"
,row.names = F)
```

Figure 19. The comparison of the relative abundance of taxa in six feeding groups, results from MaAsLin

```
metadata<-read.csv("breast_bottle_feed_past_wk_UPDATE.csv",header=T, s
tringsAsFactors = T)
Data.Subsample<-read.csv("Data.Subsample.csv",header = T)
temp<-merge(Data.Subsample, metadata,by="Group")
Data.Subsample.genus<-temp[,c(1:(ncol(Data.Subsample)))]  #N=299
metadata<-temp[,c(1,(ncol(Data.Subsample)+1):(ncol(temp)))] #N=299
Data.Subsample.genus$Group
metadata$Group

summary(metadata)
cols<-c("SEX","MD_FINAL_ROUTE","EDUC_LVL","BABY_RACE","FED_PATTERN_50C
UT")
metadata[cols]<-lapply(metadata[cols], factor)
sapply(metadata,class)

rownames(Data.Subsample.genus)<-Data.Subsample.genus$Group
rownames(metadata)<-metadata$Group
metadata<-metadata[,-1]

metadata$FED_PATTERN_50CUT<-as.character(metadata$FED_PATTERN_50CUT)
metadata$FED_PATTERN_STRING[metadata$FED_PATTERN_50CUT=="1"]<-"Breast"
metadata$FED_PATTERN_STRING[metadata$FED_PATTERN_50CUT=="2"]<-"Bottle"
metadata$FED_PATTERN_STRING[metadata$FED_PATTERN_50CUT=="3"]<-"Mix"
metadata$FED_PATTERN_STRING[metadata$FED_PATTERN_50CUT=="4"]<-"Breastm
ilk>50"
metadata$FED_PATTERN_STRING[metadata$FED_PATTERN_50CUT=="5"]<-"Breastm
ilk<50"
metadata$FED_PATTERN_STRING[metadata$FED_PATTERN_50CUT=="6"]<-"Formula
"

Data.Subsample.genus<-Data.Subsample.genus[,-c(1:3)]
Data.Subsample.genus<-t(Data.Subsample.genus)
row.names(Data.Subsample.genus)<-TaxName
Data.Subsample.genus<-t(Data.Subsample.genus)
Data.Subsample.genus<-as.data.frame(Data.Subsample.genus)

metadata$FED_PATTERN_STRING<-as.factor(metadata$FED_PATTERN_STRING)

Maaslin_multi<-Maaslin2(
  input_data = Data.Subsample.genus,
 input_metadata = metadata,
  output = "Maaslin_6_feeding_groups_multivariate_Breast_control_UPDAT
E",
  fixed_effects = c("FED_PATTERN_STRING","SEX","MD_FINAL_ROUTE","EDUC_
LVL","BABY_RACE","ESTWKSGEST","PRE_BMI","Has.baby.had.antibiotics.sinc
```

```
e.birth.","age_enrollment"),
  reference = c("FED_PATTERN_STRING,Breast"))

Maaslin_multi<-Maaslin2(
 input_data = Data.Subsample.genus,
  input_metadata = metadata,
  output = "Maaslin_6_feeding_groups_multivariate_Bottle_control_UPDAT
E",
  fixed_effects = c("FED_PATTERN_STRING","SEX","MD_FINAL_ROUTE","EDUC_
LVL","BABY_RACE","ESTWKSGEST","PRE_BMI","Has.baby.had.antibiotics.sinc
e.birth.","age_enrollment"),
  reference = c("FED_PATTERN_STRING,Bottle"))

Maaslin_multi<-Maaslin2(
  input_data = Data.Subsample.genus,
  input_metadata = metadata,
  output = "Maaslin_6_feeding_groups_multivariate_Mix_control_UPDATE",
  fixed_effects = c("FED_PATTERN_STRING","SEX","MD_FINAL_ROUTE","EDUC_
LVL","BABY_RACE","ESTWKSGEST","PRE_BMI","Has.baby.had.antibiotics.sinc
e.birth.","age_enrollment"),
  reference = c("FED_PATTERN_STRING,Mix"))

Maaslin_multi<-Maaslin2(
  input_data = Data.Subsample.genus,
  input_metadata = metadata,
  output = "Maaslin_6_feeding_groups_multivariate_Breastmilk>50_contro
l_UPDATE",
  fixed_effects = c("FED_PATTERN_STRING","SEX","MD_FINAL_ROUTE","EDUC_
LVL","BABY_RACE","ESTWKSGEST","PRE_BMI","Has.baby.had.antibiotics.sinc
e.birth.","age_enrollment"),
  reference = c("FED_PATTERN_STRING,Breastmilk>50"))

Maaslin_multi<-Maaslin2(
  input_data = Data.Subsample.genus,
  input_metadata = metadata,
  output = "Maaslin_6_feeding_groups_multivariate_Breastmilk<50_contro
l_UPDATE",
  fixed_effects = c("FED_PATTERN_STRING","SEX","MD_FINAL_ROUTE","EDUC_
LVL","BABY_RACE","ESTWKSGEST","PRE_BMI","Has.baby.had.antibiotics.sinc
e.birth.","age_enrollment"),
  reference = c("FED_PATTERN_STRING,Breastmilk<50"))

Maaslin_multi<-Maaslin2(
  input_data = Data.Subsample.genus,
  input_metadata = metadata,
  output = "Maaslin_6_feeding_groups_multivariate_Formula_control_UPDA
TE",
```

```
  fixed_effects = c("FED_PATTERN_STRING","SEX","MD_FINAL_ROUTE","EDUC_
LVL","BABY_RACE","ESTWKSGEST","PRE_BMI","Has.baby.had.antibiotics.sinc
e.birth.","age_enrollment"),
  reference = c("FED_PATTERN_STRING,Formula"))

#combine the data
#Breast as control.Every level compare to reference
all_breast_control<-read.table("/Users/busihan/Desktop/Thesis_aim3/Maa
slin_6_feeding_groups_multivariate_Breast_control_UPDATE/all_results.t
sv", header = T)
all_breast_control$feature<-gsub("\\."," ", all_breast_control$feature
)
all_breast_control<-filter(all_breast_control, metadata=="FED_PATTERN_
STRING")
Maaslin_NB_taxa_breast<-merge(NB_taxa,all_breast_control,by.x="o", by.
y="feature")

#Bottle vs Breast
Maaslin_NB_taxa_breast$value[Maaslin_NB_taxa_breast$value=="Bottle"]<-
"Bottle vs Breast"
Bottle_vs_Breast<-filter(Maaslin_NB_taxa_breast,value=="Bottle vs Brea
st")

#Mix vs Breast
Maaslin_NB_taxa_breast$value[Maaslin_NB_taxa_breast$value=="Mix"]<-"Mi
x vs Breast"
Mix_vs_Breast<-filter(Maaslin_NB_taxa_breast,value=="Mix vs Breast")

#Breast>50% vs Breast
Maaslin_NB_taxa_breast$value[Maaslin_NB_taxa_breast$value=="Breastmilk
>50"]<-"Breastmilk>50% vs Breast"
large50_vs_Breast<-filter(Maaslin_NB_taxa_breast,value=="Breastmilk>50
% vs Breast")

#Breast<=50% vs Breast
Maaslin_NB_taxa_breast$value[Maaslin_NB_taxa_breast$value=="Breastmilk
<50"]<-"Breastmilk≤50% vs Breast"
less50_vs_Breast<-filter(Maaslin_NB_taxa_breast,value=="Breastmilk≤50%
 vs Breast")

#########################################################
#Bottle as control. Every level compare to reference
all_bottle_control<-read.table("/Users/busihan/Desktop/2023Mar27_Aim3_
double_check/Maaslin_6_feeding_groups_multivariate_Bottle_control_UPDA
TE/all_results.tsv", header = T)
all_bottle_control$feature<-gsub("\\."," ", all_bottle_control$feature
```

```
)
all_bottle_control<-filter(all_bottle_control, metadata =="FED_PATTERN
_STRING")
Maaslin_NB_taxa_bottle<-merge(NB_taxa,all_bottle_control,by.x="o", by.
y="feature")

#Mix vs Bottle
Maaslin_NB_taxa_bottle$value[Maaslin_NB_taxa_bottle$value=="Mix"]<-"Mi
x vs Bottle"
Mix_vs_Bottle<-filter(Maaslin_NB_taxa_bottle,value=="Mix vs Bottle")

#Breast>50% vs Bottle
Maaslin_NB_taxa_bottle$value[Maaslin_NB_taxa_bottle$value=="Breastmilk
>50"]<-"Breastmilk>50% vs Bottle"
large50_vs_Bottle<-filter(Maaslin_NB_taxa_bottle,value=="Breastmilk>50
% vs Bottle")

#Breast<=50% vs Breast
Maaslin_NB_taxa_bottle$value[Maaslin_NB_taxa_bottle$value=="Breastmilk
<50"]<-"Breastmilk≤50% vs Bottle"
less50_vs_Bottle<-filter(Maaslin_NB_taxa_bottle,value=="Breastmilk≤50%
 vs Bottle")

#########################################################
#mix control. Every level compare to reference
#########################################################
all_mix_control<-read.table("/Users/busihan/Desktop/2023Mar27_Aim3_dou
ble_check/Maaslin_6_feeding_groups_multivariate_Mix_control_UPDATE/all
_results.tsv", header = T)
all_mix_control$feature<-gsub("\\."," ", all_mix_control$feature)
all_mix_control<-filter(all_mix_control, metadata =="FED_PATTERN_STRIN
G")
Maaslin_NB_taxa_mix<-merge(NB_taxa,all_mix_control,by.x="o", by.y="fea
ture")

#Breast>50% vs Mix
Maaslin_NB_taxa_mix$value[Maaslin_NB_taxa_mix$value=="Breastmilk>50"]<
-"Breastmilk>50% vs Mix"
large50_vs_Mix<-filter(Maaslin_NB_taxa_mix,value=="Breastmilk>50% vs M
ix")

#Breast<=50% vs Mix
Maaslin_NB_taxa_mix$value[Maaslin_NB_taxa_mix$value=="Breastmilk<50"]<
-"Breastmilk≤50% vs Mix"
less50_vs_Mix<-filter(Maaslin_NB_taxa_mix,value=="Breastmilk≤50% vs Mi
x")
```

```
#######################################################
# breastmilk<50. Every level compare to reference
#######################################################
all_breast_less50_control<-read.table("/Users/busihan/Desktop/2023Mar27_Aim3_double_check/Maaslin_6_feeding_groups_multivariate_Breastmilk<50_control_UPDATE/all_results.tsv", header = T)
all_breast_less50_control$feature<-gsub("\\."," ", all_breast_less50_control$feature)
all_breast_less50_control<-filter(all_breast_less50_control, metadata =="FED_PATTERN_STRING")
Maaslin_NB_taxa_less50<-merge(NB_taxa,all_breast_less50_control,by.x="o", by.y="feature")

#Breast>50% vs Breast≤50%
Maaslin_NB_taxa_less50$value[Maaslin_NB_taxa_less50$value=="Breastmilk>50"]<-"Breastmilk>50% vs Breastmilk≤50%"
Breast_more50_vs_Breast_less50<-filter(Maaslin_NB_taxa_less50,value=="Breastmilk>50% vs Breastmilk≤50%")

#######################################################
# Formula. Every level compare to reference
#######################################################
all_formula_control<-read.table("/Users/busihan/Desktop/2023Mar27_Aim3_double_check/Maaslin_6_feeding_groups_multivariate_formula_control_UPDATE/all_results.tsv", header = T)
all_formula_control$feature<-gsub("\\."," ", all_formula_control$feature)
all_formula_control<-filter(all_formula_control, metadata =="FED_PATTERN_STRING")
Maaslin_NB_taxa_formula<-merge(NB_taxa,all_formula_control,by.x="o", by.y="feature")

#Breast vs Formula
Maaslin_NB_taxa_formula$value[Maaslin_NB_taxa_formula$value=="Breast"]<-"Breast vs Formula"
Breast_vs_Formula<-filter(Maaslin_NB_taxa_formula,value=="Breast vs Formula")

#Bottle vs Formula
Maaslin_NB_taxa_formula$value[Maaslin_NB_taxa_formula$value=="Bottle"]<-"Bottle vs Formula"
Bottle_vs_Formula<-filter(Maaslin_NB_taxa_formula,value=="Bottle vs Formula")

#Mix vs Formula
```

```
Maaslin_NB_taxa_formula$value[Maaslin_NB_taxa_formula$value=="Mix"]<-"
Mix vs Formula"
Mix_vs_Formula<-filter(Maaslin_NB_taxa_formula,value=="Mix vs Formula"
)

#Breastmilk>50% vs Formula
Maaslin_NB_taxa_formula$value[Maaslin_NB_taxa_formula$value=="Breastmi
lk>50"]<-"Breastmilk>50% vs Formula"
Breastmilk_large50_vs_Formula<-filter(Maaslin_NB_taxa_formula,value=="
Breastmilk>50% vs Formula")

#Breastmilk<50% vs Formula
Maaslin_NB_taxa_formula$value[Maaslin_NB_taxa_formula$value=="Breastmi
lk<50"]<-"Breastmilk≤50% vs Formula"
Breastmilk_less50_vs_Formula<-filter(Maaslin_NB_taxa_formula,value=="B
reastmilk≤50% vs Formula")

all_comparsion<-rbind(Bottle_vs_Breast,Mix_vs_Breast,large50_vs_Breast
,less50_vs_Breast,Mix_vs_Bottle,large50_vs_Bottle,less50_vs_Bottle,lar
ge50_vs_Mix,less50_vs_Mix,Breast_more50_vs_Breast_less50,Breast_vs_For
mula,Bottle_vs_Formula,Mix_vs_Formula,Breastmilk_large50_vs_Formula,Br
eastmilk_less50_vs_Formula)

write.csv(all_comparsion,"Maaslin_all_comparsion_correct_order_USE_THI
S_UPDATE.csv", row.names = F)

#all_comparsion<-read.csv("Maaslin_all_comparsion_correct_order_USE_TH
IS_UPDATE.csv", header = T)

all_comparsion$value<-factor(all_comparsion$value, levels=c("Breastmil
k>50% vs Breastmilk≤50%","Breastmilk≤50% vs Mix","Breastmilk>50% vs Mi
x","Breastmilk≤50% vs Bottle","Breastmilk>50% vs Bottle","Mix vs Bottl
e","Breastmilk≤50% vs Breast","Breastmilk>50% vs Breast","Mix vs Breas
t","Bottle vs Breast","Breastmilk≤50% vs Formula","Breastmilk>50% vs F
ormula","Mix vs Formula","Bottle vs Formula","Breast vs Formula"))


ggplot(all_comparsion,aes(x=o,y=value, fill=coef))+
  geom_tile()+
  scale_fill_gradient2(low = "#2166ac",high = "#b2182b")+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank(),
        panel.background = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, linewid
th=1))+
  labs(x= "", y="",fill='Beta \ncoefficient')+
```

```
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=0.97,
size=11),axis.text.y = element_text(size=10))+
  theme(plot.margin = margin(0.5,0.05,0.05,3, "cm"))

ggsave("Heatmap_UPDATE_UPDATE.png",width = 8, height = 5)
```