I, Shannon C. Conrey, hereby submit this original work as part of the requirements for the degree of Master of Science in Nutrition.

It is entitled:
The relationship between infant feeding practices and the development of the gut microbiota

Student’s name: Shannon C. Conrey

This work and its defense approved by:

Committee chair: Abigail Pearls, Ph.D.

Committee member: Ardythe Morrow, Ph.D.

Committee member: Nicholas Ollberding
The relationship between infant feeding practices and the development of the gut microbiota

A thesis submitted to the Graduate School of the University of Cincinnati in partial fulfillment of the requirements for the degree of Master of Science

In the Department of Nutritional Sciences of the College of Allied Health

by

Shannon Conrey
B.A. University of Michigan

July, 2017

Committee: Abigail Peairs, PhD (Chair)
Ardythe Morrow, PhD
Nicholas Ollberding, PhD
Abstract

The microbial colonies of the human gut play an integral role in immune function, nutrient digestion and absorption, protection against pathogens, and biomolecule synthesis. Many factors, particularly birth method and duration and intensity of breastfeeding, have been shown to influence the microbiota in infancy. Correlations between dietary intake and bacterial phenotypes have been found in older children and adults, but little is known of the relationship between dietary patterns and the development of the microbiota in infancy. Subjects were drawn from the Cincinnati arm of the Global Exploration of Human Milk Study. Multivariable linear regression compared 16S rRNA-sequenced fecal samples obtained at 2, 4, 13, 26, 52, and 104 weeks to weekly 24-hour food recalls, including breastfeeding status and intensity and quantified dietary diversity, to explore differences in microbial abundance, diversity, and specific bacterial taxa. Significant predictive value was found between breastfeeding cessation and bacterial abundance at 52 weeks, breastfeeding intensity and bacterial abundance and diversity at 26 and 52 weeks, and dietary and bacterial diversity at 26 weeks. Results suggest that early high intensity breastfeeding may predict similar patterns as exclusive breastfeeding, as well as lower bacterial abundance and diversity in infancy, and may attenuate changes due to birth method by one month of age. Taxonomic patterns were consistent with extant research, with higher ratios of Actinobacteria and Proteobacteria in younger, highly breastfed infants, increased abundance of Firmicutes associated with breastfeeding cessation and consumption of a mixed diet, and families associated with the metabolism of human milk declining as breastfeeding waned. Dietary diversity per se was not found to significantly predict bacterial composition, but future research with specific dietary intake is needed to clarify relationships. As the microbial profile by 2 years of age is roughly similar that of adulthood, and dysfunctional microbiota is associated with a host of inflammatory, autoimmune, and atopic diseases, establishing relationships between feeding patterns and the development of microbial phenotypes during infancy could provide clues to dietary interventions targeted to alter microbial composition.
Table of Contents

I. Introduction ........................................................................................................................................... 1

II. Review of the Literature ....................................................................................................................... 1
   A. Perinatal and Early Microbial Colonization of the Infant Gut .......................................................... 1
   B. C-section Reduces Likelihood of Exclusive Breastfeeding .............................................................. 3
   C. Human Milk as a Pre- and Pro-biotic Food ....................................................................................... 4
   D. Increased Microbial Diversity with Introduction of Complementary Foods ..................................... 6
   E. Definition and Role of Dietary Diversity in Infant health ................................................................. 7
   F. The Importance of Bacterial Richness and Diversity ...................................................................... 9
   G. Dietary Diversity as a Contributor to Microbial Diversity ............................................................... 10
   H. Importance of Understanding Factors Contributing to Microbial Development ....................... 11

III. Methods ................................................................................................................................................ 13
   A. Study Population ............................................................................................................................... 13
   B. Demographic and Dietary Data Collection ..................................................................................... 14
   C. Fecal Sample Collection and Sequencing ....................................................................................... 15
   D. Sample Qualification ....................................................................................................................... 16

IV. Statistical Analysis ............................................................................................................................... 16
   A. Demographic Data Calculation ....................................................................................................... 16
   B. Alpha Diversity Calculations and Measures .................................................................................... 17
   C. Regression Models ........................................................................................................................... 17
   D. Beta Diversity Calculations and Measures ..................................................................................... 18
   E. Statistical Software ........................................................................................................................... 19

V. Results ................................................................................................................................................... 19
   A. Participant Characteristics ................................................................................................................. 19
   B. Breastfeeding by Study Week .......................................................................................................... 20
   C. Dietary Diversity by Study Week ..................................................................................................... 21
   D. Dietary Diversity by Breastfeeding Intensity .................................................................................. 22
   E. Bacterial Richness (Observed Sequence Variants) .......................................................................... 22
   F. Bacterial Diversity (Shannon Index) ................................................................................................. 24
   G. Relative Abundance .......................................................................................................................... 27
   H. Beta Diversity: Bray-Curtis Dissimilarity and Ordination ............................................................... 31

VI. Discussion ............................................................................................................................................ 34

VII. Conclusion ........................................................................................................................................... 53

VIII. Bibliography ....................................................................................................................................... 56
List of Tables and Figures

Table 1: Demographic Data of Qualifying Subjects .............................................................. 20
Table 2: Breastfeeding Status by Study Interval ................................................................. 21
Table 3: Subject mMDD by Study Interval ........................................................................... 21
Table 4: Bacterial Richness (Observed Sequence Variants) .................................................. 24
Table 5: Bacterial Diversity (Shannon Index) ...................................................................... 26

Figure 1: Breastfeeding Intensity Predicts Lower Dietary Diversity ..................................... 22
Figure 2: Bacterial Richness by Breastfeeding Status, unadjusted ...................................... 23
Figure 3: Bacterial Diversity by Breastfeeding Status, unadjusted ...................................... 25
Figure 4: Phyla-Level Relative Abundance by Breastfeeding Status, All Intervals .............. 27
Figure 5: Family-Level Relative Abundance by Breastfeeding Status, Week 2 ................. 28
Figure 6: Family-Level Relative Abundance by Breastfeeding Status, Week 4 ................. 29
Figure 7: Phyla-Level Relative Abundance by Breastfeeding Status, Weeks 26-104 ........... 30
Figure 8: Family-Level Relative Abundance by Breastfeeding Status, All Intervals .......... 33
Figure 9: Ordination Plots by Study Interval, Breastfeeding Status .................................... 34
I. Introduction

The human body hosts over 100 trillion bacteria and an estimated one quadrillion viruses representing taxa originating from a wide variety of environmental and dietary sources (1). The largest colonies of bacteria, collectively called the microbiota, reside in the human gut. Evidence indicates that the gut microbiota play an integral role in immune function, nutrient digestion and absorption, protection against pathogens, and biomolecule synthesis (1). Research has been published on the impact of breastfeeding (2-4) and the timing and type of complementary foods (5) on the colonization of the infant gut microbiota. Associations have been found between types of foods consumed and microbiota composition in children (6), and between dietary diversity and specific microbial profiles in adults (7). However, little is known of the relationship between dietary diversity and the establishment of bacterial colonies during the infant’s transition to complementary foods in the first year of life. An examination of the factors that influence the development of the infant’s microbiota, including breastfeeding status and intensity and the diversity of the complementary diet, will be the focus of this study.

II. Review of the Literature

A. Perinatal and Early Microbial Colonization of the Infant Gut

Although originally believed to be sterile in utero, meconium analysis has shown that the prenatal infant gut is minimally colonized by facultative anaerobes such as Staphylococcus, Enterobacteria, Lactobacillus, and Bifidobacteria (8). However, the neonate receives its first major exposure to microbial colonies either during passage through the birth canal, or from environmental and skin contact during a C-section,
making birth method an important contributor to early bacterial colonization (9). Infants born by C-section exhibit different microbial profiles, especially during the first year, than those born vaginally (2, 4, 9). Studies of factors influencing infant microbiota have found that birth method is a significant contributor to bacterial colonization across multiple body habitats of the newborn (9), and a significant predictor of gut bacterial profile at four months of age (4). Neonates born vaginally exhibit profiles more similar to those found in the mother’s vagina, with higher rates of Bacteroidales, Clostridiales, Enterobacteriales, and Bifidobacteriales taxa than those born by C-section (10). A study of the patterns of colonization by birth method found that the microbiota of infants born by C-section were initially colonized with taxa associated with the mother’s skin and the hospital environment, rather than vaginal or fecal-associated taxa (9). These differences are more pronounced in elective rather than acute C-section, as acute C-section is likely to occur after membrane rupture and include some exposure to the mother’s vaginal and fecal microbiota (11).

In addition to reduced exposure to vaginal flora, the routine administration of antibiotics to the mother prior to and after a C-section influences the infant’s microbiota. Maternal antibiotic use prenatally, administered during a C-section, or taken during breastfeeding, has been associated with a disruption of bacterial flora (12). Maternal pre-natal antibiotic use was a significant predictor of microbial composition at six weeks (13) and was associated with decreased richness of *Bifidobacterium* and *B. fragilis* at one month of age (14). Another study found that early antibiotic use predicted delayed microbial maturation in months 6-12, but not after one year of age. This effect, however, was smaller than the effect due to either delivery mode or age (10).
B. C-section Reduces Likelihood of Exclusive Breastfeeding

Due to the constraints of the operating theater, C-section reduces the opportunity for immediate skin-to-skin contact (SSC) between the mother and infant after birth. Immediate SSC has been shown to shorten the time to breastfeeding initiation, increase duration of breastfeeding exclusivity, and increase odds of breastfeeding to six months of age (15). The WHO includes immediate SSC to promote prompt (within 60 minutes) initiation of breastfeeding in its Baby Friendly Hospital Initiative “Ten Steps for Successful Breastfeeding” (16). An electronic medical records review of 60 births (30 vaginal births, 30 C-section births) in a Texas birthing hospital prior to an intervention designed to improve SSC post C-Section found that 93% (28/30) of vaginally born infants experienced immediate SSC compared to 23% (7/30) of C-section infants. Following the intervention, the vaginal birth SSC rate remained constant at 93% (139/150) while the SSC rate for C-section improved to 77% (116/150), still significantly less than the vaginal SSC rate ($p<0.001$) (17).

The relationship between reduced SSC and initiation of breastfeeding has been explored in C-section births. A prospective cohort study of 19 California hospitals ($n=21,842$) examining the relationship between immediate SSC after a C-section to breastfeeding initiation found a dose-response relationship between the duration of skin to skin contact and breastfeeding exclusivity in the hospital, ranging from 15 minutes of contact (OR 1.376, 95%CI 1.189-1.593) to > 1 hour (OR 3.145, 95%CI 2.905-3.405) when compared to no contact immediately following surgery (15).

Other evidence supports a reduction in breastfeeding initiation and duration with C-section. A subsection ($n=100,000$) of a large retrospective cohort study of C-section
as a risk factor for pediatric IBD in Denmark found that, compared with mothers who delivered vaginally, C-section mothers were less likely to initiate breastfeeding (4% vs. 2%), more likely not to breastfeed exclusively (10% vs. 4%), less likely to exclusively breastfeed until age five months (27% vs 29%) and more likely to cease breastfeeding altogether prior to five months of age (29% vs. 23%), while controlling for the IBD status of the mother (18). This study did not, however, examine the reasons for reduced breastfeeding with C-section.

Timely initiation of breastfeeding is important, as early initiation of breastfeeding has been shown to be a positive predictor for exclusive breastfeeding during the hospital stay (19), at six weeks of age (20), and at six months of age (21), while later initiation has been found to be a positive predictor of cessation of breastfeeding prior to six months (19). Exclusive early breastfeeding has been shown to both attenuate the microbial changes associated with C-section (22) and to be the most significant predictor of microbial composition in infancy and early childhood (2). Thus, decreasing the likelihood of successful early exclusive breastfeeding further increases birth method’s impact on microbial colonization.

C. Human Milk as a Pre- and Pro-biotic Food

The timely initiation and continuation of exclusive breastfeeding is important for many reasons, including human milk’s integral role in the establishment of bacterial colonies. Far from sterile, human milk is an extremely complex fluid, containing a wide variety of bacteria and viruses as well as macro and micronutrients (23). Both colostrum and human milk provide probiotic bacteria for the colonization of the infant’s gut (24) with an average exposure of approximately $10^5$ to $10^7$ bacteria per 800 mL (25). Despite
enormous variation between mothers in bacterial profiles, studies have shown that a “universal core” of nine genera is present, in differing ratios, in human milk regardless of region, diet, or health status and form a basic human milk bacterial phenotype in the growing infant’s developing gut microbiota (24). These bacteria are major early colonizers of the infant’s gut, as demonstrated by a study of ten breastfeeding mother-infant pairs which identified gut-associated genera in all samples of the mothers’ milk, representing between 70-88% of the genera identified in the infants’ feces (26).

In addition to its probiotic qualities, human milk contains a large percentage of human milk oligosaccharides (HMO), complex structures unique to humans of between 3-32 monomers in size. Although not digestible by the infant, HMO represent approximately 1/3 of the solid components of human milk, speaking to the important role they play (23). The structure of HMO includes a lactose core at the reducing end, elongated by N–acetyllactosamine units and diversified by the addition of fucose and/or sialic acid residues at the terminal ends, resulting in structures of intense diversity in size, sequence, and charge (27). To date over 200 distinct molecular species have been identified, although it is estimated that a few of the most abundant species account for over 80% of the total abundance of HMO in human milk (27). Studies have shown that there is enormous variation, between mothers and even within feedings, in HMO composition and abundance (27).

Functionally, HMO affect bacterial composition in two distinct ways. HMO have been shown to act as “decoys” for the pathogen receptors on the infant’s intestinal surface, thereby offering a protective effect from colonization by harmful pathogens (23). Secondly, HMO have a prebiotic quality and encourage the growth of beneficial
bacteria, especially *Bifidobacteria* (phylum Actinobacteria) and *Bacteroides* (phylum Bacteroidetes), two genera correlated with the production of short chain fatty acids, which alter the pH of the gut and have other beneficial, systemic properties (28).

Due to the pre- and pro-biotic qualities of human milk, breastfeeding duration and intensity have been found to have a significant impact on establishment of the gut microbiota (4, 9, 10). A meta-analysis examining the link between birth method and microbiota development found that differences in microbiota due to birth method were attenuated by feeding practices, especially early exclusive breastfeeding (minimum of three months), by one year of age (22). Bäckhed et al (2015) did not find the microbiota of exclusively breastfed and partially breastfed vaginally-delivered infants to be statistically dissimilar at one week, but did find that by four months, infants who were partially breastfed had more “mature” microbial profiles than those who were exclusively breastfed, while there were clear, significant differences between the microbiota of exclusively- and not breastfed infants at the same periods (29).

D. Increased Microbial Diversity with Introduction of Complementary Foods

Gut colonization progresses through infancy, with an explosion of microbial diversity commensurate with complementary food feeding around six months of age and reaching mature ratios at approximately two to three years (30). The introduction of complementary foods has been shown to be the period of greatest change in bacterial composition in infants (31). However, the impact of human milk on microbial colonization continues after the introduction of a mixed diet. Bergstrom et al (2014) examined the microbiota of infants after the introduction of complementary foods (from 9 months to 3 years) and found that the most significant predictive variable in infant
microbiota colonization at all time-points was the timing of cessation of breastfeeding (2). Fallani et al (2011) reported that early breast feeding intensity, categorized as exclusive, any, or no breastfeeding, was a significant determinant of microbiota profiles four weeks after introduction of complementary foods (5) while Azad et al (2014) found bacterial communities differed significantly based on breastfeeding status both at three months (prior to complementary food introduction), and at one year (after complementary foods had been established), while controlling for birth method and antibiotic use (32).

Breastfeeding can, however, influence both the timing and the diversity of the complementary diet. While breastfeeding is associated with a willingness to try novel foods in infancy and increased dietary diversity in childhood (33), high intensity breastfeeding has been correlated to reduced dietary diversity and nutrient intake before age one in South African infants (34). In a prior analysis of the Cincinnati cohort (n=120) of the Global Exploration of Human Milk (GEHM) study, the parent study from which data for this study was drawn, breastfeeding intensity > 50% was associated with a higher risk for low dietary diversity at one year (35).

E. Definition and Role of Dietary Diversity in Infant health

A diverse diet is recognized as important to infant and child health. The WHO included meeting or exceeding its Minimum Dietary Diversity (MDD) recommendation in its eight core indicators for adequate feeding practices for infants and young children (36). MDD is defined as a complementary diet consisting of a minimum of four of seven food categories, defined as 1.) Grains, roots, and tubers, 2.) Legumes and nuts, 3.) Dairy foods (milk, yogurt, cheese), 4.) Flesh foods (meat, fish, poultry, and liver/organ
meats), 5.) Eggs, 6.) Vitamin-A rich fruits and vegetables, and 7.) Other fruits and vegetables, consumed on the previous day (36).

The WHO’s MDD recommendation is based on the association of dietary diversity with increased diet quality and the likelihood that, in consuming a minimum of four of these seven defined groups, the child was likely to have consumed foods from the animal-sourced proteins, fruits and vegetables, and staple grains food groups (36). Achieving MDD, particularly in the developing world, increases the chances that children achieve adequate intake of the macro and micronutrients, improving health and preventing stunting and deficiency diseases (36).

Overall dietary diversity has long been recognized as a predictor for improved diet quality, especially in children. A varied diet at weaning has been associated with higher diet quality in later childhood, including increased acceptance of vegetables (37). Faber et al (2014) found that meeting the WHO’s MDD recommendation was associated with higher nutrient density in the diet of young children in South Africa (34) and the intake of a diverse diet has been positively associated with increased height for age in young children in 11 developing countries in Africa and Asia (38).

Consuming a high quality diet, especially high in fiber and low in saturated fat content, has been implicated in microbial health as well. In observational studies, dietary patterns have been associated with specific bacterial phenotypes, with high fiber/low fat diets associated with higher species abundance and diversity and the decreased fiber/high fat pattern associated with the Western diet associated with decreased microbial richness and diversity (39). The microbiota is particularly sensitive to dietary changes, as drastic changes have been shown to affect bacterial health,
richness, and diversity in as little as 24 hours (40). Few intervention studies of this effect have been completed on humans, but an intervention using human-microbiota colonized mice, switched to a low-fiber diet, showed significant changes in the microbial composition of the gut compared to mice that maintained a high-fiber diet for seven weeks and remained distinctive from the control mice after returning to a high fiber diet for six weeks (41).

F. The Importance of Bacterial Richness and Diversity

Although what defines a “healthy” microbial profile remains a subject of debate, a functional gut microbiota is integral in aiding in digestion and absorption of nutrients, vitamin synthesis, and acts as a first-line defense for opportunistic infections through short chain fatty acid production and competitive exclusion (1). Dysfunctional gut microbiota, known as dysbiosis, has been linked to a variety of pediatric and adult disorders, including asthma, wheezing, atopic dermatitis, inflammatory bowel disease, food allergies, and obesity (42).

Generally, in older children and adults, a healthy microbiota is robust in both abundance and diversity, with higher Bacteroidetes and lower Firmicutes ratios. Research has shown that a diet rich in fruits, vegetables, and fiber correlates to a highly diverse gut bacteria with a higher Bacteroidetes to Firmicutes ratio, with the converse true in the case of a higher fat, lower fiber dietary pattern (6, 39). Evidence that low Bacteroidetes to Firmicutes ratios is implicated in increased risk for obesity, metabolic syndrome, and Type II Diabetes Mellitus has been reported (43-45), however a recent meta-analysis pooling these results did not find a statistically significant link between measures of alpha or beta diversity and obesity (46). Other major phyla commonly
found, albeit in much lower ratios, in the adult gut include Actinobacteria, which are involved in short chain fatty acid production as well as protection against gut pathogens, and Proteobacteria, which is highly correlated with dysbiosis and malnutrition (47).

In infants, however, these patterns are variable at best. Highly breastfed infants exhibit microbiota typically low in abundance and diversity, with heavy representation in the protein-fermenting Proteobacteria, the lactic acid-producing Actinobacteria, especially *Bifidobacteria*, and species of Bacteroidetes and Firmicutes typically not found in abundance in adult guts (26). The composition of each mother’s milk is unique to her, and her infant’s colonies are uniquely suited to metabolize it, with bacteria specializing in hydrolyzing bonds found in human milk proteins and taxa which ferment the unique composition of HMO found in their mother’s milk (24).

The introduction of alternative sources of nourishment (formula and/or complementary foods) “matures” the microbiota, gradually increasing bacterial diversity and abundance and reducing the species associated with human milk metabolism (29). Ratios of colonization of the infant microbiota by the end of the first year is broadly similar to the composition of microbes they will maintain through adulthood (31), although bacterial richness by one year is only approximately 1/6 that of an adult. The rapid expansion in both richness and diversity during the transitional period of weaning and the introduction of a mixed diet results in a profile resembling approximately 2/3 the adult profile by age two to three (48).

G. Dietary Diversity as a Contributor to Microbial Diversity

While dietary diversity has been convincingly linked to dietary quality, and a high-quality diet is implicated in microbiota health, the direct effect of a diverse diet on the
microbiota has been less thoroughly explored, especially in infants. The microbiota of the infant, prior to the introduction of complementary foods, is difficult to characterize due to its instability in response to influencing agents. The landmark Burkina Faso study examined the relationship between diet and microbial composition and found associations between types of foods consumed, particularly fiber content, and microbiota composition in young children in rural Africa compared to urban Italy, but only included children older than two years and did not explore dietary diversity per se (6). Laursen et al (2016) explored the composition of the complementary diet of nine and 18 month old infants of healthy and obese mothers, finding that high protein and high fiber foods significantly affected microbial composition, alpha and beta diversity, and bacterial richness. Laursen and colleagues, however, only examined two time points, 9 and 18 months, both after complementary foods were established and again, did not look specifically at the diversity of the diet as a factor (49).

Studies focused on the relationship between dietary diversity and the microbiota have centered on adults. Salazar et al. (2014) correlated dietary diversity with increased bacterial richness and diversity and studies on Finnish adult monozygotic twins indicated that similar intake patterns was predictive of similar bacterial profiles (50). Non-human studies have linked dietary diversity to bacterial diversity and abundance in mammal (51), fish (52), and insect (53) models.

H. Importance of Understanding Factors Contributing to Microbial Development

The importance of understanding the factors that contribute to the early establishment of a robust and functional bacterial profile cannot be overstated. The period between the introduction of complementary foods and two years of age is
integral for establishing a microbiota framework that is maintained through adulthood. In adults, specific microbial profiles have been correlated to a range of metabolic, autoimmune, and inflammatory diseases and there is evidence that disorders from obesity to bipolar disorder may have a gut microbiota connection (54). Despite its important role in so many functions of maintaining human health, what is known of the very early development of the gut microbiota is limited.

It has been shown that microbial development is highly correlated with age, with heterogeneous colonization early and a more or less conserved vector of homogenous colonization by age six (55, 56). It has also been established that early breastfeeding intensity is a significant predictor of microbial composition both before and after the introduction of complementary foods (4, 5) and continues to be a significant predictor up to three years of age (2). It is well established that the introduction of complementary foods is associated with significant changes in bacterial composition in infants (31) and that microbial diversity expands commensurate with complementary food feeding occurring around six months of age (30).

While substantial research has been published on the effect of breastfeeding, both duration and intensity, on microbial development, most have considered it in the context of breastfeeding vs formula feeding or breastfeeding as a covariate with birth method and/or antibiotic use. No studies to our knowledge have considered breastfeeding duration and intensity in the context of dietary diversity in an attempt to clarify its relationship to microbial composition. And, although dietary diversity has been established as a significant contributor to microbial diversity in both adult and animal
models (7, 52), few, if any, studies have quantified it and traced its influence on the infant microbiota.

This study, using data obtained from the Cincinnati arm of the GEHM study, offers a unique opportunity to examine highly breastfed infants for relationships between consumption of human milk, dietary diversity, and the development of the gut microbiota. Using GEHM’s intensive, quantified dietary data and fecal samples collected at regular intervals in the first year, with additional samples collected at two years, will allow for exploration of the role of dietary diversity and breastfeeding intensity and duration in the development of the gut microbiota in terms of richness, diversity, and taxonomic composition during infancy and early childhood.

III. Methods

A. Study Population

Subjects were drawn from the Cincinnati cohort of the Global Exploration of Human Milk study (GEHM), a prospective cohort of urban breastfeeding mother-infant pairs (n=120), recruited from January, 2007 through December, 2008. Participants were recruited through a single, large, birth hospital and through community advertisement. Eligible mothers were between 18-49 years and intending to breastfeed a minimum of 75% of feedings for the first three months. Infants were singleton, full term, with a birth weight > 2500 g, and had no medical conditions that would prevent breastfeeding. Eligibility was initially assessed at two weeks post-partum and reassessed at four weeks post-partum, with mothers not achieving the 75% breastfeeding requirement excluded. Mothers provided written informed consent and the study was approved by the Institutional Review Board of Cincinnati Children’s Hospital Medical Center.
B. Demographic and Dietary Data Collection

GEHM followed qualifying mother-infant pairs for two years, with eligibility and baseline questionnaires at two weeks of age, and five in-person visits at weeks 4, 13, 26, 52, and 104. Duration and intensity of breastfeeding were validated by weekly 24 hour food frequency recalls via telephone surveillance during the first year, with breastfeeding cessation date collected at the 104 week visit for those continuing to breastfeed into the second year. Maternal recall provided pregnancy details, including pre-pregnancy weight, gestational weight gain, gestational diabetes, and any medications administered during pregnancy. Birth method, gestational age, birth weight, and length of hospital stay were obtained from the medical record (35).

Breastfeeding intensity was defined as the number of times the infant had been fed human milk divided by the total number of items fed, multiplied by 100, with 100% defined as exclusive breastfeeding (EBF), any amount of breastfeeding <100% defined as partial breastfeeding (PBF), and human milk comprising 0% of feeds defined as no breastfeeding (NBF). Breastfeeding cessation was defined as the infant’s age when the mother reported breastfeeding cessation, cross-validated with the last 24-hour food frequency recall reporting human milk consumption (35).

Dietary intake was measured using a 24-hour food frequency recall recording the number of times in the previous 24 hours the infant had consumed food or liquids, classified into 21 categories, but did not assess portion size or nutritional value of intake. Using a modified World Health Organization (WHO) Minimum Dietary Diversity (mMDD) Metric (35), infant intake was categorized into six food groups [1.] Grains, roots, and tubers, 2.) Legumes and nuts, 3.) Dairy products, (milk, cheese, and yogurt),
4.) Flesh foods (meat, fish, poultry, liver, organ meats), 5.) Eggs, and 6.) Fruits and vegetables. An mMDD score was assigned at each interval representing the highest number of mMDD food categories the infant consumed in any of the weekly surveys during that interval. Human milk and formula were not included in the mMDD score (35).

C. Fecal Sample Collection and Sequencing

Mothers were asked to collect, seal in plastic, and refrigerate a soiled diaper on the previous or same day for collection at the scheduled clinic visit. Fecal specimens were collected by scraping the sample from the diaper and vortexing in a phosphate buffered saline solution. Samples were then aliquoted and stored at -80°C. All samples were linked to a unique Subject ID by digital barcode.

DNA was isolated by first thawing, then centrifuging the stool samples for 10 minutes at 4000xg to pellet the fecal matter. Supernatant was removed and 100 µL of TE buffer with lysozyme and proteinase K was added to 0.24 g of thawed stool. The samples were then vortexed for 10 minutes. A total of 1.2 mL of Buffer RLT (Qiagen) with β-mercaptoethanol was added to the sample. The sample was then transferred to bead beating tubes containing 0.3 grams of 0.1-mm glass beads and homogenized for three minutes in a bead beater. Samples were centrifuged at 4000xg for 5 minutes to pellet the debris. Samples were then transferred to a clean microcentrifuge tube and spun an additional 2 minutes to remove remaining debris. Supernatant was applied to a Qiagen AllPrep DNA/RNA Mini Kit (Qiagen, Valencia, CA, USA) column and DNA was isolated following the manufacturer directions. The V4 region of the 16S rRNA gene was amplified using the protocol and primers were prepared on the Illumina MiSeq platform (Illumina, San Diego, CA, USA).
Read processing and error correction was performed using the open-source software package DADA2 (57) in R. Read processing included primer trimming using cutadapt (58) followed by quality filtering and length trimming (fastqPairedFilter), dereplication (drepFastq), error correction (57), merging of paired-end reads (mergePairs), and chimera removal (removeBimeraDenovo). Taxonomic classification was implemented using the assignSpecies and assignTaxonomy functions and DADA2-formatted RDP reference databases allowing for exact matching. A de-novo phylogenetic tree was constructed in R using the phangorn package (59) to construct a neighbor-joining tree and then a GTR+G+I maximum likelihood tree was fit using the neighbor-joining tree as the starting point (60).

D. Sample Qualification

Subjects included in this analysis provided a minimum of one normal/non-diarrheal stool sample, containing a minimum of 5000 sequence reads, which was collected during an interval in which they also provided dietary data. As fecal samples were not taken from every subject at each sample interval, the data was considered as cross sectional data at each interval and analyzed accordingly.

IV. Statistical Analysis

A. Demographic Data Calculation

Demographic data was calculated for all subjects (n=88) included in this analysis and reported as mean ± standard deviation, or as a proportion, dependent on the nature of the data. As each interval contained a unique subset of the whole, demographic data was also calculated at each interval and reported as mean ± standard deviation, or as a
proportion, dependent on the nature of the data. Due to the unique composition of each interval, time points were treated as sequential, cross-sectional studies.

B. Alpha Diversity Calculations and Measures

Two measures of alpha diversity were calculated. Bacterial richness was calculated using the number of observed sequence variants (SV), after subsampling to 5k reads per sample. Observed SV is defined as the count of identified SV present in the sample. Observed SV counts each SV, regardless of the frequency with which it appears, so reflects an estimate of the total number of unique SV identified. Bacterial diversity was calculated using the Shannon Index. The Shannon Index is a weighted α diversity measure in which the proportion of each SV relative to the total number of SV is calculated, and then multiplied by the natural logarithm of this proportion. The resulting product is summed across SV, and multiplied by -1, with a higher Shannon Index score indicating greater bacterial diversity. As this measure adjusts for the proportion of taxa, it is influenced by both the number of SV and their evenness of distribution.

C. Regression Models

Multivariable linear regression assessed the relationship between Observed SV and Shannon Index and breastfeeding status, intensity, and quantified dietary diversity (mMDD score), unadjusted and adjusted by the infant’s sex, birth method (vaginal or C-section), gestational age (in weeks), length of hospital stay at birth (in days), race (White, Black, Asian, Native American/Alaska Native, Native Hawaiian/Pacific Islander, or “Other”), ethnicity (Hispanic, non-Hispanic, or “Unknown”), and the presence of other children in the household (“Only” or “Not Only”). As no income information was available
in the data, WIC enrollment at birth was used as a proxy variable for socioeconomic status (SES). Each model was run with Observed SV or Shannon Diversity as the outcome variable and were designed as follows: Model 1: breastfeeding status (Exclusive (EBF: human milk comprised 100% of feeds), Partial (PBF: subject received human milk, comprising < 100% of feeds) or No (NBF: human milk comprised 0% of feeds) breastfeeding) ± covariates, Model 2: breastfeeding intensity (defined as percentage of feeds consisting of human milk) ± covariates, and Model 3: dietary diversity score (mMDD: highest number of modified Minimum Dietary Diversity categories subject consumed during interval) ± covariates. Breastfeeding intensity was previously negatively correlated to dietary diversity in the GEHM cohort (35). To test for this relationship in the present study, simple linear regression assessed the relationship between breastfeeding intensity and dietary diversity at the intervals with most subjects consuming a complementary diet and that contained dietary diversity data (Weeks 26 and 52). To control for possible confounding by breastfeeding intensity, Model 4 added breastfeeding intensity to mMDD as an additional covariate. Results were reported as β coefficient, with a significance level of p<0.05 and trends identified at p<0.10.

D. Beta Diversity Calculations and Measures

Beta diversity was calculated by first rarifying the data to 5000 reads per sample and comparing relative abundance by phyla and/or family. Bray-Curtis dissimilarity was calculated and compared using multivariate ANOVA to determine clustering by group, based on breastfeeding status and time. Ordination plots were created using nonmetric multidimensional scaling (NMDS) to compare bacterial richness by study week and
breastfeeding status. Comparisons of dispersion were made using the *permutest* function in the *vegan* package for R (61).

E. Statistical Software

All analysis was performed using R statistical software, version 3.3.3 (62), *vegan* (61) and *phyloseq* (63) packages. Additional graphing functions were performed using Microsoft Excel 2013 for Windows 7 Professional (64).

V. Results

A. Participant Characteristics

A total of 194 samples from 88 subjects met criteria for this analysis (Table 1). The mean age of the mothers of the qualifying subjects was 31.5 ±5.8 years. All but one of the mothers had achieved at a minimum of a high school education, and the majority (70.5%) were college graduates (data not shown). The infants were mostly delivered vaginally (76.1%), majority female (58.0%), and non-Hispanic (94.3%), with 80.7% identifying as White, 12.5% identifying as Black, zero identifying as Asian, Native American/Alaskan, or Native Hawaiian/Pacific Islander, and 6.8% identifying as “Other.” One quarter of the children were the only children in the household, while on average 2.4 ±1.1 children lived in the home. Enrollment in the Federal Women, Infants, and Children (WIC) supplemental food program included 22.7% of the infants.

As samples were not taken from every subject at every interval and the data was analyzed as sequential cross-sectional studies, demographic data was also calculated at each study interval. The demographics of each interval were similar to the
demographics of all of the subjects, with the exceptions of proportion of vaginal births in Week 2 (55%) and Week 104 (100%).

**Table 1: Demographic Data of Qualifying Subjects**

<table>
<thead>
<tr>
<th>Week #</th>
<th>2</th>
<th>4</th>
<th>13</th>
<th>26</th>
<th>52</th>
<th>104</th>
<th>All Qualifying</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>μ/n</td>
<td>SD/</td>
<td>μ/n</td>
<td>SD/</td>
<td>μ/n</td>
<td>SD/</td>
<td>μ/n</td>
</tr>
<tr>
<td>Mother Age (at Delivery) Years</td>
<td>32.04 ±6.67</td>
<td>31.04 ±6.32</td>
<td>31.51 ±5.14</td>
<td>31.98 ±4.17</td>
<td>32.24 ±5.31</td>
<td>31.86 ±5.26</td>
<td>31.50 ±5.37</td>
</tr>
<tr>
<td>Delivery Type</td>
<td>Vaginal</td>
<td>16</td>
<td>0.55</td>
<td>23</td>
<td>0.66</td>
<td>20</td>
<td>0.65</td>
</tr>
<tr>
<td>Gestational Age</td>
<td>39.39 ±0.99</td>
<td>39.55 ±1.13</td>
<td>39.51 ±1.05</td>
<td>39.50 ±1.04</td>
<td>39.77 ±1.27</td>
<td>39.50 ±1.13</td>
<td></td>
</tr>
<tr>
<td>Birth Weight (kg)</td>
<td>3.48 ±0.40</td>
<td>3.45 ±0.35</td>
<td>3.53 ±0.41</td>
<td>3.54 ±0.40</td>
<td>3.59 ±0.44</td>
<td>3.42 ±0.47</td>
<td>3.54 ±1.05</td>
</tr>
<tr>
<td>Length of Hosp. stay (days)</td>
<td>2.55 ±1.24</td>
<td>2.42 ±1.31</td>
<td>2.32 ±1.25</td>
<td>2.25 ±1.25</td>
<td>2.00 ±0.80</td>
<td>1.67* ±0.59</td>
<td>2.13 ±1.05</td>
</tr>
<tr>
<td>Infant Sex</td>
<td>Female</td>
<td>17</td>
<td>0.59</td>
<td>21</td>
<td>0.60</td>
<td>17</td>
<td>0.55</td>
</tr>
<tr>
<td>Infant Race</td>
<td>White</td>
<td>24</td>
<td>0.83</td>
<td>25</td>
<td>0.71</td>
<td>25</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>4</td>
<td>0.14</td>
<td>7</td>
<td>0.20</td>
<td>3</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>1</td>
<td>0.03</td>
<td>3</td>
<td>0.09</td>
<td>3</td>
<td>0.10</td>
</tr>
<tr>
<td>Infant Ethnicity</td>
<td>Hispanic</td>
<td>1</td>
<td>0.03</td>
<td>3</td>
<td>0.09</td>
<td>1</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Non-Hispanic</td>
<td>27</td>
<td>0.93</td>
<td>30</td>
<td>0.86</td>
<td>29</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>1</td>
<td>0.03</td>
<td>2</td>
<td>0.06</td>
<td>1</td>
<td>0.03</td>
</tr>
<tr>
<td>Enrolled in WIC?</td>
<td>No</td>
<td>23</td>
<td>0.79</td>
<td>23</td>
<td>0.66</td>
<td>23</td>
<td>0.74</td>
</tr>
<tr>
<td>Only Child in household</td>
<td>Yes</td>
<td>8</td>
<td>0.28</td>
<td>10</td>
<td>0.29</td>
<td>9</td>
<td>0.29</td>
</tr>
<tr>
<td>Number of children in household</td>
<td>2.38 ±1.11</td>
<td>2.51 ±1.27</td>
<td>2.26 ±1.00</td>
<td>2.61 ±1.10</td>
<td>2.56 ±1.25</td>
<td>2.22 ±0.88</td>
<td>2.38 ±1.13</td>
</tr>
</tbody>
</table>

Results are reported as mean (μ) ± standard deviation (SD) or as the number of subjects/samples (n) and the proportion (p̂), as indicated by the data.

Demographic data of study subjects, by interval and all qualifying. A total of 194 samples from 88 subjects were included in this study. Qualifying samples contained a minimum of 5000 sequence variants from an interval in which the subject completed a 24-hour dietary recall. Each interval, then, contains a unique group of subjects, drawn from the Cincinnati arm of the Global Exploration of Human Milk study.

**B. Breastfeeding by Study Week**

Due to inclusion criteria mandating an intention to exclusively breastfeed for a minimum of three months, the data in terms of breastfeeding status is highly homogeneous during the first two intervals, with the majority (Week 2: 27/29, Week 4: 37/35).
30/35) EBF, and all those categorized as PBF exceeding the 75% intensity threshold (Week 2, both=92.3%, Week 4 all>85.7%)
(Table 2).

Although the first subject to cease breastfeeding appears in Week 13, all others were breastfeeding at a high intensity in that interval (27 EBF, 3 PBF> 83.3%). At 26 weeks, we see more separation of the data, with 5 subjects still EBF, the majority PBF (27/36), with an intensity range of 37.5-90.9%, and four NBF. No subjects were EBF by 52 Weeks, but 37/43 still partially breastfed (range: 8.3-63.2%). Four subjects were still PBF in Week 104, but breastfeeding at a low intensity (range: 3.3-33.3%).

C. Dietary Diversity by Study Week

Complementary foods were first introduced by Week 4 to two subjects and by Week 13 to five subjects (Table 3). Both subjects in Week 4 consumed foods from only one category. In Week 13, four of the five subjects consuming a mixed diet scored an mMDD of 1, while one subject consumed foods from two categories. Most subjects (29/35) had introduced complementary foods

<table>
<thead>
<tr>
<th>Table 2: Breastfeeding Status by Study Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interval #</td>
</tr>
<tr>
<td>Status</td>
</tr>
<tr>
<td>EBF</td>
</tr>
<tr>
<td>PBF</td>
</tr>
<tr>
<td>NBF</td>
</tr>
</tbody>
</table>

EBF=Exclusive breastfeeding (100% of feeds)
PBF=Partial breastfeeding (any breastfeeding < 100%)
NBF=No breastfeeding (0% of feeds)
n=number of subjects in that interval at breastfeeding status

<table>
<thead>
<tr>
<th>Table 3: Subject mMDD* by Study Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interval #</td>
</tr>
<tr>
<td>mMDD=0</td>
</tr>
<tr>
<td>mMDD=1</td>
</tr>
<tr>
<td>mMDD=2</td>
</tr>
<tr>
<td>mMDD=3</td>
</tr>
<tr>
<td>mMDD=4</td>
</tr>
<tr>
<td>mMDD=5</td>
</tr>
<tr>
<td>mMDD=6</td>
</tr>
</tbody>
</table>

*mMDD=modified Minimum Dietary Diversity Score defined as the highest number of modified WHO Minimum Dietary Diversity food categories consumed in the interval
n=the number of subjects achieving mMDD score in that interval

Number of subjects achieving mMDD scores by study interval. Although most subjects were consuming a complementary diet by Week 26, dietary diversity in this interval was low, with only 1 subject achieving the WHO’s recommended 4 or more food categories. By Week 52, most subjects were achieving WHO recommendations.
by Week 26, although only one subject met the WHO’s recommendation of ≥4 categories, and most (28/35) consumed ≤ 2 food categories, with a mean of 1.60 mMDD. By one year, all subjects were consuming a mixed diet, and the majority (31/39) had an mMDD ≥4, with five subjects achieving the top mMDD of 6 and a mean mMDD of 4.60. Dietary diversity data was not available for week 104.

D. Dietary Diversity by Breastfeeding Intensity

Breastfeeding intensity (percentage of feeds) was significantly predictive of lower dietary diversity in Weeks 26 and 52 (Figure 1), with higher percentage of breastfeeding predicting lower dietary diversity ($\beta=-3.99, p<0.001$). This relationship remained constant when examining the intervals individually, however only Week 52 remained statistically significant.

E. Bacterial Richness (Observed Sequence Variants)

In the unadjusted analysis (Figure 2), bacterial richness was significantly predicted by breastfeeding cessation only at 52 weeks, with cessation of breastfeeding predicting higher bacterial richness ($\beta=20.12, p=0.02$). However, when adjusted for birth method, gestational age, sex, race/ethnicity, length of hospital stay, WIC enrollment, and the presence of other children in the home, breastfeeding status no longer
significantly predicted bacterial richness at any interval, with a trend in higher richness with NBF only at Week 52 (β=19.92, p=0.06) (Model 1, Table 4).

Breastfeeding intensity (percentage of feeds of human milk) was not a significant predictor of bacterial richness in the unadjusted analysis at any time point, but a trend was detected at 52 weeks (β=-35.37, p=0.06), with higher intensity of breastfeeding predicting lower bacterial richness. When adjusted for the covariates (Model 2, Table 4), breastfeeding intensity significantly predicted bacterial richness at 26 (β=-39.59, p=0.004) and 52 (β=-44.03, p=0.04) weeks. At both time points, higher breastfeeding intensity predicted lower bacterial richness.

The mMDD score was not a significant predictor of bacterial richness at any time point (Model 3, Table 4). In Model 4, breastfeeding intensity was added as a covariate in the regression (Model 4, Table 4). In this model, mMDD was still not predictive of bacterial richness in any interval. However, in Week 26, this model explained more of the variance (adjusted r²=0.29) than either Model 2 (adjusted r²=0.19) or Model 3 (adjusted r²=-0.04).
In the covariate analysis of bacterial richness, infant sex and type of delivery were significant predictors in Week 2 in both Models 1 and 2, with male sex predicting higher (β=34.89, p=0.0079) and vaginal delivery predicting lower (β=-24.18, p=0.030) bacterial richness in both analyses. When used as a covariate in the analysis of mMDD (Model 4), breastfeeding intensity was a significant predictor of Observed SV at 26 weeks (β=-46.34, p=0.003) and trending toward significance at 52 weeks (β=-52.81, p=0.062). In both instances, higher percentage of breastfeeding predicted lower bacterial richness. No other covariates significantly predicted bacterial richness at any time point.

**Table 4: Bacterial Richness (Observed Sequence Variants)**

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Breastfeeding Status</th>
<th>Exclusive</th>
<th>Partial</th>
<th>None</th>
<th>Ref</th>
<th>16.62</th>
<th>NA</th>
<th>-19.64</th>
<th>Ref</th>
<th>-2.54</th>
<th>20.03</th>
<th>Ref</th>
<th>9.49</th>
<th>25.87</th>
<th>NA</th>
<th>19.92*</th>
<th>NA</th>
<th>40.98</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 2</td>
<td>Breastfeeding intensity</td>
<td>% Breastmilk</td>
<td>-21.05</td>
<td>166.81</td>
<td>-19.99</td>
<td>-39.59**</td>
<td>-44.03**</td>
<td>63.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 3</td>
<td>Dietary Diversity</td>
<td>mMDD#</td>
<td>6.55</td>
<td>-7.55</td>
<td>3.72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 4</td>
<td>Dietary Diversity</td>
<td>mMDD#</td>
<td>4.21</td>
<td>-4.803</td>
<td>-3.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Breastfeeding intensity</td>
<td>% Breastmilk</td>
<td>-16.66</td>
<td>-48.34**</td>
<td>-52.81*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<0.1, **p<0.05  #modified Minimum Dietary Diversity score  § Insufficient dietary diversity data available

Multivariable regression analysis of bacterial richness (measured by number of observed sequence variants) by breastfeeding status, breastfeeding intensity, quantified dietary diversity, and a combined model of breastfeeding intensity/quantified dietary diversity while controlling for infant sex, ethnicity, race, gestational age, WIC enrollment status, home crowding, birth method, and days hospitalized at birth.

**F. Bacterial Diversity (Shannon Index)**

In the unadjusted analysis (Figure 3), PBF was significantly predictive of Shannon Index score vs EBF at 13 Weeks (β=-0.63, p=0.04) and breastfeeding cessation was significantly predictive at 52 Weeks (β=0.514, p=0.02), with both intervals indicating that reduced breastfeeding predicted higher bacterial diversity.
When adjusted for the covariates birth method, gestational age, sex, race/ethnicity, length of hospital stay, WIC enrollment, and the presence of other children in the home, a trend in reduced breastfeeding predicting higher bacterial diversity was detected in Weeks 4 and 26 (Week 4, PBF vs EBF: $\beta=-0.59$, $p=0.098$, Week 26, NBF vs EBF: $\beta=0.82$, $p=0.099$) (Model 1, Table 5). Breastfeeding cessation was significantly predictive in Model 1 at 52 weeks ($\beta=0.51$, $p=0.04$) with no breastfeeding predicting higher bacterial diversity than partial breastfeeding. In Model 2, higher breastfeeding intensity predicted lower bacterial diversity at 26 weeks ($\beta=-1.28$, $p=0.004$) and 52 weeks ($\beta=-1.32$, $p=0.007$) with a trend toward significance in Week 4 ($\beta=5.68$, $p=0.08$) (Model 2, Table 5). The dietary diversity score (mMDD) significantly predicted Shannon Index only at 26 weeks, both in the unadjusted ($\beta=-0.238$, $p=0.009$) and the adjusted models ($\beta=-4.78$, $p=0.008$) (Model 3, Table 5).

The regression model examining mMDD while controlling for breastfeeding intensity (Model 4, Table 5) explained more of the variability in Shannon diversity than the previous model with mMDD alone in Weeks 26 (Model 4: adjusted $R^2=0.47$).
3: adjusted R²=0.018) and 52 (Model 4: adjusted R²=0.31, Model 3: adjusted R²=0.12).

At 26 weeks, Model 4 significantly predicted bacterial diversity by both breastfeeding intensity (β=-1.44, p=0.001) and mMDD score (β=-0.29, p=0.01), with both increased breastfeeding and dietary diversity predicting lower bacterial diversity.

In the analysis of the covariates, Shannon Index was significantly predicted in Week 4 by infant race “Other” and gestational age in Model 1 (race “Other:” β=0.92, p=0.045, gestational age: β=0.30, p=0.009) and Model 2 (race “Other:” β=0.95, p=0.04, gestational age: β=0.31, p=0.008), and at Week 52 in Model 4 (infant race “Other:” β=-1.03, p=0.03, gestational age: β=-0.20, p=0.03). No other covariates significantly predicted Shannon Diversity at any time point.

### Table 5: Bacterial Diversity (Shannon Index)

<table>
<thead>
<tr>
<th>Results reported as β coefficient</th>
<th>Wk 2$</th>
<th>Wk 4$</th>
<th>Wk 13</th>
<th>Wk 26</th>
<th>Wk 52</th>
<th>Wk 104$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breastfeeding status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exclusive</td>
<td>Ref -0.097</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partial</td>
<td>Ref -0.59*</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td></td>
<td></td>
<td>Ref 0.58</td>
<td>Ref -0.070</td>
<td>Ref 0.82*</td>
<td>NA</td>
</tr>
<tr>
<td><strong>Model 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breastfeeding Intensity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Breastmilk</td>
<td>1.26</td>
<td>5.68*</td>
<td>-0.91</td>
<td>-1.28**</td>
<td>-1.32**</td>
<td>2.20</td>
</tr>
<tr>
<td><strong>Model 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dietary Diversity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mMDD#</td>
<td>0.176</td>
<td>-4.78**</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dietary Diversity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mMDD#</td>
<td>0.054</td>
<td>-0.29**</td>
<td>-0.073</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breastfeeding Intensity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Breastmilk</td>
<td>-0.87</td>
<td>-1.44**</td>
<td>-1.55**</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<0.1, **p<0.05  #modified Minimum Dietary Diversity score  § Insufficient dietary diversity data available

Multivariable regression analysis of bacterial diversity (measured by Shannon Index) by breastfeeding status, breastfeeding intensity, quantified dietary diversity, and a combined model of breastfeeding intensity/quantified dietary diversity while controlling for infant sex, ethnicity, race, gestational age, WIC enrollment status, home crowding, birth method, and days hospitalized at birth. Results reported as β coefficient.
G. Relative Abundance

A total of 1102 distinct taxa, from nine phyla, were identified in the samples. Four phyla, Actinobacteria, Firmicutes, Proteobacteria, and Bacteroidetes, represented over 99.5% of all phyla at all time points (Figure 4). The groups when separated by breastfeeding status appear roughly similar in ratios during the first two intervals. Although no significant differences were found in bacterial richness or diversity at Weeks 2 or 4 in Models 1-3, an examination of the groups at the taxonomic family level revealed compositional differences both by interval and breastfeeding status.

![Relative abundance of phyla comparing bacterial profiles in each study week by breastfeeding status. Four phyla, Actinobacteria, Bacteroidetes, Firmicutes, and Proteobacteria, account for >99.5% of all taxa at all intervals.](image)
Of the Week 2 samples, 27 were EBF and 2 were PBF, but breastfeeding at a very high intensity (both 92.3%). However, the bacterial profiles at the family level exhibited some notable differences (Figure 4). The Week 2 EBF profiles were dominated by Bifidobacteriaceae (26.33%) (Actinobacteria) and Enterobacteriaceae (31.76%) (Proteobacteria), but registered a total of 56 families, including 12 families with ≥1% abundance and 44 families combined to 2.43% of the taxa. The Week 2 PBF group was less varied in its make-up, with a total of 33 families represented, including eight families ≥1%, dominated by Bifidobacteriaceae (41.98%) and Staphylococcaceae (38.67%) (Firmicutes), with the remaining 25 families combined for 5.90% of the taxa.

Analysis of the relative abundance by family at Week 4 (Figure 6) reveals more similarity in composition between the EBF and PBF groups, especially between the...
families representing ≥1% of taxa. The Week 4 profiles comparing these families are highly similar, with nearly identical representation of families ≥1% in relative abundance and parallel proportions. The EBF group exhibited more overall family diversity (EBF: 62, PBF: 25), but most of these families were only represented in trace amounts (<0.1%).

The first interval with representation in all three breastfeeding categories is Week 13, with the first subject to cease breastfeeding. Although breastfeeding status was the only significant predictor of bacterial diversity at this interval (PBF: $\beta=0.63$, $p=0.04$), and only in the unadjusted model, there were some notable differences in phylum level relative abundance among the three groups (data not shown). First, both breastfed groups had high ratios of Actinobacteria (EBF: 49.11%, PBF: 40.06%), while the NBF subject had greatly diminished proportions of this phyla (NBF: 8.05%). A similar pattern was seen in Firmicutes.
(EBF: 24.83%, PBF: 21.22%, NBF: 89.66%) and Proteobacteria (EBF: 21.14%, PBF: 20.95, NBF: 1.97%), with the PBF and the EBF ratios similar and the NBF subject radically different.

At 26 Weeks, the microbial profiles begin to mature and assume more adult-like ratios of phyla (Figure 7). At Week 26, migrations from high Actinobacteria and Proteobacteria ratios to higher Firmicutes ratios appear. The profiles at Weeks 52 and 104, exhibit a noticeable shift and similarity in composition, with additional decreases in Actinobacteria and Proteobacteria and further increases in Firmicutes.

<table>
<thead>
<tr>
<th>Week</th>
<th>EBF, n=5</th>
<th>PBF, n=27</th>
<th>PBF, n=37</th>
<th>PBF, n=37</th>
<th>NBF, n=4</th>
<th>NBF, n=8</th>
<th>NBF, n=14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actinobacteria</td>
<td>23.96%</td>
<td>62.82%</td>
<td>27.73%</td>
<td>9.95%</td>
<td>18.24%</td>
<td>5.68%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Bacteroidetes</td>
<td>7.04%</td>
<td>6.83%</td>
<td>9.60%</td>
<td>17.13%</td>
<td>8.53%</td>
<td>17.94%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Firmicutes</td>
<td>37.47%</td>
<td>23.68%</td>
<td>59.28%</td>
<td>69.40%</td>
<td>72.06%</td>
<td>74.08%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>31.52%</td>
<td>6.63%</td>
<td>1.64%</td>
<td>3.06%</td>
<td>0.26%</td>
<td>1.25%</td>
<td>0.00%</td>
</tr>
<tr>
<td>All others</td>
<td>0.01%</td>
<td>0.04%</td>
<td>1.75%</td>
<td>0.46%</td>
<td>0.91%</td>
<td>1.04%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

**Figure 7**: Taxonomic differences in phyla level composition as complementary foods are introduced, showing gradual maturation of infant microbiota from Week 26, with higher ratios of Actinobacteria and Proteobacteria and lower ratios of Bacteroidetes and Firmicutes, to the more adult-like and highly similar ratios of Weeks 52 and 104.
An examination at the family level from 2-104 weeks demonstrates these patterns, with the gradual diminishment of families of phyla Actinobacteria (Figure 7, shades of blue) and increases in families from Firmicutes (Figure 8, shades of green & brown) over the length of the study. Of particular interest are families specifically associated with breastfeeding, such as Lactobacillaceae and Bifidobacteriaceae. In these samples, Lactobacillaceae (phylum Firmicutes) were found in abundance ≥1% in Week 13 EBF (1.41%), Week 26 EBF (2.12%), and Week 26 PBF (1.94%), but then appear only as trace quantities in any sample that had ceased breastfeeding or in any of the Week 52 or 104 samples. Like the Lactobacillaceae, the Bifidobacteriaceae (phylum Actinobacteria) begin to wane with decreased intensity and cessation of breastfeeding with 61.58% and 24.67% in Week 26, 26.72% and 9.37% in Week 52, and only 16.68% and 5.32% in Week 104 of the PBF and NBF taxa, respectively.

Of the phyla Firmicutes, Staphylococcaceae appear in both of the Week 2 groups, but are not found in any group after Week 2 in abundance ≥1%. Another Firmicutes family to note is Lachnospiraceae. This family was found in abundance ≥1% in all intervals, was lower in breastfeeding groups in intervals 13-104, and exhibited a pattern of progressive increases by time in all categories. The exception to this pattern was the single Week 13 NBF subject, ID#181, whose profile was dominated by Lachnospiraceae at this interval (84.21%).

H. Beta Diversity: Bray-Curtis Dissimilarity and Ordination

Bray-Curtis dissimilarity was calculated between each pair of samples in each interval in terms of bacterial richness, with a score of 0 indicating identical samples and 1 indicating no similarity. Nonmetric Multidimensional Scaling (NMDS) was used to
generate an ordination, based on analysis of each possible pair in the interval (Figure 9). Clustering by time was confirmed with multivariate ANOVA, which found that the plots were significantly different between the intervals ($p=0.001$). However, a permutation test for homogeneity of multivariate dispersions (*permutest*) found significant heterogeneity in dispersion by comparing the average distance to the mean in each study week (Average distance to mean, Week 2: 0.62, Week 4: 0.56, Week 13: 0.50, Week 26: 0.47, Week 52: 0.54, Week 104: 0.54, $p=0.04$). Multivariate ANOVA assessed clustering within study weeks by breastfeeding status, finding significant clustering by breastfeeding status in Weeks 4 ($p=0.01$), 13 ($p=0.04$), 26 ($p=0.006$), and 52 ($p=0.03$) (*ellipses*, Figure 9). *Permutest* analysis indicated heterogeneity in dispersion by breastfeeding status only in Week 2 (average distance to the mean EBF: 0.61, PBF: 0.48, $p=0.001$).
Figure 8: All interval relative abundance by breastfeeding status at the taxonomic family level, shows the patterns in colonization as the infant ages and breastfeeding wanes. Families heavily represented in young, exclusively breastfed infants, such as Bifidobacteriaceae and Enterococcaceae, gradually decrease in prominence as families associated with mature profiles, such as Lachnospiraceae and Ruminococcaceae, increase.
VI. Discussion

Duration and intensity of breastfeeding are well established as early significant contributors to bacterial composition in infants and the composition of the diet, including
dietary diversity, has been associated with specific bacterial profiles in both children and adults. This analysis examined the relationship between feeding practices, including breastfeeding duration and intensity and dietary diversity, and bacterial richness, diversity, and composition in infancy, a time when both the microbiota and dietary intake are rapidly changing. Since both breastfeeding intensity and microbial composition are generally confounded by time (55, 56), and each interval in this analysis contains a unique mixture of subjects, intervals were considered separately, with comparisons only made generally based on expected findings as reported in the literature.

The GEHM study's inclusion criteria mandated that the mothers intended to exclusively breastfeed for the first three months post-partum, and excluded infants who were not meeting a minimum of 75% of human milk feeds at four weeks. These criteria resulted in a highly homogeneous cohort, particularly for the first two intervals. With so little variation in breastfeeding status or intensity, it is unsurprising that no significant differences in predictive value of Observed SVs or Shannon diversity emerged in terms of infant feeding in Weeks 2 or 4. Rather, in these early intervals, covariates not directly related to feeding practices, but rather related to the birth environment and genetic characteristics were better predictors of bacterial abundance and diversity.

In Week 2, birth method and infant sex were the only significant predictors of bacterial richness. Previous research supports that birth method is highly predictive of bacterial composition while controlling for breastfeeding status, with infants born by C-section exhibiting microbial profiles significantly different from those born vaginally (2, 4, 9). These differences are due to several factors, including exposure to prophylactic antibiotic use, lack of exposure to vaginal flora, and decreased initiation of
breastfeeding. With overall C-section rates in this cohort consistent with national averages (65), it was expected to find statistical differences in bacterial composition associated with birth method, at least in the first three intervals. However, birth method was a significant predictor only in Week 2. There are several possible explanations for this finding.

Of the subjects that provided samples in Week 2, fewer (55.17%) were born vaginally compared to any other interval (Table 1). This higher proportion of C-section could have made differences in bacterial composition associated with C-section more easily detected in this interval for several reasons. First, antibiotic administration is standard practice with C-section and antibiotic use by the mother prenatally or during breastfeeding has been associated with a significant disruption of bacterial flora up to 12 months of age (10, 11). Maternal antibiotic data was not included in this study, but it may be that the use of antibiotics associated with the higher proportion of C-section rates may have influenced bacterial composition in Week 2, a relationship that was undetected in any other interval.

Secondly, although birth method has been shown to predict bacterial composition, significant differences are more readily detected between elective C-section, rather than acute C-section, and vaginal birth. As most acute C-section are performed after membrane rupture, infants born by acute C-section are typically exposed to vaginal microbes and have microbiota more closely resembling those born vaginally than elective C-section infants (11). Data for type of C-section (acute or elective) was not available for this study, resulting in a failure to differentiate C-section by type for the entire data set, possibly introducing a misclassification bias and thereby
understating the association. It may be that the increased proportion of births by C-section in Week 2 compared to the other weeks increased the number of children born by elective C-section, and allowed for detection of a difference between the groups that was masked in other intervals.

Perhaps most importantly, little consensus exists regarding the duration for which differences in bacterial composition due to birth method remain. Research has shown that exclusive early breastfeeding attenuates, at least partially, the effects of C-section on the microbiota by 3-12 months (2, 22, 66), although significant differences in microbial profiles have been detected up to seven years of age (67). It is possible that the high intensity of breastfeeding in this cohort may have mitigated differences in bacterial composition associated with birth method before Week 4.

Although no significant differences were found between EBF and PBF groups in terms of Observed SV or Shannon diversity, differences were found in terms of family-level relative abundance. Further analysis of Week 2 PBF group revealed that both of the PBF infants were born by C-section. Several of the taxonomic families identified in the Week 2 PBF subjects represent skin and air rather than feces or vagina-associated bacteria, particularly Micrococcaceae and the high percentage of Staphylococcaceae. This finding is consistent with the literature, as C-section born infants have been shown to exhibit microbiota more similar to their mother’s skin and the hospital environment than their mother’s vaginal or fecal taxa, particularly in species of Staphylococcaceae (9). These differences were not noted in Week 4, nor were any taxa belonging to Micrococcaceae or Staphylococcaceae detected in significant (≥1%) amounts after that point, regardless of birth method. Four of the five Week 4 PBF infants were born by C-
section, so the shifts in bacterial composition cannot be explained simply by an increase in vaginal births in this interval compared to Week 2. That the differences in the family level composition between the PBF groups in Weeks 2 and 4 were so distinct suggests that differences associated with a surgical birth may have been, at least partially, assuaged by high breastfeeding intensity by Week 4 in this study, months earlier than the literature would support.

The PBF groups in Weeks 2 and 4 are only comprised of two and five subjects, respectively, while the EBF groups contain 27 and 30 subjects, so relative abundance comparisons can be misleading. As all subjects in Weeks 2 and 4 were breastfeeding at a high intensity (all >85.0%), the groups were split by birth method and reexamined. The Week 2 C-section group (n=13) still exhibited higher ratios of Staphylococcaceae (9.8%) compared to the vaginal birth group (n=16) (1.6%). Micrococcaceae were only found in the vaginal birth group in trace quantities (0.08%), and the relative abundance of Micrococcaceae in the C-section group fell to only 0.6%, which is lower than the amount recorded in the Week 2 PBF groups, but still 13X the vaginal birth group. When dividing the groups by birth method and reexamining these ratios at Week 4, the same pattern emerges as examining these groups by breastfeeding status: relative abundance of Staphylococcaceae falls below 1% for both the C-section and vaginal birth groups, and Micrococcaceae nearly disappears (Vaginal: 0.03%, C-section: 0.001%).

It should also be noted here that GEHM excluded subjects who were not meeting the 75% breastfeeding intensity threshold at Week 4. As C-section has been shown to decrease the likelihood of exclusive breastfeeding and increase the likelihood of early breastfeeding cessation (i.e. prior to five months) (18), it is possible that the most
persistent changes found to be associated with C-section are due to its influence on feeding practices rather than initial exposure at birth. Although most studies examining the relationship between birth method and microbiota control for breastfeeding status or intensity, the establishment of breastfeeding is an incredibly complex process involving biological, physical, and psychosocial factors. Many of the risk factors for C-section (maternal obesity, chronic disease, preterm birth, inadequate prenatal care) (68) are also predictors for lower rates of breastfeeding success (69). By excluding infants not meeting the 75% threshold at Week 4, GEHM ensured that all of the subjects in this study had the benefit of high intensity or exclusive breastfeeding during the first four weeks, perhaps establishing and nurturing colonies of human milk-associated bacteria and attenuating the changes associated with exposure at birth in a very short period of time. This is an encouraging finding, as birth-method associated differences in the microbiota have been associated with a variety of pediatric and adult inflammatory and autoimmune diseases (45, 70). Future studies of high intensity or exclusive breastfeeding with larger sample sizes, controlling for type of C-section, and including antibiotic data are needed to clarify these relationships.

In this study, male sex was predictive of greater bacterial richness and exhibited a trend toward greater bacterial diversity in Week 2 in Models 1 and 2. Studies examining differences in the microbiome by sex are few, especially within a healthy infant population, with results inconsistent between studies. Martin et al (2016) found infant sex to be a significant predictor of bacterial composition of the gut microbiota in the first six months, with boys exhibiting a higher bacterial count at all intervals measured from two days to three months of age (8). However, Cong et al (2016) examined microbial richness,
diversity, and composition by gender in a longitudinal study of healthy pre-term infants over the first 30 days of life and found that boys had lower α diversity during the entire study period (71).

When examining sex-related differences in order-level relative abundance, both of these studies reported similar findings, while this study differed. Martin et al found that relative abundance of Lactobacillales was significantly higher in girls compared to boys (8). Cong et al found that boys were more likely to have profiles dominated by Enterobacteriales, while girls were more likely to have profiles dominated by Clostridiales and Lactobacillales (71). However, in the present study, Lactobacillales (15.2%) and Clostridiales (19.1%) were observed to be higher in the boys’ samples than the girls’ (6.8% and 6.1%, respectively) at two weeks, a finding counter to both Martin and Cong.

It is possible that the sex-related differences in the present study were a result of the higher proportion of births by C-section in Week 2. It has been postulated that boys’ developing microbiota may be more susceptible to the changes introduced by C-section than girls’ (72). A finding that sex predicted differences in bacterial richness and composition during an interval with a higher rate of C-section would be consistent with this hypothesis. While Cong et al confirmed that the C-section rates were not statistically different between the genders, neither Martin nor Cong controlled for birth method as a potential confounder in their analysis of gender-related differences in their respective studies, so this may possibly explain these inconsistencies.

In all models in Week 4, membership in the racial category “Other” was significantly predictive of higher Shannon diversity when compared to “Black,” but not “White.” This finding was unexpected, as only three of 35 subjects identified as “Other” in the interval and
all three were exclusively breastfed. While race/ethnicity (defined as African American or White/Other) was found to predict significant differences in microbiota in both neonates and 6 month old infants in a study of inner-city Detroit mother/infant pairs (73), no difference in predictive value was detected between Black and White groups in this study.

The explanation may lay in race and ethnicity-associated differences in human milk oligosaccharide (HMO) composition. Structural analysis of HMO shows that there is enormous diversity between mothers and throughout lactation in their composition. One factor related to this variation is secretor status (27). Secretors possess an active allele of the fucyltransferase-2 (FUT2) gene, producing a distinct Fucose 1-2 motif in secreted oligosaccharides, which is absent in the HMO of non-secretors (27). Most notably, secretors produce both a wider variety and greater abundance of fucosylated and sialylated HMO, giving the milk of secretor mothers higher HMO abundance and diversity, which have been shown to support the growth of a variety of microorganisms, \textit{Bifidobacteria spp.} in particular (27). Incidence of this polymorphism has been correlated to racial and ethnic differences (74) and may provide an explanation of this finding. While the Detroit study classified Black as one group and White/Other race as a second group, GEHM divided race into six groups (White, Black, Asian, American Indian/Alaskan Native, Native Hawaiian/Pacific Islander, or “Other”), three of which were represented in this study; in the process possibly separating the groups by race-associated differences in secretor status. Further analysis controlling for the secretor status of the mothers in the GEHM study is needed to clarify this relationship, but is beyond the scope of this study.

At 13 weeks, the comparison between the two breastfeeding groups (EBF, PBF) and the sole non-breastfed subject is interesting for many reasons. Partial breastfeeding
was a significant predictor of increased bacterial diversity compared to exclusive breastfeeding in this interval in the unadjusted analysis ($\beta=0.63$, $p=0.04$), but disappears when adjusted for the covariates. However, compositional differences emerge when examining relative abundance of the groups by breastfeeding status. The EBF and PBF groups were remarkably similar. The majority of subjects in this interval were exclusively breastfed, and the three partially breastfed subjects were breastfeeding at a high intensity ($>83\%$), so the similarity in the ratios of major phyla between the partially and exclusively breastfed infants is perhaps unsurprising. This suggests that high intensity breastfeeding coupled with occasional formula supplementation does not result in radical differences in microbial composition at three months of age, a finding that many mothers experiencing difficulties with exclusive breastfeeding will find encouraging.

It is tempting to ascribe the radical differences between in the sole NBF subject, ID#181, and the EBF and PBF groups to the cessation of breastfeeding. However, as the infant microbiota during this time is highly variable and subject to influence by a wide variety of factors (4), this assumption should not and cannot be made. There are many reasons to assume that these changes are due to factors other than the cessation of breastfeeding. First, this subject provided two early fecal samples, at 2 Weeks and 13 Weeks. At 2 Weeks, ID#181 was exclusively breastfed. Compared to all EBF infants in this interval, ID#181 had much higher proportion of Bacteroidetes (49.59%) and much lower Firmicutes (8.91%) than the EBF group as a whole (8.71% and 31.61%, respectively). Secondly, due to the inclusion criteria it can be assumed that ID#181 was breastfeeding $\geq 75\%$ at four weeks, but by Week 13 ID#181 had ceased breastfeeding
and began consuming complementary foods (mMDD=1). Firmicutes, at this point, increased by nearly tenfold from Week 2 (Week 13: 89.66%), while Bacteroidetes were greatly diminished (Week 13: 0.32%), a pattern that was not seen in the EBF or PBF groups. Finally, the adjusted $r^2$ for all models in this interval were negative, indicating that models examining breastfeeding status and intensity explained very little, if any, of the variance.

It is unknown what other factors may have influenced the radical change in phyla proportions in ID #181. It is possible that this subject experienced a medical situation, such as a course of antibiotics, responsible for the remarkable shift in phyla ratios. As this subject cannot be considered representative of non-breastfed infants and due to these gaps in knowledge of other potential influences on the microbiota, it is reasonable to conclude that ID#181’s differences in bacterial composition are the result of factors other than breastfeeding cessation. This data, then, should be considered observational only, and is intended to demonstrate the high variability of the microbiota of the infant during this time.

Reduced breastfeeding intensity, including breastfeeding cessation, predicted lower bacterial richness and diversity in Weeks 26 and 52 in the adjusted analysis. Previous studies have shown that breastfeeding duration and intensity have been found to have a significant impact on the establishment of the gut microbiota in studies ranging from birth to 12 months with higher intensity breastfeeding consistently predicting lower bacterial abundance and diversity (4, 9, 10). This study only detected this relationship after the introduction of complementary foods to the majority of subjects (Weeks 26 and 52).
As previously discussed, with the exception of the single non-breastfed subject in Week 13, all subjects were breastfeeding at a high intensity through the first three intervals (all > 83.3%). As no power analysis was performed for this study, it is possible that the homogeneity of the data in intervals 2, 4, 13, and 104 resulted in an effect size too low to be detected at the power level of each interval. Week 26 is the first interval in which the majority of subjects were PBF (n=27) rather than EBF (n=5), and the first with multiple NBF subjects (n=4). The intensity range of the PBF subjects at this interval was also broader than in previous intervals (37.5-90.9%), allowing for more variation in the data. These differences may have resulted in an effect size sufficient for detection at the power level of this interval and at Week 52 (PBF=37, Intensity: 8.3-63.2%, NBF=6).

Few subjects by Week 26 were exclusively breastfeeding and most had introduced complementary foods. As dietary diversity has been shown to predict bacterial diversity in adults (7), the finding that higher dietary diversity predicted lower bacterial diversity (Model 3: β=-0.38, p=0.008, Model 4: β=-0.29, p=0.01) at Week 26 in this study was unexpected. One possible explanation may lie in the composition of the diet at this time point. Most of the infants were consuming ≤ 2 food groups, with a mean mMDD of 1.60. While the actual content of the diet of the infants in this study is not included, prior analysis of the entire Cincinnati cohort found that most of the infants at this interval were consuming foods only from the grains/roots/tubers category, mostly from the breads and cereals group, with increases in the fruits, vegetables, and meats categories occurring later (9-12 months) (35). Iron-rich infant cereals, commonly fed as a first food to infants, have been shown to influence enteric development (75). However, these cereals are sterile, low in fiber or other complex carbohydrates, and as such do little to encourage bacterial diversity. In their
analysis, Krebs et al showed that the nutrient content of the cereal, whether iron-only or iron + zinc fortified, resulted in significant differences in bacterial groups, especially in the Bifidobacteriaceae and Lactobacillaceae families, with iron-only cereals resulting in decreased abundance in these families (75). Dependent on the types of infant cereals consumed, exclusive consumption of the breads and cereals category could possibly lead to a decrease in bacterial richness and diversity and explain this finding. However, this can only be considered speculative, as this study was limited by its constraint of examining dietary diversity only in terms of an mMDD score. Further studies of the relationship between dietary diversity, including specific food components, and its impact on the developing microbiota are needed to explore this relationship.

It was also surprising that the mMDD score did not significantly predict either bacterial richness or diversity at any other time-point, including Week 52, when the cohort was, on average, meeting the WHO’s MDD recommendations. These results would suggest that dietary diversity per se does not influence microbial diversity in infants. To further explore this, a multiple regression model with a binary variable “MDD met/MDD not met” in place of the mMDD score in Model 4 was run. Although meeting MDD was not a significant predictor ($\beta=0.087, p=0.74$), this model in its entirety was significantly predictive ($p=0.01$) and explained more of the variance (adjusted $r^2=0.33$) in Shannon diversity than any other model at this interval. This model would imply that higher dietary diversity is a contributor to microbial diversity in conjunction with breastfeeding intensity while controlling for infant sex, race, and ethnicity, number of children in the home, WIC enrollment, birth method, and length of hospital stay. No other study to our knowledge has explored quantified dietary diversity as a predictor of bacterial abundance or diversity in infants. This
finding warrants future studies of a longitudinal nature to explore the relationship between increasing dietary diversity and the changes in microbial composition.

Another possible explanation for the lack of predictive value of dietary diversity for microbial composition may be the study design itself. Dietary diversity is a progressive phenomenon, but no samples were taken between six months and one year in this study as the subjects’ complementary diet developed. In addition, each interval contained a unique group of subjects, making comparisons between the intervals for changes in microbial diversity and richness invalid. To truly explore the influence of dietary diversity on the developing microbiota requires a longitudinal design, with sampling intervals during the period of complementary diet development to track changes associated with the addition of food groups on an individual and group levels rather than predictability of bacterial composition associated with consumption.

The WHO recommends achieving MDD as a way to improve overall diet quality and ensure adequate intake in a variety of macro and micronutrients. The links between dietary diversity and diet quality are well established and form the basis of the WHO’s recommendations (36). Although this study did not find a positive relationship between dietary diversity and microbial richness and diversity, it is important to remember that no direct correlation has been established between microbial diversity and richness and any marker of health and wellness in the infant. Thus, the negative predictive value at 26 weeks between dietary diversity and microbial diversity does not imply or suggest any detrimental influence of dietary diversity in the development of a healthy, diverse, and abundant microbiota later in life. As previous research has linked diet quality and diversity in infancy and early childhood to the likelihood of a more diverse and higher quality diet later in life
and a high quality, diverse diet has been linked to healthy, diverse, and abundant microbial profiles in adulthood (7), one could argue that the impact of a diverse diet on the infant's microbiota at the least is a tangential, long term association rather than a direct and immediate one.

Although all infants in the GEHM study were full term, gestational age was a significant predictor of bacterial composition at several time points. Gestational age was significantly predictive of Shannon Index in Week 4 (Models 1 and 2) in Week 52 (Model 4), and trended toward significance in Week 26 (Model 2, Observed SV, Models 3 & 4, Shannon Index) and Week 52 (Models 1 & 3, Shannon Index). Gestational age has been previously shown to influence the microbiota of both breastmilk and the infant gut and is associated with differences in the gut microbiota, even in full term infants (55, 73, 76). Evidence indicates this relationship is particularly important in the abundance of Bifidobacteriaceae and Enterococcaceae (76), two families present in the profiles of all infants at all intervals in this study.

The examination of the changes related to breastfeeding status in relative abundance in this study illustrate the highly changeable and malleable nature of the infant gut microbiota and support the growing evidence that high intensity breastfeeding attenuates changes wrought by other factors. As expected, this study found that the ratios of bacterial profiles of the infants were dominated by Firmicutes, Bacteroidetes, Proteobacteria, and Actinobacteria at all intervals, were highly variable in the first six months, experienced an increase in microbial exposure with the introduction of complementary foods and decreased breastfeeding around Week 26, and settled into more mature, adult-like profiles between ages one and two. Little change was noted between the
ratios of phyla, when grouped by breastfeeding status, from Week 52 to Week 104.

Although generally, the age-associated development of the microbiota resulted in increases in taxa from Firmicutes and Bacteroidetes and decreases in Actinobacteria and Proteobacteria, a family-level examination of the taxa provides a clearer picture of the maturation patterns over the first two years.

Of particular interest in this study are the changing ratios of families specifically associated with breastfeeding, such as Lactobacillaceae (Firmicutes) and Bifidobacteriaceae (Actinobacteria), as the profiles mature. Both Lactobacillaceae and Bifidobacteriaceae are families of lactic-acid and short-chain fatty acid producing bacteria, are associated with a variety of host benefits, including protection against pathogens and short chain fatty acid and digestive enzyme production, particularly lactase (77), and have been used as probiotics for various digestive and health benefits, including improved lactose tolerance and improved intestinal regularity. Their patterns of decline in this study were not unexpected, as both of these families are associated with breastfeeding and infancy, but these declining ratios should be considered in the context of the current application of probiotic supplements.

As breastfed infants experience lowered risk of food allergies and other atopic diseases (78), introducing species found predominantly in breastfed infants’ profiles as a therapeutic option is currently being explored as a treatment for these disorders. For example, a subspecies in the Lactobacillaceae family associated with lactase production, *L. rhamnosus GG (LGG)*, has recently been studied as a probiotic supplement in infant formula to treat cow’s milk allergy and has shown promising results, with significantly increased tolerance to cow’s milk in the LGG group compared to a control group and
significant increases in butyrate production (77). However, Lactobacillaceae are heavily associated with breastfeeding and early infancy and are rarely found in any significant quantities outside of these conditions. The present study isolated Lactobacillaceae, including *L. rhamnosus*, in significant quantities only in currently breastfeeding subjects, consistent with previous findings. It is possible that probiotic supplementation with these organisms is not sustainable outside of the context of breastfeeding or without the addition of appropriate prebiotics to promote their growth and development. Additional research is needed to better understand these bacteria's nutritional needs to ensure their development, maintenance, and growth and to ensure benefits associated with probiotic supplementation are sustainable in the non-breastfed and/or mature human gut.

The increases noted in this study in the Lachnospiraceae family are consistent with previously established patterns of microbial maturation. Lachnospiraceae are found in high abundance in adult humans in all regions of the world (79) and different species belonging to the Lachnospiraceae family have been associated with both beneficial, probiotic properties and have been investigated for a possible link to metabolic disorders, including poor glucose control and obesity (79). In general, Lachnospiraceae were lower in breastfed groups in intervals 13-104 than NBF groups, and exhibited a pattern of progressive increases by time in all categories. The exception to this pattern was the single Week 13 NBF subject, whose profile was dominated by Lachnospiraceae (84.21%). A genus-level examination of the Lachnospiraceae taxa present in this study was unable to differentiate the majority of genera of this family, so a deeper exploration of the relationships between breastfeeding status, dietary diversity, and the increases in this family of bacteria is not possible.
Any interpretation of microbial composition in relation to feeding practices in infants must contend with the maturation of the microbiota over time. As symbionts, the bacteria in the human gut matures with its host and both microbial composition and breastfeeding intensity have been shown to be cofounded by time (55, 56). This was demonstrated by the Bray-Curtis distance analysis and the ordination plots. Clustering by both study week and breastfeeding status was confirmed by multivariate ANOVA, suggesting that infant microbial development is related to both time and infant dietary intake. However, additional analysis showed that the dispersion patterns were significantly heterogeneous between the intervals \((p=0.004)\), so it cannot be assumed that the differences by interval are due only to time or breastfeeding status, but possibly also influenced by differences in dispersion caused by the cross-sectional nature of the data. While the findings of differences by study interval and breastfeeding status are consistent with current research, longitudinal studies are needed to confirm these relationships.

The GEHM study benefits from extensive, detailed infant feeding data allowing for comparisons by multiple feeding practices. The subjects for this study were drawn from a small regional area, reducing the possibility of environmental confounding due to region-specific environmental bacteria or cultural practices. In addition, since all subjects were breastfeeding at a very high intensity through the first two intervals, and most through the first six months, it is possible to examine the attenuation of factors known to influence the early infant gut bacteria due to high intensity breastfeeding. The use of a quantified dietary diversity measure to examine microbial development is also a novel approach to examine the influence of dietary diversity in infant gut microbial development.
A number of limitations should be considered. Although the homogeneity of the data in terms of breastfeeding status and intensity provided some advantages, it reduces the generalizability of these findings and possibly reduced the probability of detecting differences related to dietary diversity in later intervals due to early colonization by human milk-associated bacteria and the high intensity and duration of breastfeeding through 52 weeks and the continuation of partial breastfeeding through 104 weeks by a number of subjects.

The data was homogeneous and atypical in other ways as well. The mothers of the infants included in this study were highly educated, mostly white and non-Hispanic, and, at 31.50 ±5.37 years old, older than average US mothers for first, second, or third births (80). In addition, all of the subjects in this study were from an urban environment, so these findings may not be applicable to residents of a suburban or rural area, where increased exposures to farm animals and a rural diet may affect bacterial composition.

Although WIC enrollment at birth was used as a proxy for socioeconomic status (SES), no direct income information was available, so a true SES variable could not be included. SES has repeatedly been shown to be a negative predictor of both diet quality and dietary diversity (36), and has been associated with microbial changes in infancy (81), so the inability to control for SES may confound the results. WIC enrollment at birth was a trending covariate at Week 26 when examining Shannon Diversity, which may indicate that SES could be a factor in this cohort.

As previously discussed, no data on type of C-section was included with this analysis. The failure to find significant predictability beyond Week 2 for birth method may indicate that this failure introduced a misclassification bias, as prior research suggests that
bacterial changes due to birth method would survive past two weeks. Indeed, the differences in family-level microbial composition in Week 2 associated with breastfeeding status and when sorted by birth method suggest that birth method was a major influence on microbial composition in this group, but the lack of significance at Week 4 is either due to the inability to separate acute from elective C-section in the group or attenuation of changes due to high intensity breastfeeding. Without the ability to separate the data by C-section type, the reasons for these differences are only speculative.

While it may be assumed that the mothers of all C-section infants were administered routine antibiotics, the lack of antibiotic data for all mothers and subjects may also confound the results. An example of this may be ID#181. The radical change in this subject’s profile from Week 2 to Week 13 could be due to breastfeeding cessation or the early introduction of complementary foods, but could also be the result of a medical situation or routine administration of antibiotics.

While this study’s goal was to examine the effect of dietary diversity per se on the infant microbiota, the lack of specificity in the dietary data limited the depth of investigation possible. This study did find a significant negative relationship between mMDD score and dietary diversity at 26 weeks, a time when complementary foods were introduced to the majority of the subjects, but with low dietary diversity and high intensity breastfeeding and a time in which prior analysis determined the majority of subjects in this cohort were consuming only foods from the cereals group (35). While previous research did find that type of infant cereal could influence enteric bacteria (75), this study did not include specific foods data, limiting the interpretation of these findings.
Finally, no power analysis was performed to determine the necessary sample size to detect the relationships explored in this study. Although many studies of the microbiota have small sample sizes, the homogeneity of the data in this study may have limited effect size to such that was undetectable at the power of this study. A large number of covariates also reduced the degrees of freedom in each model, increasing the effect size required to detect these relationships.

VII. Conclusion

This study examined the relationships between infant feeding practices, including breastfeeding status and intensity and quantified diversity of the complementary diet, and the development of the infant microbiota. While the data was highly homogeneous in terms of breastfeeding status and intensity, especially during the first two intervals, significant differences due to a variety of factors were detected.

Findings of particular importance are that birth method was not a significant predictor of bacterial richness or diversity at any point past two weeks of age and that no significant differences in bacterial abundance or diversity were found between exclusively and high-intensity partially breastfed subjects prior to the introduction of complementary foods. In addition, taxonomic family-level compositional differences detected in this study at two weeks were largely mitigated by Week 4, both when examining the groups by breastfeeding status and by birth method. Although high intensity breastfeeding may be implicated in these findings, the cross sectional nature of this data prevents us from drawing any conclusions regarding the cause of these changes. Future longitudinal studies are needed to investigate the relationship between
high intensity breastfeeding and the development of the microbiota during this period of rapid change.

The surprising finding at 26 weeks that higher mMDD score predicted lower bacterial diversity is intriguing and warrants additional study. While prior research indicates that this cohort was consuming a diet almost exclusively from the breads/grains/cereals group at this time and evidence supports that an iron-enriched cereal diet may reduce bacterial diversity and abundance, the lack of specific dietary data limits the ability to further investigate this connection. Future studies examining quantified dietary diversity including specific WHO food groups consumed, with sampling periods between the introduction of complementary foods and the achievement of MDD, are needed to adequately explore this relationship.

Finally, the finding that racial differences predicted significant differences in bacterial diversity in Week 4, despite the high intensity of breastfeeding and racial homogeneity in the cohort, supports the growing body of evidence that racially associated genetic factors, such as secretor status, may be a significant predictor of microbial composition. The composition of human milk is an area of ongoing and exciting research and understanding the relationship between the mother’s genetic profile, HMO composition and diversity, and the infant’s microbiota may help us better understand the relationships between prebiotics and the growth and maintenance of a healthy and robust gut microbiota. This information could be useful for the development of effective pro+prebiotic (or synbiotic) formulas for use by infants unable to receive feedings of human milk and to treat diseases associated with dysfunctional bacterial composition.
This study used a sequential cross-sectional design, and did not include important potential confounders, such as C-section type and maternal and infant antibiotic use. Due to the potential implications of this study’s findings, further investigation using a prospective cohort design controlling for these confounders and with more sampling frequency are warranted. At present, our understanding of the development, maintenance, and roles of the microbiota is in its infancy. Future studies examining the influence of high intensity breastfeeding and diversity of the complementary diet on the infant’s gut microbiota may help to clarify these relationships.
VIII. Bibliography


60. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and


