JAMA Pediatrics | Original Investigation

Association of Exposure to Formula in the Hospital and Subsequent Infant Feeding Practices With Gut Microbiota and Risk of Overweight in the First Year of Life

Jessica D. Forbes, PhD; Meghan B. Azad, PhD; Lorena Vehling, MSc; Hein M. Tun, PhD; Theodore B. Konya, MPH; David S. Guttman, PhD; Catherine J. Field, PhD; Diana Lefebvre, PhD; Malcolm R. Sears, MBChB; Allan B. Becker, MD; Piushkumar J. Mandhane, MD, PhD; Stuart E. Turvey, MBBS, DPhil; Theo J. Moraes, MD, PhD; Padmaja Subbarao, MD, MSc; James A. Scott, PhD; Anita L. Kozyrskyj, PhD; for the Canadian Healthy Infant Longitudinal Development (CHILD) Study Investigators

IMPORTANCE The effect of neonatal and infant feeding practices on childhood obesity is unclear. The gut microbiome is strongly influenced by feeding practices and has been linked to obesity.

OBJECTIVE To characterize the association between breastfeeding, microbiota, and risk of overweight during infancy, accounting for the type and timing of supplementary feeding.

DESIGN, SETTING, AND PARTICIPANTS In this study of a subset of 1087 infants from the prospective CHILD pregnancy cohort, mothers were recruited between January 1, 2009, and December 31, 2012. Statistical analysis was performed from February 1 to December 20, 2017.

MAIN OUTCOMES AND MEASURES Feeding was reported by mothers and documented from hospital records. Fecal microbiota at 3 to 4 months (from 996 infants) and/or 12 months (from 821 infants) were characterized by 16S ribosomal RNA sequencing. Infants with a weight for length exceeding the 85th percentile were considered to be at risk for overweight.

RESULTS There were 1087 infants in the study (507 girls and 580 boys); at 3 months, 579 of 1077 (53.8%) were exclusively breastfed according to maternal report. Infants who were exclusively formula fed at 3 months had an increased risk of overweight in covariate-adjusted models (53 of 159 [33.3%] vs 74 of 386 [19.2%]; adjusted odds ratio, 2.04; 95% CI, 1.25-3.32). This association was attenuated (adjusted odds ratio, 1.33; 95% CI, 0.79-2.24) after further adjustment for microbiota features characteristic of formula feeding at 3 to 4 months, including higher overall richness and enrichment of *Lachnospiraceae*. A total of 179 of 579 infants who were exclusively breastfed (30.9%) received formula as neonates; this brief supplementation was associated with lower relative abundance of *Bifidobacteriaceae* and higher relative abundance of *Enterobacteriaceae* at 3 to 4 months but did not influence the risk of overweight. At 12 months, microbiota profiles differed significantly according to feeding practices at 6 months; among partially breastfed infants, formula supplementation was associated with a profile similar to that of nonbreastfed infants (higher diversity and enrichment of *Bacteroidaceae*), whereas the introduction of complementary foods without formula was associated with a profile more similar to that of exclusively breastfed infants (lower diversity and enrichment of *Bifidobacteriaceae* and *Veillonellaceae*). Microbiota profiles at 3 months were more strongly associated with risk of overweight than were microbiota profiles at 12 months.

CONCLUSIONS AND RELEVANCE Breastfeeding may be protective against overweight, and gut microbiota may contribute to this effect. Formula feeding appears to stimulate changes in microbiota that are associated with overweight, whereas other complementary foods do not. Subtle microbiota differences emerge after brief exposure to formula in the hospital. These results identify important areas for future research and distinguish early infancy as a critical period when transient gut dysbiosis may lead to increased risk of overweight.

JAMA Pediatr. 2018;172(7):e181161. doi:10.1001/jamapediatrics.2018.1161
Published online June 4, 2018. Corrected on July 2, 2018.

Supplemental content

Author Affiliations: Author affiliations are listed at the end of this article.

Group Information: The Canadian Healthy Infant Longitudinal Development (CHILD) Study investigators are listed at the end of this article.

Corresponding Author: Meghan B. Azad, PhD, Developmental Origins of Chronic Diseases in Children Network (DEVOTION), Department of Pediatrics and Child Health, Children's Hospital Research Institute of Manitoba, University of Manitoba, 501G-715 McDermot Ave, Winnipeg, MB R3E 3P4, Canada (meghan.azad@umanitoba.ca).

Obesity originates early in life,[1] and breastfeeding appears to be protective against obesity.[2] Hypothesized mechanisms for this protection include the promotion of self-regulation in breastfed infants and the lower protein content of breast milk compared with infant formula.[3] Another potential mechanism involves modification of the developing gut microbiota, which contributes to nutrient acquisition, energy regulation, and fat storage.[4] Microbiota shifts have been associated, albeit inconsistently,[5] with obesity in adults, including lower diversity, enrichment of *Ruminococcus gnavus*,[6] and a higher ratio of Firmicutes to Bacteroidetes.[7] Microbiota transplant experiments in mice suggest that these associations are causal,[8] and studies of children[9-12] suggest that they originate early in life, although few studies have been conducted for infants. Breastfeeding is among the most influential factors shaping the infant gut microbiome because breast milk contains prebiotic oligosaccharides and probiotic microorganisms, including bifidobacteria.[13]

Despite this evidence, we do not fully understand how infant feeding practices affect the developing microbiota and influence weight gain. Studies often do not differentiate between partially breastfed infants receiving formula vs those receiving complementary foods, yet these forms of nutrition clearly provide very different substrates for microbiota. The definition of *exclusive breastfeeding* also varies, and few studies have accessed hospital records to confirm exclusivity in the neonatal period. To address these knowledge gaps, we characterized these specific infant feeding practices in the Canadian Healthy Infant Longitudinal Development (CHILD) birth cohort and examined their association with gut microbiota and risk of overweight in the first year of life.

Key Points

Question How do infant feeding practices influence gut microbiota and risk of overweight?

Findings Among 1087 infants from the Canadian Healthy Infant Longitudinal Development (CHILD) cohort, earlier cessation of breastfeeding and supplementation with formula (more so than complementary foods) were associated with a dose-dependent increase in risk of overweight by age 12 months; this association was partially explained by specific gut microbiota features at 3 to 4 months. Subtle but significant microbiota differences were observed after brief exposure to formula limited to the birth hospital stay, but these differences were not associated with overweight.

Meaning Breastfeeding may contribute to protection against overweight by modifying the gut microbiota, particularly during early infancy.

Methods

Study Design

We accessed data from the CHILD birth cohort (http://childstudy.ca) of 3495 families across 4 sites in Canada.[14] Women were recruited between January 1, 2009, and December 31, 2012, and remained eligible if they delivered a healthy, full-term infant. This study included 1087 infants enrolled in the general cohort at the Manitoba, Edmonton, and Vancouver sites. This subset is a representative selection of infants with fecal samples analyzed at 3 to 4 months (from 996 infants) and/or 12 months (from 821 infants), of which 730 infants had samples analyzed at both times (eFigure 1 in the Supplement). The rates of breastfeeding, overweight, and other demographics in this subset were similar to those of the general cohort (eTable 1 in the Supplement). The Human Research Ethics Boards at McMaster University, University of Manitoba, University of Alberta, University of Toronto, and University of British Columbia approved this study. Parents provided written consent at the time of enrollment.

Overweight

At 12 months of age (mean [SD] age, 12.4 [1.3] months), infants were weighed and measured by CHILD Study staff. Age- and sex-specific weight for length *z* (WFL*z*) scores were calculated according to World Health Organization standards.[15] A WFL*z* score greater than the 97th percentile was considered overweight, and a WFL*z* score greater than the 85th percentile was considered at risk for overweight[16]; these 2 groups were combined into a composite outcome for logistic regression analyses.

Infant Feeding

Mothers completed questionnaires at 3, 6, and 12 months post partum, reporting on breastfeeding and the introduction of formula and complementary foods. At 3 months, breastfeeding status was classified as exclusive (breast milk only), partial (breast milk and formula), or none (formula only). Using hospital data, we further classified infants as exclusively breastfed after hospital discharge if they briefly received formula in the hospital but were exclusively breastfed after hospital discharge. At 6 months, feeding was defined as exclusively breastfed (breast milk only), partially breastfed with formula (breast milk and formula, with or without complementary foods), partially breastfed without formula (breast milk and complementary foods), or not breastfed (formula with or without complementary foods). The duration of breastfeeding was determined from the earliest report of cessation of breastfeeding. For microbiota analyses, breastfeeding status was determined on the date of collection of the fecal sample. In this study, *breastfeeding* refers to feeding the infant breast milk, whether at the breast or from a bottle.

Covariates

Mode of birth, parity, gestational diabetes, infant sex, birth weight, and hospital-administered antibiotics to the mother or neonate were documented from hospital records.[17] Oral antibiotic use was reported by parents. As described previously,[18] the quality of the maternal diet was estimated using the Healthy Eating Index,[19] and the maternal prepregnancy body mass index was self-reported and validated against medical records. Data on maternal race/ethnicity, smoking status, educational level, and pet ownership were self-reported during pregnancy.

Fecal Microbiota Analysis

Fecal samples were collected at a home visit (3-4 months; mean [SD], 3.7 [1.0] months) and a clinic visit (12 months; mean [SD], 12.3 [1.2] months); DNA was extracted using the QIAamp DNA

Table 1. Crude and Adjusted Association of Infant Feeding Practices With Infant Weight Status at 12 Months

Breastfeeding Exposure	Prevalence of Overweight, No. (%)	Crude OR (95% CI) (n = 1020)	Adjusted OR (95% CI) With Multiple Imputation of Missing Data (N = 1087)[a]
Breastfeeding at 3 mo			
None (formula only)	53/159 (33.3)	2.11 (1.39-3.19)	2.02 (1.18-3.45)
Partial (breast milk and formula)	84/304 (27.6)	1.61 (1.13-2.30)	1.63 (1.09-2.44)
Exclusive after hospital discharge	35/171 (20.5)	1.09 (0.68-1.69)	1.13 (0.68-1.89)
Exclusive (breast milk only)	74/386 (19.2)	1 [Reference]	1 [Reference]
Breastfeeding at 6 mo		(n = 1001)	
None (formula with or without food)	77/249 (30.9)	2.11 (1.33-3.42)	1.59 (0.92-2.74)
Partial with formula (breast and formula with or without food)	81/296 (27.4)	1.77 (1.13-2.85)	1.43 (0.87-2.37)
Partial without formula (breast milk and food)	55/279 (19.7)	1.16 (0.71-1.90)	0.96 (0.57-1.64)
Exclusive (breast milk only)	31/177 (17.5)	1 [Reference]	1 [Reference]
Breastfeeding duration		(n = 978)	
<6 mo[b]	68/219 (31.1)	2.02 (1.39-2.93)	1.64 (1.06-2.52)
6 to <12 mo	85/309 (27.5)	1.70 (1.21-2.41)	1.47 (0.99-2.18)
≥12 mo	82/450 (18.2)	1 [Reference]	1 [Reference]

Abbreviation: OR, odds ratio.

[a] Adjusted for maternal body mass index, smoking, postsecondary education, race/ethnicity, cesarean delivery, dog in household, infant sex, any oral antibiotics between 0 and 12 mo, and study site.

[b] Excludes infants who were never breastfed. Breastfeeding refers to breast milk feeding regardless of feeding mode (at the breast or from a bottle).

Stool Mini Kit (Qiagen); and the 16S ribosomal RNA gene, hypervariable region V4, was amplified and sequenced by Illumina MiSeq (eAppendix in the Supplement). Using QIIME, version 1.8.0,[20] reads were assembled, demultiplexed, filtered against the Greengenes reference database, version 13.8,[21] and clustered at 97% similarity. After filtering, a total of 265 095 597 reads were retained (median, 235 623 per sample [range, 13 134-833 392]), representing 939 unique operational taxonomic units. For subsequent analyses, data were rarefied to 13 000 sequences per sample and summarized at the family taxonomic level.

Statistical Analysis

Statistical analysis was performed from February 1 to December 20, 2017. Covariates were tabulated against feeding and overweight and compared by use of the χ^2 test. Multivariable regression was used to investigate associations between feeding and overweight. Models were adjusted for suspected confounders selected a priori or identified in univariate analyses, grouped as maternal body mass index, other maternal factors (educational level, smoking status, ethnicity, and study site), and microbiota-related factors (cesarean delivery, dog ownership, infant sex, and antibiotics). Sensitivity analyses were conducted to adjust for birth weight, exclude never-breastfed infants, and evaluate continuous WFL*z* scores as an alternative outcome. Results are presented as crude odds ratios (ORs) and adjusted ORs (aORs) or differences in WFL*z* scores (SDs with 95% CIs). Multiple imputation (20 imputed data sets) was performed for all covariates using the R package mice.[22] Microbiota alpha diversity was assessed using the abundance-based coverage estimator and Chao1 indices of species richness and the Simpson and Shannon indices of diversity. Microbiota measures were compared between feeding or weight status groups by use of nonparametric Kruskal-Wallis tests and post hoc Dunn tests with false discovery rate (FDR) correction for multiple comparisons. Microbiota community structures were compared by permutational analysis of variance (PERMANOVA) on UniFrac[23] distance matrices and visualized by principal coordinate analysis. Microbiota composition and diversity (classified in quartiles) were further investigated in multivariable logistic regression models to evaluate their influence on the association between breastfeeding and risk of overweight. All analyses were performed in R, version 3.3.3 (R Development Core Team). $P < .05$ (2-sided) after FDR correction was considered significant.

Results

Study Population

Most mothers were white (817 of 1078 [75.8%]) and delivered vaginally (790 of 1064 [74.2%]); 408 of 1025 mothers (39.8%) were overweight or obese (eTable 1 in the Supplement). The breastfeeding initiation rate was 95.5% (1032 of 1081) (eTable 2 in the Supplement). At 3 months, 53.8% of infants (579 of 1077) were exclusively breastfed, including 37.1% (400 of 1077) who were exclusively breastfed since birth and 16.6% (179 of 1077) who briefly received formula in the hospital. The remaining infants were partially breastfed (323 of 1077 [30.0%]) or not breastfed (175 of 1077 [16.2%]). By 6 months, the rate of exclusive breastfeeding had decreased to 17.6% (183 of 1040), and partial breastfeeding had increased to 54.6% (593 of 1087), including 28.2% (307 of 1087) who received formula with or without food and 26.3% (286 of 1087) who received food but not formula. At 12 months, 42.2% of infants (459 of 1087) were still breastfeeding; the mean (SD) WFL*z* score was 0.29 (1.08), and 22.9% of infants (249 of 1087) were overweight or at risk for overweight.

Infant Feeding and Risk of Overweight

Breastfeeding was associated with a lower risk of overweight at 12 months, with dose responses observed according to breastfeeding exclusivity and duration (**Table 1**). Among infants who

were exclusively breastfed at 3 months, 19.2% (74 of 386) were overweight or at risk of overweight by 12 months compared with 27.6% of infants (84 of 304) who were partially breastfed (OR, 1.61; 95% CI, 1.13-2.30) and 33.3% of infants (53 of 159) who were not breastfed (ie, exclusively formula fed) (OR, 2.11; 95% CI, 1.39-3.19). There was no increase in risk of overweight among exclusively breastfed infants who briefly received formula in the hospital (35 of 171 [20.5%] at risk; OR, 1.09; 95% CI, 0.68-1.69). These associations were largely unaffected by adjustment for maternal body mass index, education, smoking, and other potential confounders (eTable 3 in the Supplement) (partial breastfeeding: aOR, 1.63; 95% CI, 1.09-2.44; exclusive formula feeding: aOR, 2.02; 95% CI, 1.18-3.45; exclusive breastfeeding after hospital discharge: aOR, 1.13; 95% CI, 0.68-1.89) (Table 1).

At 6 months, partial breastfeeding supplemented with formula was associated with an increased risk of overweight when adjusting individually for maternal body mass index (aOR, 1.60; 95% CI, 1.01-2.59), other maternal factors (aOR, 1.65; 95% CI, 1.03-2.68), or microbiota-related factors (aOR, 1.64; 95% CI, 1.02-2.70), although statistical significance was lost in the fully adjusted model (aOR, 1.43; 95% CI, 0.87-2.37) (Table 1). In contrast, partial breastfeeding without formula (ie, with foods only) was not associated with risk of overweight (aOR, 0.96; 95% CI, 0.57-1.64). Earlier cessation of breastfeeding was associated with an increased risk of overweight (before 6 months: aOR, 1.64; 95% CI, 1.06-2.52; between 6 and 12 months: aOR, 1.47; 95% CI, 0.99-2.18 compared with 12 months or longer). Sensitivity analyses using the WFL*z* score as a continuous outcome, adjusting for infant birth weight or excluding infants who never received breast milk, followed similar patterns of association (eTable 4 in the Supplement).

Infant Feeding and Gut Microbiota

As expected, breastfeeding was strongly associated with the richness, diversity, and composition of gut microbiota at 3 to 4 months, with clear dose responses according to exclusivity (**Figure 1** and eTables 5 and 6 in the Supplement). The richness and diversity of microbiota were highest in infants who were not breastfed, lower in partially breastfed infants, and lowest in exclusively breastfed infants (Figure 1A). The community structure of microbiota also differed significantly (overall $P = .001$, pseudo F, 10.9 [unweighted UniFrac]; $P = .001$, pseudo F, 12.4 [weighted UniFrac], determined by use of PERMANOVA; eTable 6 in the Supplement), with principal coordinate analysis (Figure 1D and eFigure 2A and B in the Supplement) showing clear separation between the exclusively breastfed and nonbreastfed groups. The group that briefly received formula in the hospital overlapped almost completely with the exclusively breastfed group ($P = .24$, pseudo F, 0.24, determined by use of pairwise PERMANOVA) (Figure 1D and eTable 6 in the Supplement), indicating similar microbiota community structures.

Nearly all phyla and families demonstrated disproportional abundances across breastfeeding groups, and significant dose responses were observed with particular taxa (Figure 1B and C and eTable 5 in the Supplement). Increasing exclusivity of breastfeeding was associated with increasing relative abundance of *Bifidobacteriaceae* and *Enterobacteriaceae* and decreasing relative abundance of *Lachnospiraceae*, *Veillonellaceae*, and *Ruminococcaceae*. Although most taxa were similarly abundant between infants who were exclusively breastfed from birth and those exclusively breastfed after hospital discharge, the relative abundance of *Bifidobacteriaceae* was significantly lower after brief exposure to formula in the hospital (median, 4.3% vs 8.3% of total microbiota; FDR $P = .03$) and the relative abundance of *Enterobacteriaceae* was higher (29.8% vs 24.5% of total microbiota; FDR $P = .05$) (Figure 1C and eTable 5 in the Supplement).

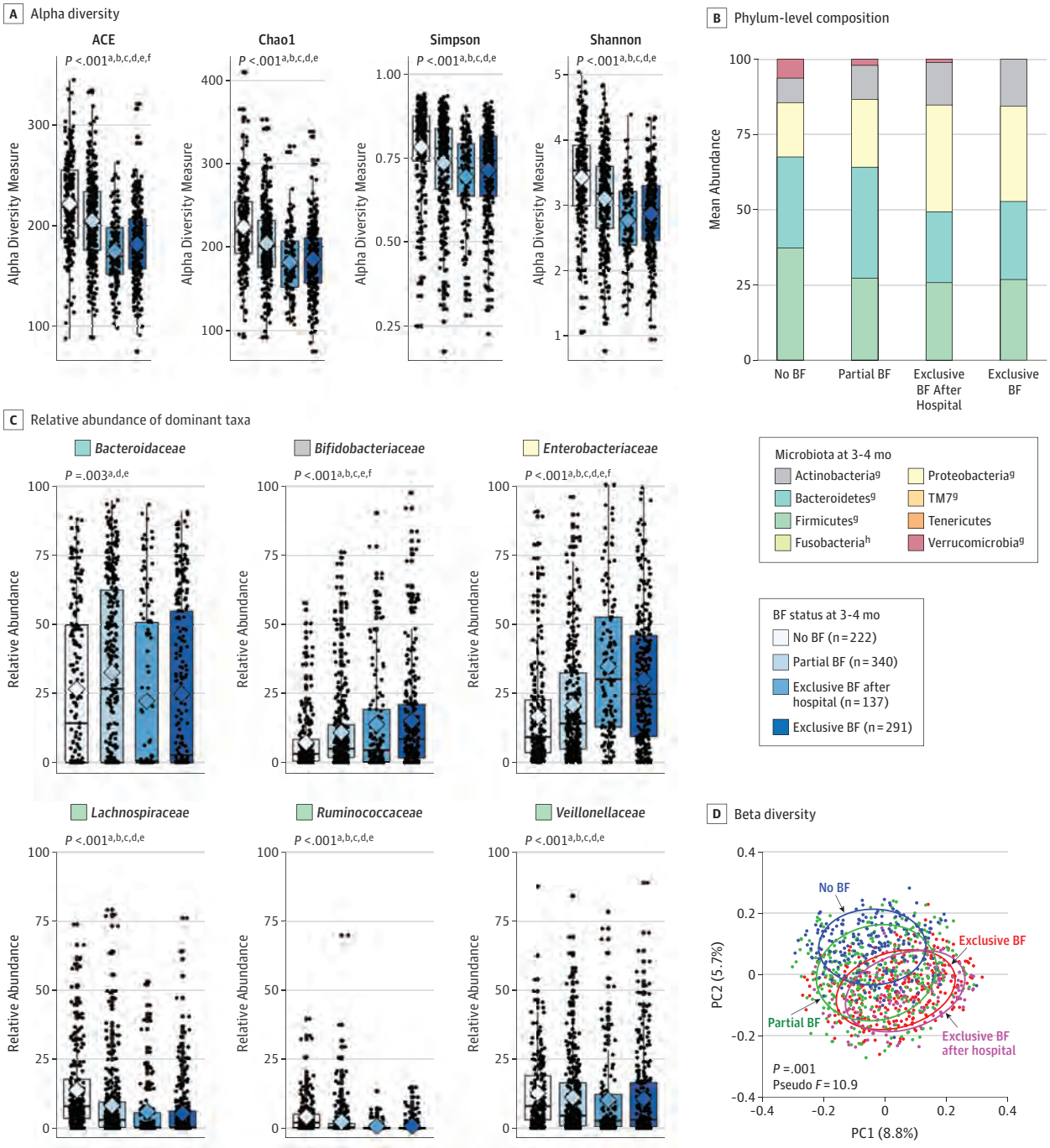
Twelve-month microbiota profiles were more homogeneous overall, but significant differences were still detectable according to dietary exposures at 6 months (**Figure 2**A-D, eFigure 2C and 2D, and eTables 6 and 7 in the Supplement). Richness was significantly higher among formula-fed infants (whether or not they were also receiving breast milk) compared with breastfed infants (whether or not they were receiving complementary foods) (Figure 2A). The relative abundances of Actinobacteria and Proteobacteria were highest in exclusively breastfed infants and lowest in nonbreastfed infants (Figure 2B). Several differences were observed between the partial breastfeeding groups, including significantly higher relative abundance of *Bifidobacteriaceae* and *Veillonellaceae* in those receiving complementary foods without formula (Figure 2C). Overall, the microbiota of partially breastfed infants who did not receive formula were similar to the microbiota of exclusively breastfed infants (no significant differences by 12 months; $P = .78$, pseudo $F = 0.40$, determined by use of pairwise PERMANOVA), whereas the microbiota of those who received formula were more similar to the microbiota of nonbreastfed infants (Figure 2D and eTable 6 in the Supplement).

The duration of breastfeeding was also associated with gut microbiota at 12 months (eFigure 3 and eTable 8 in the Supplement). Richness and diversity were lowest among infants who were still breastfeeding at 12 months and highest among those who had weaned before 6 months. *Bifidobacteriaceae*, *Veillonellaceae*, and Proteobacteria were enriched among infants who were still breastfeeding and depleted among infants who had never been breastfed. In contrast, *Lachnospiraceae*, *Ruminococcaceae*, and *Porphyromonadaceae* were enriched among infants who were not breastfeeding at 12 months.

Gut Microbiota and Overweight

Infants who were overweight or at risk of overweight at 12 months had significantly higher richness of microbiota by 3 to 4 months of age (**Figure 3**A); significant differences in composition were also detected (Figure 3B and C and eTable 9 in the Supplement). The strongest association was the enrichment of *Lachnospiraceae* among infants who subsequently became overweight (median relative abundance, 5.9% of total microbiota) or at risk for overweight (median relative abundance, 4.7% of total microbiota) by 12 months compared with normal-weight infants (median relative abundance, 1.9% of total microbiota; FDR $P = .01$). We also observed significantly higher relative abundance of *Coriobacteriaceae*, *Erysipelotrichaceae*, and *Ruminococcaceae*

Figure 1. Infant Gut Microbiota at 3 to 4 Months According to Breastfeeding (BF) Status



A, Alpha diversity evaluated by richness (abundance-based coverage estimator [ACE] and Chao1) and diversity (Simpson and Shannon). Median estimates are compared across feeding groups using the Kruskal-Wallis test (nonparametric analysis of variance) and Dunn post hoc tests for multiple comparisons. Boxes indicate interquartile range, lines indicate medians, diamonds indicate means, and whiskers represent range. B, Mean phylum-level composition. C, Relative abundance of dominant taxa across feeding groups. Breastfeeding (BF) status is assessed at the time of sample collection. Breastfeeding refers to breast milk feeding regardless of feeding mode (at the breast or from a bottle). D, Principal coordinate analysis (PC1 and PC2) based on unweighted UniFrac distances, with community structure differences tested by permutational analysis of variance with 999 permuations.

P values represent false discovery rate–corrected *P* values testing for overall differences across the 4 feeding groups. Significant pairwise comparisons:

[a] No BF/partial BF;

[b] No BF/exclusive BF after hospital;

[c] No BF/exclusive BF;

[d] Partial BF/exclusive BF after hospital;

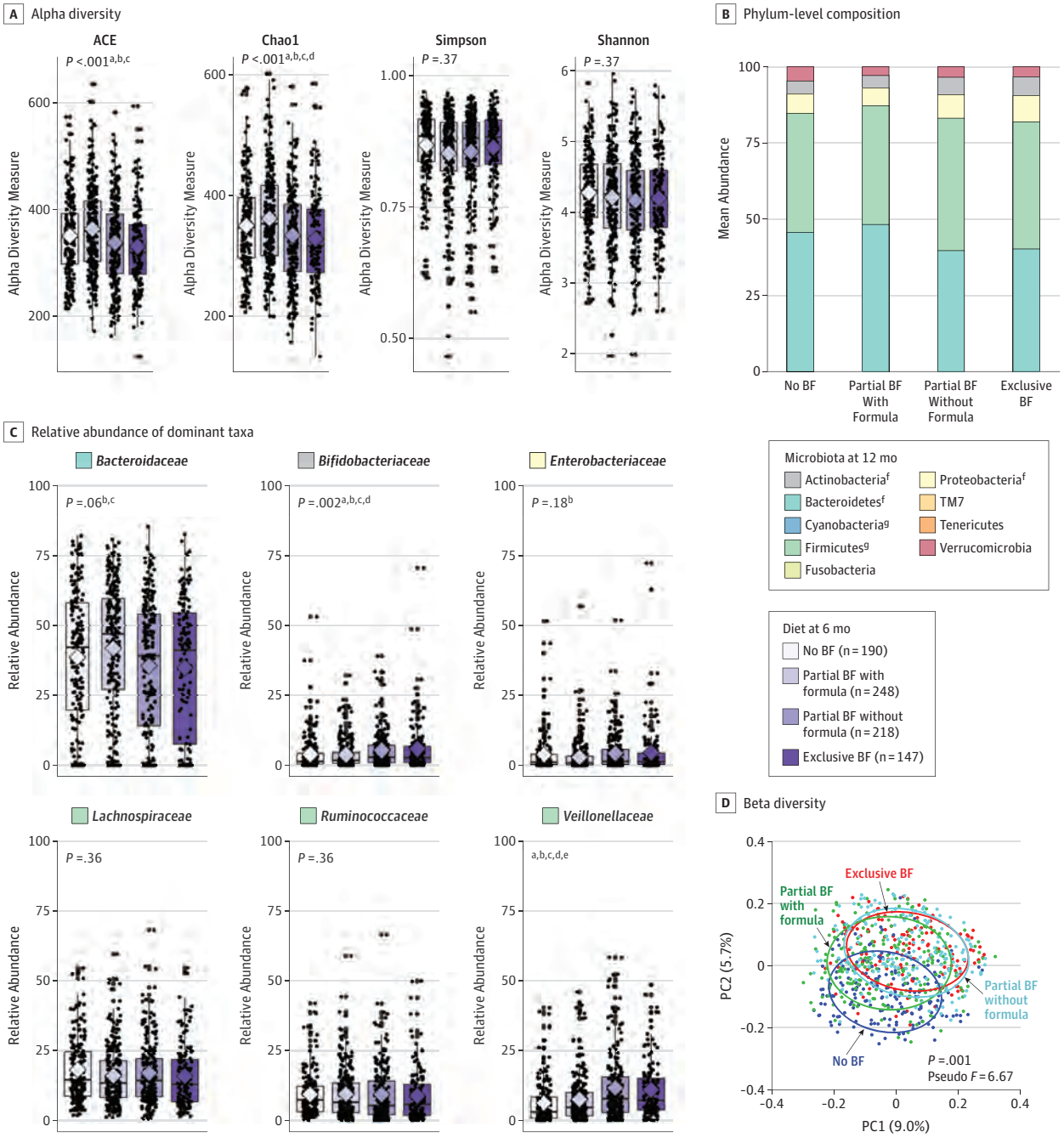
[e] Partial BF/exclusive BF;

[f] Exclusive BF after hospital/exclusive BF.

[g] $P < .001$.

[h] $P < .05$.

Figure 2. Infant Gut Microbiota at 12 Months According to Diet at 6 Months



A, Alpha diversity evaluated by richness (abundance-based coverage estimator [ACE] and Chao1) and diversity (Simpson and Shannon). Median estimates are compared across feeding groups using the Kruskal-Wallis test and Dunn post hoc tests for multiple comparisons. Boxes indicate interquartile range, lines indicate medians, diamonds indicate means, and whiskers represent range. B, Mean phylum-level composition. C, Relative abundance of dominant taxa across feeding groups. Breastfeeding (BF) refers to breast milk feeding regardless of feeding mode (at the breast or from a bottle). D, Principal coordinate analysis (PC1 and PC2) based on unweighted UniFrac distances, with community structure differences tested by permutational analysis of variance with 999 permutations.

P values represent false discovery rate–corrected *P* values testing for overall differences across the 4 feeding groups. Significant pairwise comparisons:

[a] No BF/exclusive BF;

[b] Partial BF with formula/partial BF without formula;

[c] Partial BF with formula/exclusive BF; no significant differences observed between partial BF without formula and exclusive BF;

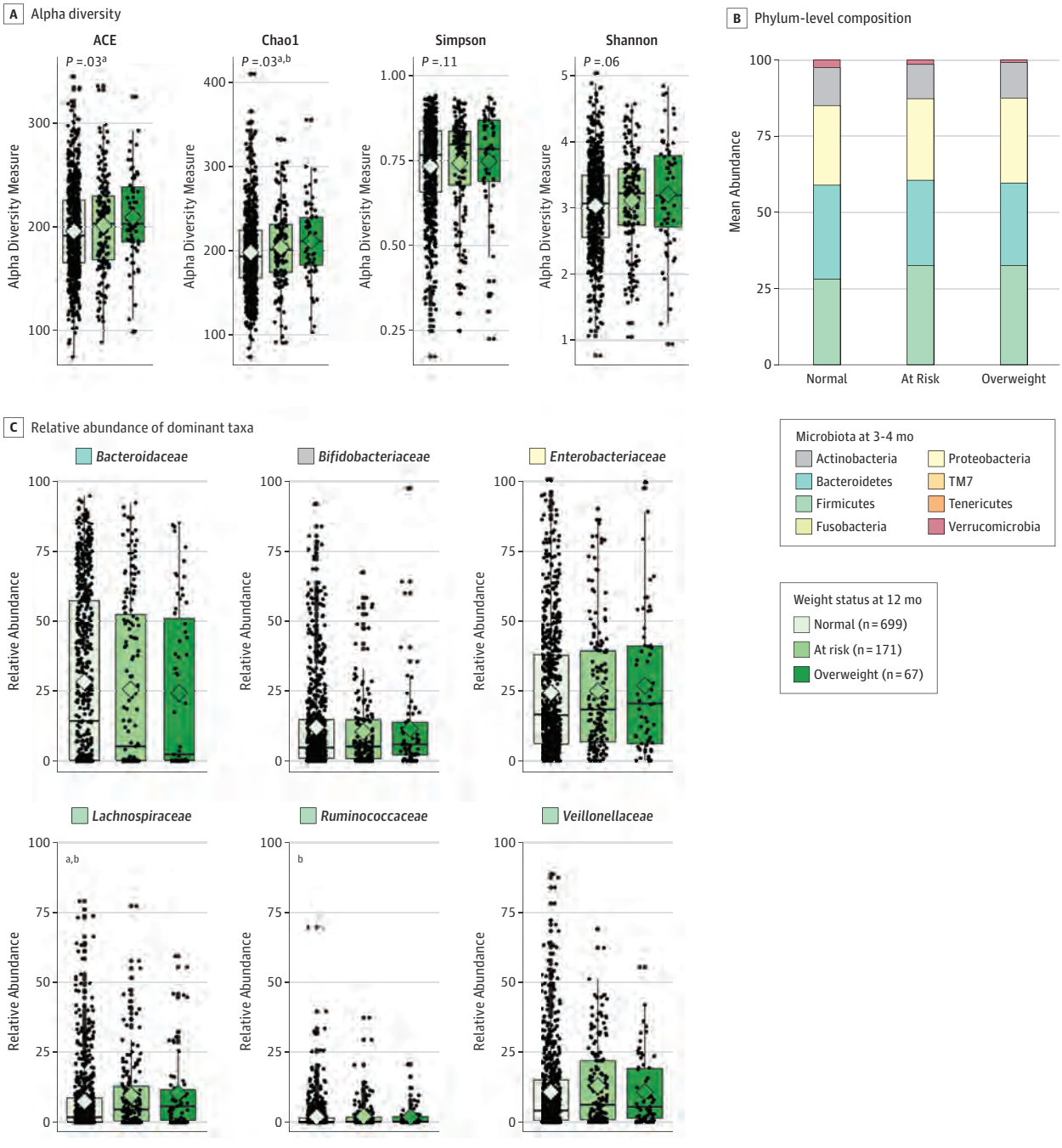
[d] No BF/partial BF without formula;

[e] No BF/partial BF with formula.

[f] $P < .01$.

[g] $P < .05$.

Figure 3. Infant Gut Microbiota Characterization at 3 Months According to Infant Weight Status at 12 Months



A, Alpha diversity evaluated by richness (abundance-based coverage estimator [ACE] and Chao1) and diversity (Simpson and Shannon). Median estimates are compared across weight status using the Kruskal-Wallis test and Dunn post hoc tests for multiple comparisons. Boxes indicate interquartile range, lines indicate medians, diamonds indicate means, and whiskers represent range. B, Mean phylum-level composition. C, Relative abundance of dominant taxa across weight status groups. Breastfeeding refers to breast milk feeding regardless of feeding mode (at the breast or from a bottle).

Significant pairwise comparisons:

[a] Normal/overweight;

[b] Normal/at risk.

at 3 to 4 months among infants who became overweight. The Firmicutes to Bacteroidetes ratio was highest in infants who became overweight at 1 year, although this difference was not significant. By 12 months, few differences in microbiota were observed according to weight status (eFigure 4 and eTable 9 in the Supplement).

To further explore the association of weight status at 12 months with the composition and diversity of gut micro-

Table 2. Association of Infant Feeding and Key Microbiota Measures at 3 and 12 Months With Weight Status at 12 Months

Breastfeeding and Microbiota Exposure	OR (95% CI) for Overweight or at Risk of Overweight (WFLz score >85th Percentile) at 12 mo					
		Mutually Adjusted				
	Adjusted for Covariates Plus Feeding or Microbiota (Individually)[a]	For Covariates, Feeding, and Chao1	For Covariates, Feeding, and Shannon	For Covariates, Feeding, and *Lachnospiraceae*	For Covariates, Feeding, and F/B Ratio	For Covariates, Feeding, and Selected Microbiota Measures[b]
Breastfeeding status at 3 mo (n = 795)						
None (formula only)	1.79 (1.09-2.93)	1.56 (0.93-2.59)	1.63 (0.98-2.70)	1.47 (0.87-2.45)	1.77 (1.07-2.91)	1.33 (0.79-2.24)
Partial (breast milk and formula)	1.49 (0.98-2.26)	1.37 (0.90-2.09)	1.41 (0.93-2.16)	1.37 (0.90-2.09)	1.52 (1.00-2.32)	1.28 (0.83-2.97)
Exclusive after hospital discharge	1.00 (0.58-1.69)	1.02 (0.59-1.73)	1.02 (0.59-1.73)	1.00 (0.58-1.69)	0.93 (0.53-1.58)	1.02 (0.59-1.73)
Exclusive (breast milk only)	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]
Microbiota measures at 3 mo (n = 795)						
Chao1 (per quartile increase)	1.25 (1.08-1.46)	1.20 (0.59-1.73)	NA	NA	NA	1.16 (0.99-1.37)
Shannon (per quartile increase)	1.18 (1.02-1.38)	NA	1.13 (0.97-1.32)	NA	NA	NA
High *Lachnospiraceae* (above median)[c]	1.82 (1.29-2.57)	NA	NA	1.66 (1.16-2.39)	NA	1.58 (1.10-2.28)
F/B ratio (per quartile increase)	1.17 (1.00-1.38)	NA	NA	NA	1.20 (1.02-1.42)	NA
Breastfeeding duration at 12 mo (n = 695)						
<6 mo	1.99 (1.23-3.22)	1.97 (1.21-3.18)	1.95 (1.20-3.15)	1.98 (1.22-3.20)	2.02 (1.25-3.27)	1.96 (1.21-3.16)
6 to <12 mo	1.59 (1.02-2.48)	1.53 (0.98-2.39)	1.57 (1.00-2.45)	1.57 (1.00-2.45)	1.60 (1.02-2.50)	1.52 (0.97-2.38)
≥12 mo	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]
Microbiota measures at 12 mo (n = 695)						
Chao1 (per quartile increase)	1.15 (0.97-1.36)	1.13 (0.96-1.35)	NA	NA	NA	1.13 (0.95-1.34)
Shannon (per quartile increase)	1.18 (1.00-1.40)	NA	1.17 (0.99-1.39)	NA	NA	NA
High *Lachnospiraceae* (above median)[c]	1.27 (0.87-1.85)	NA	NA	1.24 (0.85-1.81)	NA	1.21 (0.83-1.78)
F/B ratio (per quartile increase)	1.06 (0.90-1.26)	NA	NA	NA	1.08 (0.91-1.28)	NA

Abbreviations: F/B ratio, Firmicutes to Bacteroidetes ratio; NA, not applicable; OR, odds ratio; WFLz, weight for length z.

[a] Adjusted for maternal race/ethnicity, educational level, body mass index, smoking, cesarean delivery, dogs in household, infant sex, antibiotic exposure between 0 and 12 mo, and study site.

[b] The final model is adjusted for Chao1 and *Lachnospiraceae* because these were the strongest individual microbiota variables associated with risk of overweight; Shannon and F/B ratio were omitted to avoid multicollinearity because Shannon and Chao1 are highly correlated with each other (as 2 measures of alpha diversity), as are the F/B ratio and *Lachnospiraceae* relative abundance (*Lachnospiraceae* is a family in the Firmicutes phylum). Breastfeeding refers to breast milk feeding, regardless of feeding mode (at the breast or from a bottle). There were 795 infants for the 3-mo analyses and 695 infants for the 12-mo analyses.

[c] High relative abundance of *Lachnospiraceae*.

biota, we classified candidate microbiota measures in quartiles and conducted logistic regression analyses (eFigure 5 in the Supplement). At 3 to 4 months, higher relative abundance of *Lachnospiraceae* (above vs below median) were associated with an 89% increase in risk of overweight by 12 months (OR, 1.89; 95% CI, 1.40-2.56). Each quartile increase in the Firmicutes to Bacteroidetes ratio was associated with a 12% increase in the risk of overweight (OR, 1.12; 95% CI, 0.98-1.28). The richness of gut microbiota was also positively associated with the risk of overweight by 12 months (OR, 1.24 per quartile increase; 95% CI, 1.09-1.42 per quartile increase), as was the diversity of gut microbiota (OR, 1.21 per quartile increase; 95% CI, 1.06-1.38 per quartile increase). No comparable associations were detected for microbiota measures at 12 months.

Contribution of Gut Microbiota to Association of Infant Feeding Practices and Overweight

To examine whether gut microbiota contribute to the increased risk of overweight associated with formula feeding and shorter duration of breastfeeding, we tested these associations in mutually adjusted models. Adjustment for richness of microbiota, diversity of microbiota, or relative abundance of *Lachnospiraceae* substantially attenuated the effect estimate for cessation of breastfeeding before 3 months (**Table 2**). Simultaneous adjustment for richness of microbiota and *Lachnospiraceae* attenuated this estimate from 2.04 (95% CI, 1.25-3.32) to 1.33 (95% CI, 0.79-2.24). In contrast, associations between infant feeding and weight status were largely unaffected by adjustment for concurrent microbiota measures at 12 months.

Discussion

Our findings demonstrate a strong inverse and dose-dependent association between breastfeeding and the risk of overweight in the first year of life that is partially explained by gut microbiota. Although the effect of breast milk on the development of the gut microbiome is well known,[24-27] our

findings address important nuances that, to our knowledge, have not been explored in previous studies, identifying differences according to the type and timing of supplemental feeding. We also report novel longitudinal associations between the composition of gut microbiota at 3 to 4 months of age and weight status at 12 months of age.

Similar to previous studies,[28,29] we found a 63% increased risk of overweight among infants who were partially vs exclusively breastfed at 3 months and a 102% increased risk among exclusively formula-fed infants. As others have reported,[25,27,30] we detected significantly lower bacterial richness and diversity in breastfed infants, accompanied by enrichment of several taxa (eg, *Bifidobacteriaceae*, *Pasteurellaceae*, and *Enterobacteriaceae*) and depletion of others (eg, *Bacteroidaceae* and *Lachnospiraceae*), with dose effects according to the degree of breastfeeding exclusivity. These findings are consistent with evidence that human milk oligosaccharides function as selective substrates for particular groups of microorganisms, including *Bifidobacteriaceae*.[31-34]

Building on previous studies of adults,[35,36] children,[9-12] and infants,[37-42] our study provides new evidence linking gut microbiota with the risk of overweight in the first year of life. Prior research of infants has reported reduced relative abundance of Bifidobacteria and enrichment of streptococci and *Bacteroides fragilis* to be associated with overweight later in childhood.[37-42] Although we did not observe these particular trends, perhaps owing to cohort differences in age, geography, or feeding practices (eg, extremely high rates of initiation of breastfeeding in the CHILD Study), we identified several novel associations. Although few associations were detected between microbiota and overweight measured concurrently at 12 months, several microbiota features associated with overweight were identified at 3 to 4 months. For example, while *Lachnospiraceae* were similarly abundant in normal-weight and overweight infants at 12 months, they were significantly enriched among overweight infants at 3 to 4 months. *Lachnospiraceae* has been associated with maternal obesity and is enriched in meconium from neonates born to mothers with diabetes.[43] In our study, enrichment of *Lachnospiraceae* was associated with exposure to formula in a dose-dependent manner, along with the richness and diversity of microbiota; adjustment for these microbiota features partially explained the association between exposure to formula and the risk of overweight.

Taken together, our results suggest that the transient perturbation of microbiota in early infancy (related to feeding practices or other exposures) may influence weight gain and body composition, which may ultimately influence the risk of metabolic disease risk later in life.[44] This hypothesis (eFigure 6 in the Supplement) is consistent with studies of mice showing that the disruption of gut microbiota limited to early life has permanent metabolic effects, including elevated adiposity, despite "recovery" of the microbiota.[45] Other important mechanisms linking gut microbiota and obesity include microbial metabolites influencing levels of and sensitivity to the satiety hormone leptin.[46,47]

To our knowledge, this is the first study to evaluate the potential association of brief exposure to formula during the neonatal period as it pertains to the development of microbiota and the risk of overweight. These are clinically important questions since many neonates receive formula in the hospital, often without medical indication,[48] yet the effect of this brief intervention on the developing microbiota (and related clinical outcomes) is not known. In our cohort, 179 of 579 infants (30.9%) reported by their mothers as exclusively breastfed actually received some formula in the hospital. Overall, we found no difference in the risk of overweight among these infants. However, while their microbiota profiles at 3 to 4 months were clearly more similar to those of exclusively breastfed than partially or nonbreastfed infants, some significant differences were detected. The richness and diversity of the microbiota were lower, as was the relative abundance of *Bifidobacteriaceae*, suggesting that even brief exposure to formula may disrupt normal colonization of the infant gut. We have likely underestimated this disruption, since our first sample was not collected until 3 to 4 months after hospital discharge. It is possible that the reason for formula supplementation contributed to the observed microbiota differences, but this possibility could not be directly examined in our study because we did not systematically document reasons for supplementation.

Multiple studies have investigated the effects of breast milk on the gut microbiome[24-26,34,49,50]; however, many of these studies did not distinguish between partial breastfeeding mixed with formula vs mixed with foods. We found that breastfed infants supplemented with formula were more similar to nonbreastfed infants, whereas breastfed infants given complementary foods (without formula) were more similar to exclusively breastfed infants. These differences might explain why mixed feeding with (but not without) formula was associated with an increased risk of overweight, although more research is needed to characterize these complex associations.

Strengths and Limitations

The strengths of our study include the detailed description of infant feeding practices, repeated analysis of microbiota, and adjustment for multiple confounders. However, we lacked information about the reasons for supplementation and did not address the mode of breast milk feeding, type of formula, quantity of breast milk or formula intake, or breast milk composition. Finally, a limitation of 16S ribosomal RNA analysis is that it cannot quantify or accurately resolve individual bacterial species.

Conclusions

Our findings indicate that breastfeeding is protective against overweight and suggest that the gut microbiota contribute to this effect. Formula feeding was associated with higher microbiota diversity and enrichment of *Lachnospiraceae* at 3 to 4 months, and these microbiota features partially explained the increased risk of overweight among nonbreastfed infants. Subtle but statistically significant differences in the microbiota were observed after brief exposure to formula in the hospital, although the clinical implications of these changes are unclear. Together, these results identify important areas for future research and emphasize the importance of early infancy as a critical period during which transient gut dysbiosis is associated with the subsequent risk of overweight.

ARTICLE INFORMATION

Accepted for Publication: April 5, 2018.

Published Online: June 4, 2018. doi:10.1001/jamapediatrics.2018.1161

Correction: This article was corrected on July 2, 2018, to fix an error in a color key in all 3 figures.

Author Affiliations: Department of Pediatrics, University of Alberta, Edmonton, Alberta, Canada (Forbes, Tun, Mandhane, Kozyrskyj); Developmental Origins of Chronic Diseases in Children Network (DEVOTION), Department of Pediatrics and Child Health, Children's Hospital Research Institute of Manitoba, University of Manitoba, Winnipeg, Manitoba, Canada (Forbes, Azad, Vehling, Becker); National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Manitoba, Canada (Forbes); Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada (Konya, Scott); Department of Cell and Systems Biology, University of Toronto, Toronto, Ontario, Canada (Guttman); Department of Agriculture, Food and Nutritional Sciences, University of Alberta, Edmonton, Alberta, Canada (Field); Department of Medicine, McMaster University, Hamilton, Ontario, Canada (Lefebvre, Sears); Department of Pediatrics, University of British Columbia, Vancouver, British Columbia, Canada (Turvey); Child and Family Research Institute, BC Children's Hospital, Vancouver, British Columbia, Canada (Turvey); Department of Pediatrics and Physiology, Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada (Moraes, Subbarao).

Author Contributions: Drs Azad and Kozyrskyj had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Drs Forbes and Azad are co-first authors, and Drs Scott and Kozyrskyj are co-last authors.
Study concept and design: Azad, Becker, Mandhane, Turvey, Subbarao, Scott, Kozyrskyj.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Forbes, Azad.
Critical revision of the manuscript for important intellectual content: Azad, Vehling, Tun, Konya, Guttman, Field, Lefebvre, Sears, Becker, Mandhane, Turvey, Moraes, Subbarao, Scott, Kozyrskyj.
Statistical analysis: Forbes, Azad, Tun, Scott.
Obtained funding: Sears, Becker, Mandhane, Turvey, Subbarao, Scott, Kozyrskyj.
Administrative, technical, or material support: Tun, Konya, Guttman, Field, Lefebvre, Sears, Mandhane, Turvey, Moraes, Scott, Kozyrskyj.
Study supervision: Azad, Guttman, Turvey, Scott, Kozyrskyj.

Conflict of Interest Disclosures: Dr Azad reported holding a Canada Research Chair in the Developmental Origins of Chronic Disease. Dr Sears reported holding the AstraZeneca endowed chair in respiratory epidemiology. No other disclosures were reported.

Funding/Support: The Canadian Institutes of Health Research (CIHR) and the Allergy, Genes and Environment Network of Centres of Excellence provided core funding for the Canadian Healthy Infant Longitudinal Development (CHILD) Study. This research was specifically funded by CIHR Microbiome Initiative team grant 227312 and the Children's Hospital Foundation of Manitoba. Additional funding was provided by Health Canada, Environment Canada, Canada Mortgage and Housing Corporation, the Sick Children's Hospital Foundation, Don & Debbie Morrison, the Silver Thread Foundation, the Childhood Asthma Foundation, Research Manitoba, and the Government of Manitoba (Healthy Child Manitoba Office).

Role of the Funder/Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Group Information: The CHILD Study investigators include Padmaja Subbarao, MD, MSc (director), The Hospital for Sick Children, Toronto, Ontario, Canada; Stuart E. Turvey, MBBS, DPhil (co-director), University of British Columbia, Vancouver, British Columbia, Canada; Malcolm R. Sears, MBChB (founding director), McMaster University, Hamilton, Ontario, Canada; Sonia S. Anand, MD, PhD, McMaster University, Hamilton, Ontario, Canada; Meghan B. Azad, PhD, University of Manitoba, Winnipeg, Manitoba, Canada; Allan B. Becker, MD, University of Manitoba, Winnipeg, Manitoba, Canada; A. Dean Befus, PhD, University of Alberta, Edmonton, Alberta, Canada; Michael Brauer, ScD, University of British Columbia, Vancouver, British Columbia, Canada; Jeff R. Brook, PhD, University of Toronto, Toronto, Ontario, Canada; Edith Chen, PhD, Northwestern University, Chicago, Illinois; Michael M. Cyr, MD, McMaster University, Hamilton, Ontario, Canada; Denise Daley, PhD, University of British Columbia, Vancouver, British Columbia, Canada; Sharon D. Dell, MD, The Hospital for Sick Children, Toronto, Ontario, Canada; Judah A. Denburg, MD, McMaster University, Hamilton, Ontario, Canada; Qing L. Duan, PhD, Queen's University, Kingston, Ontario, Canada; Thomas Eiwegger, MD, The Hospital for Sick Children, Toronto, Ontario, Canada; Hartmut Grasemann, MD, The Hospital for Sick Children, Toronto, Ontario, Canada; Kent HayGlass, PhD, University of Manitoba, Winnipeg, Manitoba, Canada; Richard G. Hegele, MD, PhD, The Hospital for Sick Children, Toronto, Ontario, Canada; D. Linn Holness, MD, University of Toronto, Toronto, Ontario, Canada; Perry Hystad, PhD, Oregon State University, Corvallis; Michael Kobor, PhD, University of British Columbia, Vancouver, British Columbia, Canada; Tobias R. Kollmann, MD, PhD, University of British Columbia, Vancouver, British Columbia, Canada; Anita L. Kozyrskyj, PhD, University of Alberta, Edmonton, Alberta, Canada; Catherine Laprise, PhD, Université du Québec à Chicoutimi, Chicoutimi, Québec, Canada; Wendy Y.W. Lou, PhD, University of Toronto, Toronto, Ontario, Canada; Joseph Macri, PhD, McMaster University, Hamilton, Ontario, Canada; Piushkumar J. Mandhane, MD, PhD, University of Alberta, Edmonton, Alberta, Canada; Greg Miller, PhD, Northwestern University, Chicago, Illinois; Theo J. Moraes, MD, PhD, The Hospital for Sick Children, Toronto, Ontario, Canada; Peter Pare, MD, University of British Columbia, Vancouver, British Columbia, Canada; Clare Ramsey, MD, University of Manitoba, Winnipeg, Manitoba, Canada; Felix Ratjen, MD, The Hospital for Sick Children, Toronto, Ontario, Canada; Andrew Sandford, PhD, University of British Columbia, Vancouver, British Columbia, Canada; James A. Scott, PhD, University of Toronto, Toronto, Ontario, Canada; Jeremy Scott, PhD, University of Toronto, Toronto, Ontario, Canada; Frances Silverman, PhD, University of Toronto, Toronto, Ontario, Canada; Elinor Simons, MD, PhD, University of Manitoba, Winnipeg, Manitoba, Canada; Tim Takaro, MD, Simon Fraser University, Burnaby, British Columbia, Canada; Scott J. Tebbutt, PhD, University of British Columbia, Vancouver, British Columbia, Canada; and Theresa To, PhD, The Hospital for Sick Children, Toronto, Ontario, Canada.

Additional Contributions: We thank all the families who took part in this study and the entire CHILD team, which includes interviewers, nurses, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, and receptionists. We also acknowledge the generosity of ALK-Abello, Mississauga, Ontario, Canada, in supplying all allergens for the study, and Lincoln Diagnostics Inc, Decatur, Illinois, for supplying the Duotip-Test II devices and skin testing kits.

REFERENCES

1. Ojha S, Budge H. Early origins of obesity and developmental regulation of adiposity. In: Symonds ME, ed. *Adipose Tissue Biology*. New York, NY: Springer International Publishing; 2017:427-456.

2. Wang L, Collins C, Ratliff M, Xie B, Wang Y. Breastfeeding reduces childhood obesity risks. *Child Obes*. 2017;13(3):197-204.

3. Marseglia L, Manti S, D'Angelo G, et al. Obesity and breastfeeding: the strength of association. *Women Birth*. 2015;28(2):81-86.

4. Rosenbaum M, Knight R, Leibel RL. The gut microbiota in human energy homeostasis and obesity. *Trends Endocrinol Metab*. 2015;26(9):493-501.

5. Ley RE. Obesity and the human microbiome. *Curr Opin Gastroenterol*. 2010;26(1):5-11.

6. Le Chatelier E, Nielsen T, Qin J, et al; MetaHIT Consortium. Richness of human gut microbiome correlates with metabolic markers. *Nature*. 2013;500(7464):541-546.

7. Koliada A, Syzenko G, Moseiko V, et al. Association between body mass index and Firmicutes/Bacteroidetes ratio in an adult Ukrainian population. *BMC Microbiol*. 2017;17(1):120.

8. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*. 2006;444(7122):1027-1031.

9. Riva A, Borgo F, Lassandro C, et al. Pediatric obesity is associated with an altered gut microbiota and discordant shifts in Firmicutes populations. *Environ Microbiol*. 2017;19(1):95-105.

10. Borgo F, Verduci E, Riva A, et al. Relative abundance in bacterial and fungal gut microbes in obese children: a case control study. *Child Obes*. 2017;13(1):78-84.

11. Ignacio A, Fernandes MR, Rodrigues VAA, et al. Correlation between body mass index and faecal microbiota from children. *Clin Microbiol Infect*. 2016;22(3):258.e1-258.e8.

12. Lemas DJ, Yee S, Cacho N, et al. Exploring the contribution of maternal antibiotics and breastfeeding to development of the infant microbiome and pediatric obesity. *Semin Fetal Neonatal Med*. 2016;21(6):406-409.

13. Cerdó T, Ruiz A, Campoy C. Human gut microbiota and obesity during development. In: Gordeladze J, ed. *Adiposity—Omics and Molecular Understanding*. London, United Kingdom: InTechOpen; 2017:265-285.

14. Subbarao P, Anand SS, Becker AB, et al; CHILD Study investigators. The Canadian Healthy Infant Longitudinal Development (CHILD) Study: examining developmental origins of allergy and asthma. *Thorax*. 2015;70(10):998-1000.

15. WHO Multicentre Growth Reference Study Group. *WHO Child Growth Standards: Length/Height-for-Age, Weight-for-Age, Weight-for-Length, Weight-for-Height and Body Mass Index-for-Age: Methods and Development*. Geneva, Switzerland: World Health Organization; 2006.

16. Corby L, Secker D. Growth monitoring of infants and children using the 2006 World Health Organization [WHO] Child Growth Standards and 2007 WHO growth references: practice-based evidence in nutrition. *Acta Paediatr*. 2006;(suppl 450):76-85.

17. Azad MB, Konya T, Persaud RR, et al; CHILD Study Investigators. Impact of maternal intrapartum antibiotics, method of birth and breastfeeding on gut microbiota during the first year of life: a prospective cohort study. *BJOG*. 2016;123(6):983-993.

18. Azad MB, Sharma AK, de Souza RJ, et al; Canadian Healthy Infant Longitudinal Development Study Investigators. Association between artificially sweetened beverage consumption during pregnancy and infant body mass index. *JAMA Pediatr*. 2016;170(7):662-670.

19. Guenther PM, Casavale KO, Reedy J, et al. Update of the Healthy Eating Index: HEI-2010. *J Acad Nutr Diet*. 2013;113(4):569-580.

20. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7(5):335-336.

21. DeSantis TZ, Hugenholtz P, Larsen N, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol*. 2006;72(7):5069-5072.

22. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45(3):1-67. doi:10.18637/jss.v045.i03

23. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol*. 2005;71(12):8228-8235.

24. Paolella G, Vajro P. Childhood obesity, breastfeeding, intestinal microbiota, and early exposure to antibiotics: what is the link? *JAMA Pediatr*. 2016;170(8):735-737.

25. Pannaraj PS, Li F, Cerini C, et al. Association between breast milk bacterial communities and establishment and development of the infant gut microbiome. *JAMA Pediatr*. 2017;171(7):647-654.

26. Timmerman HM, Rutten NBMM, Boekhorst J, et al. Intestinal colonisation patterns in breastfed and formula-fed infants during the first 12 weeks of life reveal sequential microbiota signatures. *Sci Rep*. 2017;7(1):8327.

27. Milani C, Duranti S, Bottacini F, et al. The first microbial colonizers of the human gut: composition, activities, and health implications of the infant gut microbiota. *Microbiol Mol Biol Rev*. 2017;81(4):e00036-e17.

28. Yeung H, Leff M, Rhee KE. Effect of exclusive breastfeeding among overweight and obese mothers on infant weight-for-length percentile at 1 year. *Breastfeed Med*. 2017;12:39-47.

29. Bider-Canfield Z, Martinez MP, Wang X, et al. Maternal obesity, gestational diabetes, breastfeeding and childhood overweight at age 2 years. *Pediatr Obes*. 2017;12(2):171-178.

30. Castanys-Muñoz E, Martin MJ, Vazquez E. Building a beneficial microbiome from birth. *Adv Nutr*. 2016;7(2):323-330.

31. Sela DA, Mills DA. Nursing our microbiota: molecular linkages between bifidobacteria and milk oligosaccharides. *Trends Microbiol*. 2010;18(7):298-307.

32. Marcobal A, Barboza M, Froehlich JW, et al. Consumption of human milk oligosaccharides by gut-related microbes. *J Agric Food Chem*. 2010;58(9):5334-5340.

33. Azad MB, Konya T, Maughan H, et al; CHILD Study Investigators. Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. *CMAJ*. 2013;185(5):385-394.

34. Laursen MF, Andersen LBB, Michaelsen KF, et al. Infant gut microbiota development is driven by transition to family foods independent of maternal obesity. *mSphere*. 2016;1(1):e00069-e15.

35. Boulangé CL, Neves AL, Chilloux J, Nicholson JK, Dumas M-E. Impact of the gut microbiota on inflammation, obesity, and metabolic disease. *Genome Med*. 2016;8(1):42.

36. Fernandes J, Su W, Rahat-Rozenbloom S, Wolever TM, Comelli EM. Adiposity, gut microbiota and faecal short chain fatty acids are linked in adult humans. *Nutr Diabetes*. 2014;4:e121.

37. Koleva PT, Bridgman SL, Kozyrskyj AL. The infant gut microbiome: evidence for obesity risk and dietary intervention. *Nutrients*. 2015;7(4):2237-2260.

38. Scheepers LE, Penders J, Mbakwa CA, Thijs C, Mommers M, Arts ICW. The intestinal microbiota composition and weight development in children: the KOALA Birth Cohort Study. *Int J Obes (Lond)*. 2015;39(1):16-25.

39. White RA, Bjørnholt JV, Baird DD, et al. Novel developmental analyses identify longitudinal patterns of early gut microbiota that affect infant growth. *PLoS Comput Biol*. 2013;9(5):e1003042.

40. Vael C, Verhulst SL, Nelen V, Goossens H, Desager KN. Intestinal microflora and body mass index during the first three years of life: an observational study. *Gut Pathog*. 2011;3(1):8.

41. Kalliomäki M, Collado MC, Salminen S, Isolauri E. Early differences in fecal microbiota composition in children may predict overweight. *Am J Clin Nutr*. 2008;87(3):534-538.

42. Korpela K, Zijlmans MAC, Kuitunen M, et al. Childhood BMI in relation to microbiota in infancy and lifetime antibiotic use. *Microbiome*. 2017;5(1):26.

43. Singh S, Karagas MR, Mueller NT. Charting the maternal and infant microbiome: what is the role of diabetes and obesity in pregnancy? *Curr Diab Rep*. 2017;17(2):11.

44. Roy SM, Spivack JG, Faith MS, et al. Infant BMI or weight-for-length and obesity risk in early childhood. *Pediatrics*. 2016;137(5):e20153492-e20153492.

45. Cox LM, Yamanishi S, Sohn J, et al. Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. *Cell*. 2014;158(4):705-721.

46. Cani PD, Lecourt E, Dewulf EM, et al. Gut microbiota fermentation of prebiotics increases satietogenic and incretin gut peptide production with consequences for appetite sensation and glucose response after a meal. *Am J Clin Nutr*. 2009;90(5):1236-1243.

47. Schéle E, Grahnemo L, Anesten F, Hallén A, Bäckhed F, Jansson JO. The gut microbiota reduces leptin sensitivity and the expression of the obesity-suppressing neuropeptides proglucagon (Gcg) and brain-derived neurotrophic factor (Bdnf) in the central nervous system. *Endocrinology*. 2013;154(10):3643-3651.

48. Chantry CJ, Dewey KG, Peerson JM, Wagner EA, Nommsen-Rivers LA. In-hospital formula use increases early breastfeeding cessation among first-time mothers intending to exclusively breastfeed. *J Pediatr*. 2014;164(6):1339-1345.e5.

49. Amarri S, Benatti F, Callegari ML, et al. Changes of gut microbiota and immune markers during the complementary feeding period in healthy breast-fed infants. *J Pediatr Gastroenterol Nutr*. 2006;42(5):488-495.

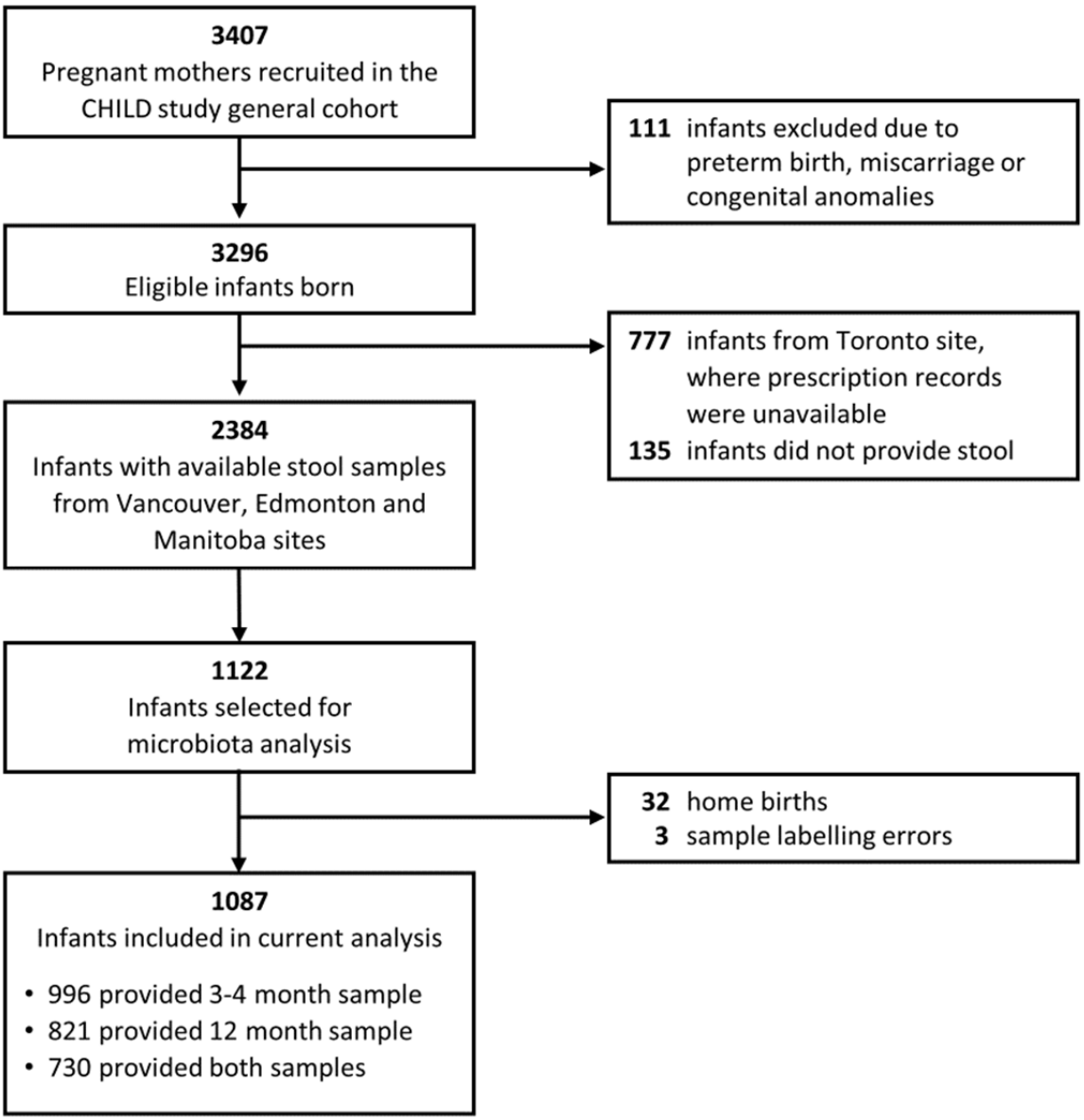
50. Qasem W, Azad MB, Hossain Z, et al. Assessment of complementary feeding of Canadian infants: effects on microbiome & oxidative stress, a randomized controlled trial. *BMC Pediatr*. 2017;17(1):54.

Supplementary Online Content

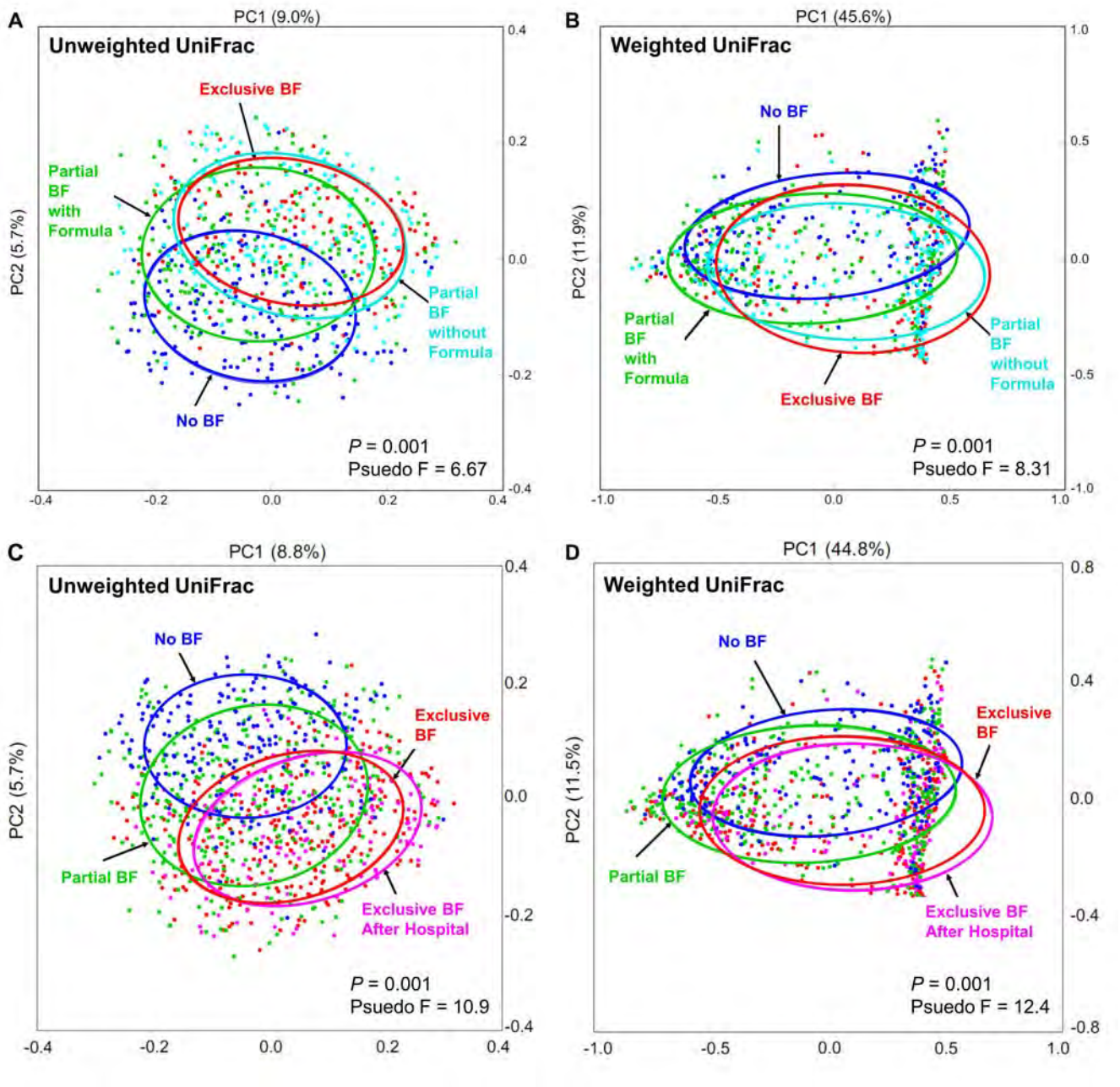
Forbes JD, Azad MB, Vehling L, et al; Canadian Healthy Infant Longitudinal Development (CHILD) Study investigators. Association of exposure to formula in the hospital and subsequent infant feeding practices with gut microbiota and risk of overweight in the first year of life. *JAMA Pediatr.* Published online June 4, 2018. doi:10.1001/jamapediatrics.2018.1161

This supplementary material has been provided by the authors to give readers additional information about their work.

eFigure 1. Flow Diagram Summarizing Selection of CHILD Study Infants Included in the Current Analysis

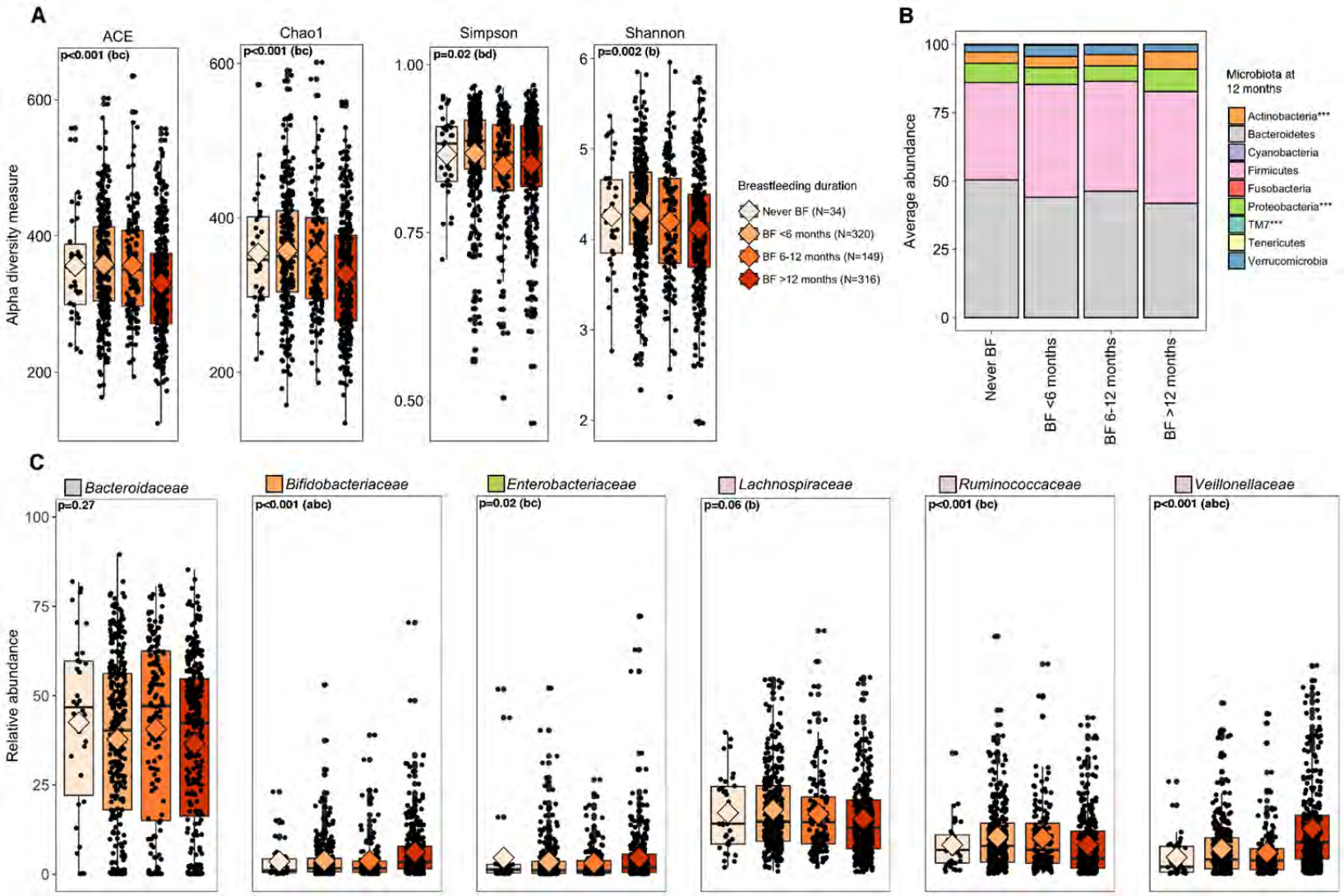


eFigure 2. Microbial Community Structure of 3-Month and 12-Month Microbiota Based on Breastfeeding Status at 3-4 Months and Infant Diet at 6 months, Respectively, as Measured by Principal Components Analysis and tested by PERMANOVA.



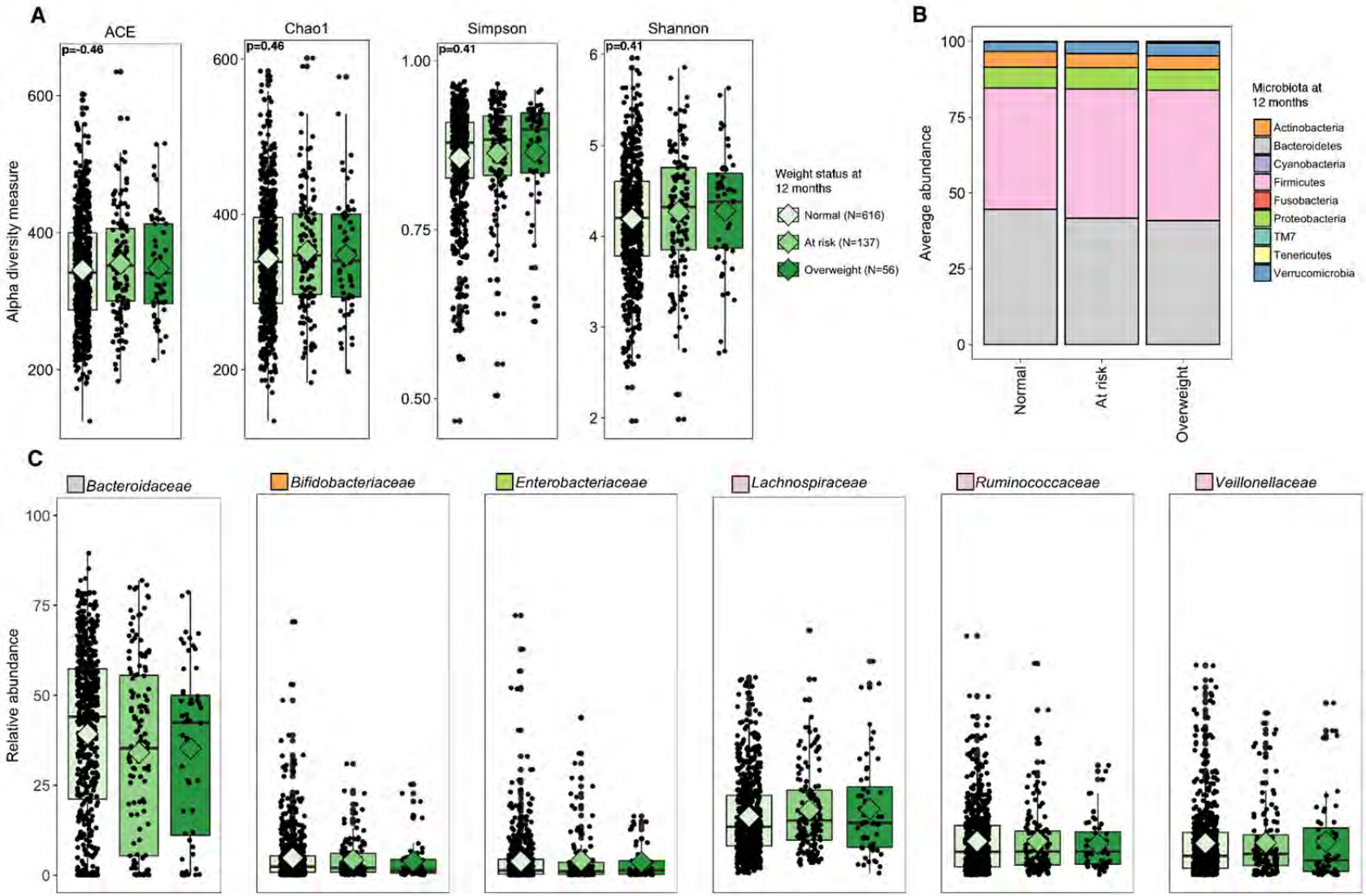
Principal coordinate analysis based on (A,C) unweighted or (B,D) weighted Unifrac distances, with community structure differences tested by PERMANOVA with 999 permutations.

eFigure 3. Infant Gut Microbiota at 12 Months According to Breastfeeding (BF) Duration[#]



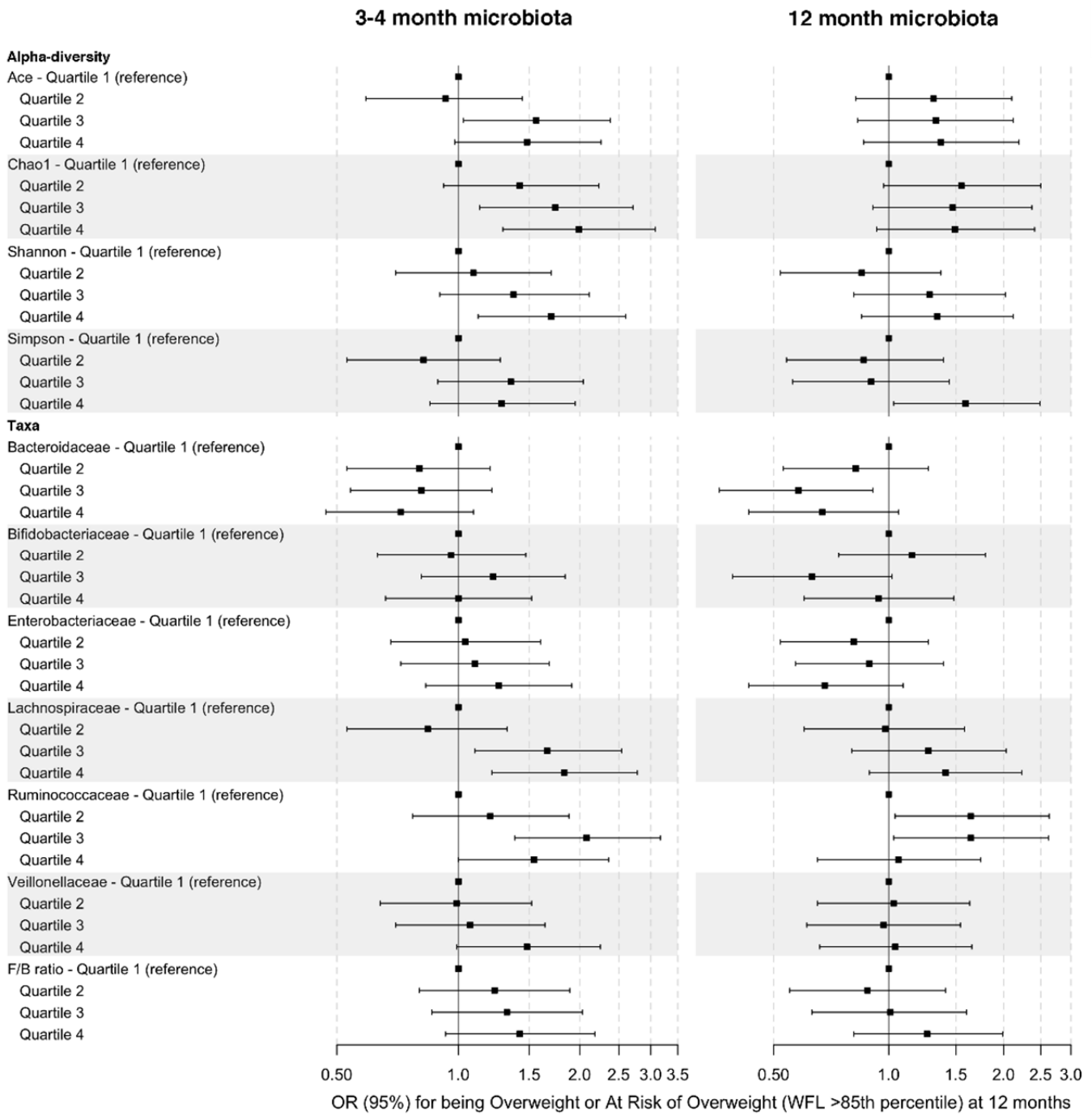
(A) Alpha diversity evaluated by richness (ACE, Chao1) and diversity (Simpson, Shannon). Median estimates compared across feeding groups using the Kruskal–Wallis test and Dunn's post-hoc tests for multiple comparisons. Boxes indicate interquartile range, lines indicate medians, diamonds indicate means and whiskers represent range. P-values represent overall FDR corrected P-values: *p<0.05, **p<0.01, ***p<0.001. (B) Mean phylum-level composition. (C). Relative abundance of dominant taxa across feeding groups. Significant pairwise comparisons for panels: [a]BF < 6 months/BF > 12 months; [b]BF 6 – 12 months/BF > 12 months; [c]Never BF/BF > 12 months; no significant differences were observed between BF <6 months/BF 6 – 12 months, Never BF/BF < 6 months or Never BF/BF 6 – 12 months. Breastfeeding refers to breast milk feeding, regardless of feeding mode (at the breast or from a bottle).

eFigure 4. Infant Gut Microbiota Characterization at 12 Months According to Infant Weight Status at 12 Months



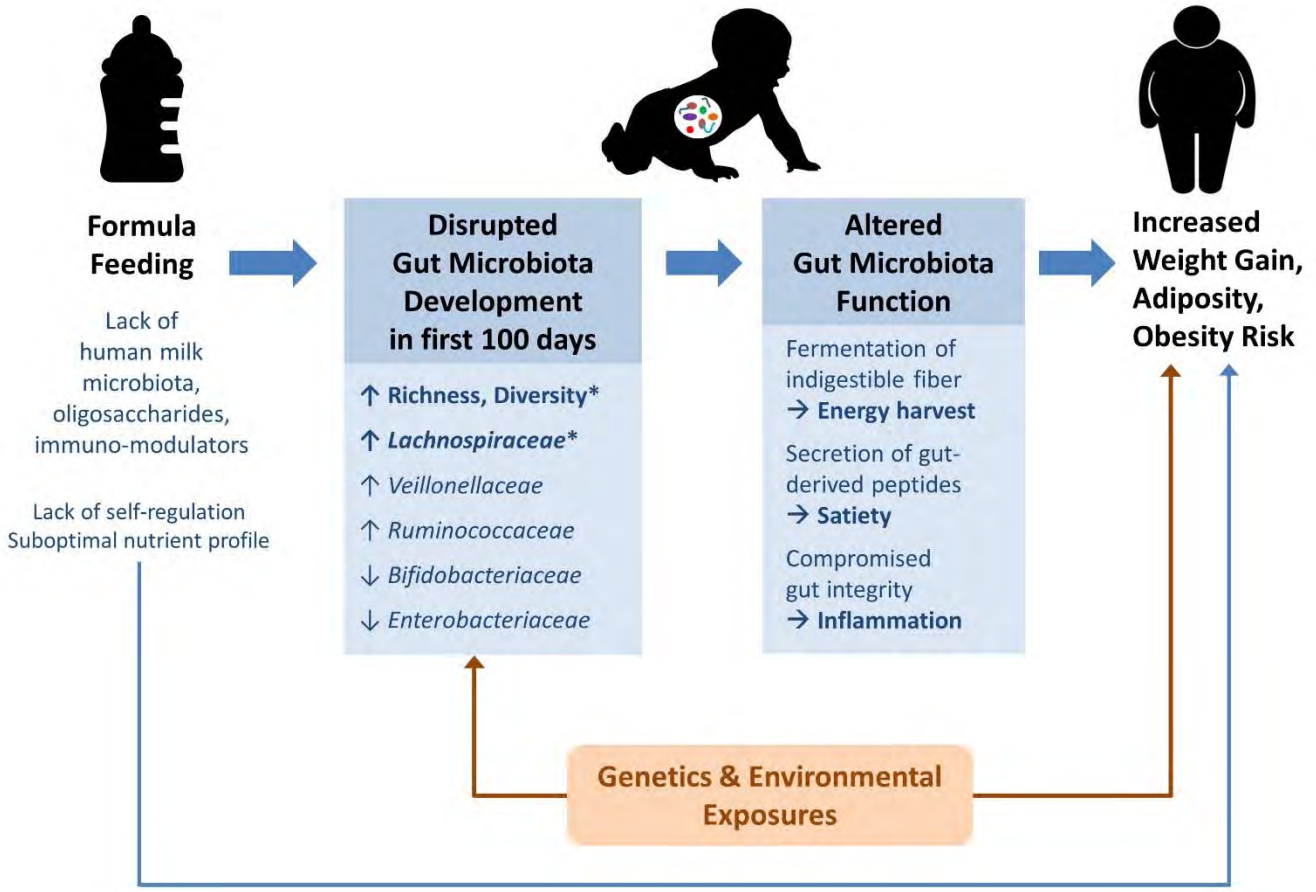
(A) Alpha diversity evaluated by richness (ACE, Chao1) and diversity (Simpson, Shannon). Median estimates compared across weight status using the Kruskal–Wallis test and Dunn's post-hoc tests for multiple comparisons. Boxes indicate interquartile range, lines indicate medians, diamonds indicate means and whiskers represent range. P-values represent overall FDR corrected P-values: *p<0.05, **p<0.01, ***p<0.001. (B) Mean phylum-level composition. (C). Relative abundance of dominant taxa across weight status groups. Significant pairwise comparisons: [a]Normal/At Risk; [b]Normal/Overweight; [c]At Risk/Overweight. Breastfeeding refers to breast milk feeding, regardless of feeding mode (at the breast or from a bottle).

eFigure 5. Association of Key Microbiota Measures at 3 and 12 Months With Infant Weight Status at 12 Months



F/B ratio, Firmicutes/Bacteroidetes ratio; CI, confidence interval; OR, unadjusted odds ratio.

eFigure 6. Associations and Hypothesized Mechanisms Linking Infant Feeding Practices, Gut Microbiota and Obesity



Results from this study show that formula-feeding is strongly associated with increased overweight risk at 1 year of age, and provide evidence that early changes in the gut microbiota (i.e. during the first 3-4 months of life) contribute to this association. Infant formulas differ from human milk in composition and bioactivity, causing changes in gut microbial communities that likely lead to altered metabolic networks affecting energy harvest, satiety and inflammation. These physiological changes, along with host genotype and environmental exposures, influence infant weight gain, adiposity and obesity risk later in life. Aside from these microbiota-related pathways, formula may influence weight gain through other mechanisms related to its composition (eg. higher protein content compared to human milk) or delivery to the infant (e.g. bottle feeding may discourage self-regulation compared to suckling at the breast). *Indicates microbiota features significantly associated with both formula feeding and risk of overweight in this study.

eTable 1. Characteristics of Participants Included in the Current Study and the General CHILD Cohort

	Subset for current study (N=1087)		General CHILD cohort (N=3296)	
Breastfeeding duration	10.26	[6.81]	10.39	[6.81]
Exclusive breastfeeding duration	2.90	[2.37]	3.16	[2.31]
Weight for length (WFL) z-score	0.29	[1.08]	0.25	[1.04]
Maternal race				
Asian	164	(15.2)	508	(15.7)
Caucasian	817	(75.8)	2359	(72.9)
FN	48	(4.5)	143	(4.4)
Other	49	(4.5)	225	(7.0)
Maternal postsecondary degree				
No	230	(21.8)	746	(23.7)
Yes	823	(78.2)	2407	(76.3)
Maternal smoking in pregnancy				
No	967	(91.5)	2897	(90.9)
Yes	90	(8.5)	290	(9.1)
Maternal Healthy Eating Index				
< 70	306	(30.4)	990	(33.1)
70 to 75	240	(23.9)	682	(22.8)
> 75	459	(45.7)	1323	(44.2)
Maternal weight class				
Underweight	31	(3.0)	103	(3.5)
Normal	586	(57.2)	1760	(59.9)
Overweight	230	(22.4)	633	(21.6)
Obese	178	(14.4)	440	(15.0)
Dog in home				
No	687	(66.9)	2142	(69.7)
Yes	340	(33.1)	930	(30.3)
Cat in home				
No	745	(72.6)	2308	(75.2)
Yes	281	(27.4)	762	(24.8)
Older siblings				
No	569	(52.3)	1772	(53.9)
Yes	518	(47.7)	1519	(46.2)
Infant sex				
Male	580	(53.4)	1726	(52.7)
Female	507	(46.6)	1550	(47.3)
Birth weight (g)				
< 3000	162	(15.2)	527	(16.5)
3000 to < 3500	419	(39.4)	1243	(38.9)
3500 to < 4000	350	(32.9)	1024	(32.0)
4000 +	133	(12.5)	403	(12.6)
Birth mode				
CS-Elective	121	(11.4)	346	(10.8)
CS-Emergency	153	(14.4)	466	(14.6)
Vaginal, IAP	237	(22.3)	688	(21.5)
Vaginal, no IAP	553	(52.0)	1695	(53.1)
Oral antibiotics before 12 months				
No	815	(79.7)	2266	(80.3)
Yes	`208	(20.3)	557	(19.7)

BMI, body mass index; WFL, weight-for-length; FN, First Nations; CS, caesarean section; IAP, intrapartum antibiotic prophylaxis. Values are n (%) or mean [standard deviation]. Percentages reflect proportion of non-missing data for each variable. Breastfeeding refers to breast milk feeding, regardless of feeding mode (at the breast or from a bottle).

eTable 2. Infant Feeding and Weight Variables Among Participants in the Subcohort (N* = 1087)

Feeding exposure variables		
Breastfeeding at 3 months		
None (formula only)	175	(16.1)
Partial (breast milk + formula)	323	(29.7)
Exclusive after hospital (brief formula supplementation in hospital)	179	(16.5)
Exclusive (breast milk only)	400	(36.8)
Missing	10	(0.9)
Breastfeeding status at time of 3-4 month sampling		
None (formula only)	225	(20.7)
Partial (breast milk + formula)	367	(33.8)
Exclusive after hospital (brief formula supplementation in hospital)	150	(13.8)
Exclusive (breast milk only)	324	(29.8)
Missing	21	(1.9)
Breastfeeding at 6 months		
None (formula +/- food)	264	(24.3)
Partial breastfeeding with formula (breast milk + formula +/- food)	307	(28.2)
Partial breastfeeding without formula (breast milk + food)	286	(26.3)
Exclusive (breast milk only)	183	(16.8)
Missing	47	(4.3)
Breastfeeding at 12 months		
No	591	(54.4)
Yes	459	(42.2)
Missing	37	(3.4)
Breastfeeding duration		
Never breastfed	49	(4.5)
Breastfed < 6 months	324	(29.8)
Breastfed 6 – 12 months	249	(22.9)
Breastfed > 12 months	459	(42.2)
Missing	6	(0.05)
Breastfeeding duration at time of 12 month sampling		
Never breastfed	43	(4.0)
Breastfed < 6 months	200	(18.4)
Breastfed > 6 months, but not currently breastfeeding	365	(33.6)
Currently breastfeeding	367	(33.8)
Missing	112	(10.3)
Breastfeeding duration (months; N = 1081)	10.26	[6.81]
Exclusive breastfeeding duration (months; N = 1058)	2.90	[2.37]
Weight outcome variables		
Weight class at 1 year		
Normal (WFL z-score < 85th percentile)	778	(71.6)
At risk (85th – 97th percentile)	178	(16.4)
Overweight (> 97th percentile)	71	(6.5)
Missing	60	(5.5)
WFL z-score at 12 months (N = 1027)	0.29	[1.08]

WFL, weight-for-length. *N = number of infants with 3 month and/or 12 month microbiota data. Values are n (%) or mean [± standard deviation]. Breastfeeding refers to breast milk feeding, regardless of feeding mode (at the breast or from a bottle).

eTable 3. Prevalence of Potential Confounders and Associations With Breastfeeding and Overweight Risk

	Overall Prevalence		Exclusive Breastfeeding at 3 months		Any Breastfeeding at 12 months		At Risk or Overweight at 12 months: WFL >85th percentile	
	n/N	(%)	%	*P* Value	%	*P* Value	%	*P* Value
Maternal race								
Asian	164/1078	(15.2)	53.7	0.13	51.6	0.10	17.6	0.007
Caucasian	817/1078	(75.8)	55.0		43.1		24.4	
FN	48/1078	(4.5)	45.8		36.2		43.2	
Other	49/1078	(4.5)	39.6		35.7		24.4	
Maternal postsecondary degree								
No	230/1053	(21.8)	45.9	0.008	30.3	<0.001	31.3	0.008
Yes	823/1053	(78.2)	56.0		47.7		22.3	
Maternal smoking in pregnancy								
No	967/1057	(91.5)	55.9	<0.001	46.1	<0.001	23.7	0.17
Yes	90/1057	(8.5)	30.3		20.9		31.3	
Maternal Healthy Eating Index 2010 Score								
< 70	306/1005	(30.4)	42.8	<0.001	33.2	<0.001	23.7	0.87
70 to 75	240/1005	(23.9)	55.8		42.0		25.3	
> 75	459/1005	(45.7)	60.3		52.1		23.6	
Maternal weight class								
Underweight	31/1025	(3.0)	54.8	<0.001	48.3	<0.001	20.7	0.05*
Normal	586/1025	(57.2)	59.2		49.4		21.7	
Overweight	230/1025	(22.4)	55.0		47.7		24.4	
Obese	178/1025	(17.4)	34.3		24.7		28.9	
Dog in home								
No	687/1027	(66.9)	57.2	0.003	47.5	0.002	24.0	0.94
Yes	340/1027	(33.1)	47.4		37.0		23.6	
Cat in home								
No	745/1026	(72.6)	54.4	0.67	44.2	0.93	23.5	0.68
Yes	281/1026	(27.4)	52.7		43.6		25.0	
Older siblings								
No	569/1087	(52.3)	51.9	0.27	44.5	0.56	25.4	0.40
Yes	518/1087	(47.7)	55.4		42.7		23.0	
Infant sex								
Male	507/1087	(46.6)	57.1	0.04	43.8	1	23.8	0.83
Female	580/1087	(53.4)	50.6		43.6		24.6	
Birth weight (g)								
< 3000	162/1064	(15.2)	51.3	0.91	43.3	0.79	13.8	<0.001
3000 to < 3500	419/1064	(39.4)	54.1		41.5		17.7	
3500 to < 4000	350/1064	(32.9)	54.0		45.1		30.5	
4000 +	133/1064	(12.5)	51.9		44.6		40.6	
Birth mode								
CS-Elective	121/1064	(11.4)	45.5	0.31	41.2	0.95	25.0	0.16
CS-Emergency	153/1064	(14.4)	54.9		43.0		27.5	
Vaginal, IAP	237/1064	(22.3)	54.7		44.2		28.0	
Vaginal, no IAP	553/1064	(52.0)	54.4		43.9		21.2	
Any oral antibiotics between birth and 12 months								
No	815/1023	(79.7)	53.6	0.55	45.9	0.05	23.8	0.63
Yes	`208/1023	(20.3)	51.0		37.9		25.7	

FN, First Nations; CS, caesarean section; IAP, intrapartum antibiotic prophylaxis, WFL, weight-for-length. Comparisons by chi-squared test or *Cochran-Armitage test for trend. Breastfeeding refers to breast milk feeding, regardless of feeding mode (at the breast or from a bottle).

eTable 4. Sensitivity Analyses: Association of Infant Feeding Practices With Infant Weight Status at 12 Months

	Model 1[1] OR (95%CI)	**Model 2[2] OR (95%CI)**	**Model 3[3] OR (95%CI)**	**Model 4[4] OR (95%CI)**	**OR (95%CI)**	**OR (95%CI)**	**Beta (95%CI)**
Outcome	WFL >85th percentile	WFL >85th percentile	WFL >85th percentile	WFL >85th percentile	WFL >85th percentile	WFL >85th percentile	WFL *z*-score
Sensitivity Analysis	Maternal BMI, Site	Other Maternal Factors: Maternal Smoking, Education, Ethnicity, Site	Microbiota Factors: Caesarean section, Dog, Infant Sex, Oral Antibiotics, Site	Maternal BMI, Other Maternal Factors, Microbiota Factors	Model 4 + Further Adjustment for Birth Weight	Model 4 Excluding Never-Breastfed Infants	Model 4 with Alternative Outcome: WFL *z*-score
Breastfeeding at 3 months	N = 990	N = 985	N = 913	N = 879	N=868	N=814	N = 879
None	2.04 (1.31 –3.19)	2.00 (1.27 – 3.12)	2.34 (1.49 –3.68)	2.04 (1.25 – 3.32)	2.15 (1.30 – 3.56)	2.33 (1.29 – 4.16)	0.30 (0.08 – 0.51)
Partial	1.55 (1.07 – 2.25)	1.53 (1.06 – 2.21)	1.67 (1.13 –2.45)	1.63 (1.10 – 2.45)	1.77 (1.17 – 2.69)	1.64 (1.09 – 2.47)	0.26 (0.09 – 0.43)
Exclusive after hospital	1.06 (0.66 – 1.69)	1.07 (0.67 – 1.70)	1.17 (0.71 – 1.91)	1.13 (0.67 – 1.87)	1.12 (0.65 – 1.89)	1.14 (0.68 – 1.89)	-0.04 (-0.25 – 0.16)
Exclusive	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	0.00 (reference)
Breastfeeding at 6 months	N = 973	N = 967	N = 901	N = 868	N = 856	N = 803	N = 868
None	1.94 (1.20 – 3.21)	1.92 (1.18 – 3.17)	1.94 (1.18 – 3.26)	1.65 (0.98 – 2.82)	1.57 (0.92 – 2.73)	1.68 (0.97 – 2.96)	0.28 (0.05 – 0.51)
Partial with formula	1.63 (1.02 – 2.65)	1.65 (1.03 – 2.68)	1.65 (1.03 – 2.74)	1.43 (0.87 – 2.39)	1.46 (0.88 – 2.48)	1.44 (0.88 – 2.41)	0.19 (-0.02 – 0.40)
Partial without formula	1.11 (0.68 – 1.85)	1.16 (0.71 – 1.92)	1.06 (0.64 – 1.80)	0.95 (0.56 – 1.63)	0.82 (0.47 – 1.43)	0.97 (0.57 – 1.66)	0.02 (-0.20 – 0.23)
Exclusive	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	0.00 (reference)
Breastfeeding duration	N = 950	N = 945	N = 876	N = 844	N = 832	N/A	N = 844
< 6 months*	1.89 (1.27 – 2.81)	1.77 (1.19 – 2.63)	1.84 (1.22 –2.77)	1.64 (1.06 – 2.52)	1.70 (1.08 – 2.66)	N/A	0.27 (0.08 – 0.45)
6 to < 12 months	1.60 (1.11 – 2.30)	1.59 (1.11 – 2.29)	1.63 (1.12 – 2.37)	1.47 (0.99 – 1.28)	1.57 (1.04 – 2.36)	N/A	0.19 (0.03 – 0.36)
≥ 12 months	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	N/A	0.00 (reference)

OR, odds ratio; CI, confidence interval; WFL, weight for length. [1]Adjusted for maternal BMI and study site. [2]Adjusted for maternal smoking, post-secondary education and ethnicity and site. [3]Adjusted for caesarean section, dog in household, infant sex, any oral antibiotics between 0 and 12 months and study site. [4]Adjusted for maternal BMI, smoking, post-secondary education and ethnicity and site, caesarean section, dog in household, infant sex and any oral antibiotics between 0 and 12 months. [*]Excludes infants who were never breastfed. Breastfeeding refers to breast milk feeding, regardless of feeding mode (at the breast or from a bottle).

eTable 5. Median Relative Abundance of Abundant† Taxa in Gut Microbiota at 3-4 Months According to Feeding Status[#]

	(A) No Breastfeeding	(B) Partial Breastfeeding	(C) Exclusive Breastfeeding after Hospital	(D) Exclusive Breastfeeding	Overall pFDR	Pairwise pFDR					
						A vs B	A vs C	A vs D	B vs C	B vs D	C vs D
Phylum *Family*	N=222	N=340	N=137	N=291							
Actinobacteria	3.829	5.468	5.368	8.456	***	a		c			f
Actinomycetaceae	0.039	0.031	0.016	0.016	**		b	c	d	e	
Bifidobacteriaceae	3.006	4.882	4.342	8.315	***	a	b	c		e	f
Coriobacteriaceae	0.079	0.054	0.023	0.016	***	a	b	c	d	e	
Micrococcaceae	0.008	0.008	0.008	0.016	***			c	d	e	
Bacteroidetes	22.446	38.887	0.908	4.061	***	a	b		d	e	
Bacteroidaceae	14.208	26.689	0.511	2.686	**	a			d	e	
Porphyromonadaceae	0.008	0.008	0.000	0.000	***		b	c	d	e	
Prevotellaceae	0.000	0.000	0.000	0.000	*			c			f
Rikenellaceae	0.000	0.000	0.000	0.000	**		b	c	d	e	
Firmicutes	31.996	19.703	18.073	20.405	***	a	b	c			
Clostridiales (other)	0.000	0.000	0.000	0.000	***	a	b	c	d	e	
Clostridiales (unclassified)	0.000	0.000	0.000	0.000	***	a	b	c	d	e	
Clostridiaceae	0.785	0.294	0.139	0.287	***	a	b	c			
Enterococcaceae	0.023	0.023	0.023	0.016							
Erysipelotrichaceae	0.215	0.023	0.000	0.008	***	a	b	c	d	e	
Gemellaceae	0.008	0.000	0.008	0.000	**	a			d		
Lachnospiraceae	7.942	3.022	0.349	0.395	***	a	b	c	d	e	
Lactobacillales (other)	0.016	0.000	0.000	0.000	***	a	b		d	e	
Ruminococcaceae	1.961	0.334	0.008	0.008	***	a	b	c	d	e	
Streptococcaceae	0.587	0.541	0.575	0.619							
Veillonellaceae	7.899	4.531	2.610	3.047	***	a	b	c	d	e	
Proteobacteria	11.271	15.922	32.261	26.492	***	a	b	c	d	e	
Alcaligenaceae	0.000	0.000	0.000	0.000							
Enterobacteriaceae	9.049	13.904	29.801	24.462	***	a	b	c	d	e	f
Pasteurellaceae	0.008	0.016	0.086	0.101	***	a	b	c	d	e	
Verrucomicrobia	0.000	0.000	0.000	0.000	***	a	b	c	d	e	
Verrucomicrobiaceae	0.000	0.000	0.000	0.000	***	a	b	c	d	e	
Firmicutes/ Bacteroidetes ratio	1.54	0.56	17.63	5.53	***	a			d	e	

FDR, false discovery rate. †Taxa with median abundance >0% in 3 month and/or 12 month microbiota. [#]Feeding status at the time of sample collection. Overall comparisons by rank-based nonparametric Kruskal–Wallis test with FDR correction for multiple comparisons; *p<0.05, **p<0.01, ***p<0.001. Significant differences are possible when all medians are 0.000 because this is a rank-based test. Pairwise comparisons by Dunn's post-hoc tests for multiple comparisons (significant pairwise differences): [a]No breastfeeding (BF)/Partial BF; [b]No BF/Exclusive BF after hospital; [c]No BF/Exclusive BF; [d]Partial BF/Exclusive BF after hospital; [e]Partial BF/Exclusive BF; [f]Exclusive BF after hospital/Exclusive BF. Breastfeeding refers to breast milk feeding, regardless of feeding mode (at the breast or from a bottle).

eTable 6. Pairwise PERMANOVA Analyses of Infant Microbiota According to Feeding Status at 3-4 Months and 6 Months

	Unweighted Unifrac		Weighted Unifrac	
	Pseudo-F	*P* Value	Pseudo-F	*P* Value
BF at 3-4 months	**Microbiota at 3-4 months**			
No BF vs Partial BF	8.11	0.001	8.46	0.001
No BF vs Exclusive BF after hospital	18.46	0.001	16.99	0.001
No BF vs Exclusive BF	22.54	0.001	18.63	0.001
Exclusive BF after hospital vs Partial BF	8.15	0.001	14.56	0.001
Partial BF vs Exclusive BF	8.16	0.001	14.03	0.001
Exclusive BF after hospital vs Exclusive BF	1.42	0.07	0.24	0.24
Diet at 6 months	**Microbiota at 12 months**			
No BF vs Partial BF with formula	6.67	0.001	5.74	0.002
No BF vs Partial BF without formula	13.37	0.001	12.74	0.001
No BF vs Exclusive BF	12.03	0.001	10.95	0.001
Partial BF with formula vs Partial BF without formula	3.76	0.001	9.72	0.001
Partial BF with formula vs Exclusive BF	3.43	0.001	9.81	0.001
Partial BF without formula vs Exclusive BF	0.93	0.59	0.40	0.78

BF, breastfeeding. Pairwise community structure differences tested by PERMANOVA with 999 permutations based on unweighted or weighted Unifrac distances.

eTable 7. Median Relative Abundance of Abundant† Taxa in Fecal Microbiota of Infants at 12 Months According to Feeding Status at 6 Months

	(A) No Breastfeeding	(B) Partial Breastfeeding with Formula	(C) Partial Breastfeeding without Formula	(D) Exclusive Breastfeeding	Overall pFDR	Pairwise pFDR					
Phylum *Family*	N=190	N=248	N=218	N=147		A vs B	A vs C	A vs D	B vs C	B vs D	C vs D
Actinobacteria	1.724	2.134	3.212	3.293	**		b	c	d	e	
Actinomycetaceae	0.008	0.008	0.008	0.008							
Bifidobacteriaceae	1.422	1.809	2.896	2.796	**		b	c	d	e	
Coriobacteriaceae	0.117	0.098	0.085	0.085							
Micrococcaceae	0.000	0.000	0.000	0.000							
Bacteroidetes	49.853	52.703	45.714	45.054	**		b	c	d	e	
Bacteroidaceae	42.131	46.911	39.098	41.060					d	e	
Porphyromonadaceae	0.125	0.082	0.016	0.008			b	c	d	e	
Prevotellaceae	0.008	0.008	0.012	0.015							
Rikenellaceae	0.008	0.008	0.004	0.008							
Firmicutes	36.420	34.671	39.050	36.364					d		
Clostridiales (other)	0.008	0.008	0.008	0.008			b				
Clostridiales (unclassified)	0.808	0.998	0.603	0.521							
Clostridiaceae	0.387	0.515	0.560	0.544							
Enterococcaceae	0.000	0.000	0.004	0.000							
Erysipelotrichaceae	0.526	0.452	0.420	0.350				c			
Gemellaceae	0.008	0.008	0.008	0.008							
Lachnospiraceae	14.531	13.372	14.275	12.894							
Lactobacillales (other)	0.008	0.015	0.008	0.008							
Ruminococcaceae	7.275	6.433	5.225	5.713							
Streptococcaceae	0.257	0.285	0.337	0.312							
Veillonellaceae	3.040	4.611	7.644	7.078	***	a	b	c	d	e	
Proteobacteria	3.611	3.782	5.104	5.706	**		b	c	d	e	
Alcaligenaceae	0.266	0.719	1.057	1.188			b	c			
Enterobacteriaceae	1.167	0.928	1.446	1.391					d		
Pasteurellaceae	0.105	0.194	0.279	0.249	*	a	b	c			
Verrucomicrobia	0.023	0.016	0.015	0.016							
Verrucomicrobiaceae	0.023	0.016	0.015	0.016							
Firmicutes/ Bacteroidetes ratio	0.71	0.68	0.87	0.85	**		b		d	e	

FDR, false discovery rate. †Taxa with median abundance >0% in 3 month and/or 12 month microbiota. Overall comparisons by rank-based nonparametric Kruskal–Wallis test with FDR correction for multiple comparisons; *p<0.05, **p<0.01, ***p<0.001. Significant differences are possible when all medians are 0.000 because this is a rank-based test. Pairwise comparisons by Dunn's post-hoc tests for multiple comparisons (significant pairwise differences): [a]No breastfeeding (BF)/Partial BF with formula; [b]No BF/Partial BF without formula; [c]No BF/Exclusive BF; [d]Partial BF with formula/Partial BF without formula; [e]Partial BF with formula/Exclusive BF; no significant differences observed between Partial BF without formula and Exclusive BF. Breastfeeding refers to breast milk feeding, regardless of feeding mode (at the breast or from a bottle).

eTable 8. Median Relative Abundance of Abundant† Taxa in Fecal Microbiota of Infants at 12 Months According to Breastfeeding (BF) Duration[#]

	(A) Never Breastfed	(B) Breastfed < 6 months	(C) Breastfed 6 – 12 months	(D) Breastfed > 12 months	Overall pFDR	Pairwise pFDR					
Phylum *Family*	N=34	N=320	N=149	N=316		A vs B	A vs C	A vs D	B vs C	B vs D	C vs D
Actinobacteria	1.010	2.060	1.667	3.637	***			a		b	c
Actinomycetaceae	0.008	0.008	0.008	0.008							
Bifidobacteriaceae	0.944	1.533	1.485	3.270	***			a		b	c
Coriobacteriaceae	0.129	0.117	0.086	0.086							
Micrococcaceae	0.000	0.000	0.000	0.000							
Bacteroidetes	50.845	48.509	52.810	47.002							c
Bacteroidaceae	46.682	40.262	47.045	42.293							
Porphyromonadaceae	0.183	0.121	0.055	0.008	*					b	c
Prevotellaceae	0.008	0.008	0.008	0.015							
Rikenellaceae	0.016	0.008	0.015	0.000	***			a		b	c
Firmicutes	33.699	37.739	33.623	36.673							
Clostridiales (other)	0.008	0.015	0.008	0.008	***					b	c
Clostridiales (unclassified)	0.638	1.049	1.160	0.433	***					b	c
Clostridiaceae	0.388	0.514	0.500	0.544							
Enterococcaceae	0.000	0.000	0.000	0.008	**			a		b	c
Erysipelotrichaceae	0.471	0.521	0.481	0.350	**			a		b	c
Gemellaceae	0.008	0.008	0.008	0.008							
Lachnospiraceae	14.006	14.595	14.457	12.861						b	
Lactobacillales (other)	0.016	0.008	0.015	0.008							
Ruminococcaceae	6.624	7.715	6.682	4.272	***					b	c
Streptococcaceae	0.277	0.283	0.281	0.319							
Veillonellaceae	1.925	3.944	3.685	8.749	***			a		b	c
Proteobacteria	3.042	3.869	3.962	5.694	***			a		b	c
Alcaligenaceae	0.012	0.704	0.721	0.956				a			
Enterobacteriaceae	1.191	0.973	0.781	1.683	*					b	c
Pasteurellaceae	0.109	0.187	0.162	0.250							
Verrucomicrobia	0.016	0.031	0.016	0.008						b	
Verrucomicrobiaceae	0.016	0.031	0.016	0.008						b	
Firmicutes/ Bacteroidetes ratio	0.66	0.78	0.64	0.78				a		b	c

FDR, false discovery rate. †Taxa with median abundance >0% in 3 month and/or 12 month microbiota. [#]Feeding status at the time of sample collection. Overall comparisons by rank-based nonparametric Kruskal–Wallis test with FDR correction for multiple comparisons; *p<0.05, **p<0.01, ***p<0.001. Significant differences are possible when all medians are 0.000 because this is a rank-based test. Pairwise comparisons by Dunn's post-hoc tests for multiple comparisons (significant pairwise differences): [a]BF < 6 months/BF > 12 months; [b]BF 6 – 12 months/BF > 12 months; [c]Never BF/BF > 12 months; no significant differences were observed between BF <6 months/BF 6 – 12 months, Never BF/BF < 6 months or Never BF/BF 6 – 12 months. Breastfeeding refers to breast milk feeding, regardless of feeding mode (at the breast or from a bottle).

eTable 9. Median Relative Abundance of Abundant† Taxa in Fecal Microbiota of Infants at 3-4 and 12 Months According to Infant Weight Status at 12 Months

	Microbiota at 3-4 months						Microbiota at 12 months					
	(A) Normal	(B) At risk	(C) Overweight	Overall pFDR	Pairwise pFDR		(A) Normal	(B) At risk	(C) Overweight	Overall pFDR	Pairwise pFDR	
Phylum *Family*	N=699	N=171	N=67		A vs B	A vs C	B vs C	N=616	N=137	N=56		A vs B	A vs C	B vs C
Actinobacteria	5.360	5.232	6.194					2.756	2.178	1.890		a		c
Actinomycetaceae	0.023	0.023	0.039					0.008	0.008	0.008				
Bifidobacteriaceae	4.591	4.967	5.754					2.290	1.915	1.197				
Coriobacteriaceae	0.039	0.055	0.085			b		0.093	0.147	0.117				
Micrococcaceae	0.008	0.008	0.015					0.000	0.000	0.000				
Bacteroidetes	21.753	8.091	3.483					50.111	45.054	48.316				
Bacteroidaceae	14.201	5.117	2.247					44.024	35.248	42.399				
Porphyromonadaceae	0.008	0.008	0.008					0.023	0.078	0.012				
Prevotellaceae	0.000	0.000	0.000					0.008	0.016	0.008				
Rikenellaceae	0.000	0.000	0.000					0.008	0.008	0.008				
Firmicutes	20.708	28.229	26.142					35.749	40.016	37.312				
Clostridiales (other)	0.000	0.000	0.000					0.008	0.016	0.008				
Clostridiales (unclassified)	0.000	0.000	0.000	*	a	b		0.641	0.814	1.199				
Clostridiaceae	0.330	0.574	0.581					0.516	0.564	0.385				
Enterococcaceae	0.023	0.023	0.024					0.000	0.000	0.008				
Erysipelotrichaceae	0.016	0.031	0.124			b		0.425	0.499	0.356				
Gemellaceae	0.000	0.000	0.008					0.008	0.008	0.008		a		
Lachnospiraceae	1.915	4.699	5.848	*	a	b		13.392	15.163	14.450				
Lactobacillales (other)	0.000	0.000	0.000					0.008	0.015	0.012				
Ruminococcaceae	0.054	0.278	0.280			b		6.362	6.460	6.454				
Streptococcaceae	0.535	0.710	0.904					0.281	0.404	0.245		a		c
Veillonellaceae	4.226	6.266	5.537					5.250	5.814	3.978				
Proteobacteria	18.196	19.921	22.815					4.448	3.887	5.758				
Alcaligenaceae	0.000	0.000	0.000					0.748	0.598	0.733				
Enterobacteriaceae	16.273	18.287	20.413					1.226	0.972	1.251				
Pasteurellaceae	0.031	0.031	0.031					0.195	0.256	0.082				
Verrucomicrobia	0.000	0.000	0.000					0.016	0.016	0.093				
Verrucomicrobiaceae	0.000	0.000	0.000					0.016	0.016	0.093				
Firmicutes/ Bacteroidetes ratio	1.29	3.08	11.36					0.73	0.84	0.76				

FDR, false discovery rate. †Taxa with median abundance >0% in 3 month and/or 12 month microbiota. Overall comparisons by rank-based nonparametric Kruskal–Wallis test with FDR correction for multiple comparisons; *p<0.05, **p<0.01, ***p<0.001. Significant differences are possible when all medians are 0.000 because this is a rank-based test. Pairwise comparisons by Dunn's post-hoc tests for multiple comparisons (significant pairwise differences): [a]Normal/At Risk; [b]Normal/Overweight; [c]At Risk/Overweight

eAppendix. Detailed Methods

DNA isolation

Following collection and aliquotting, stool samples were maintained frozen at or below –80 C prior to analysis. For the isolation of community DNA, a modification of the Qiagen Isolation of DNA from Stool for Pathogen Detection protocol was used (Qiagen Inc., Valencia CA). In this procedure, a target mass of 200 mg (acceptable range from 80–220 mg, actual mass recorded) of frozen stool was combined with 1.4 mL Qiagen Stool Lysis Buffer (ASL), vortex mixed for 1 min or until the sample appeared thoroughly thawed and homogenized, and placed in a 95 C water bath for 5 min. Samples were then vortex mixed for 15 sec and centrifuged for 2 min at 14 Krpm. A volume of 1.2 mL of supernatant was removed to a new microcentrifuge tube, combined with a tablet of InhibitEX (Qiagen) and vortex mixed continuously for 1 min or until suspended. Samples were then incubated at room temperature for 1 min and centrifuged for 5 min at 14 Krpm. The remainder of the isolation procedure was carried out using a QIAcube robot following the "Pathogen Detection" program modified for a 60 µL elution volume rathewr than the standard 200 µL elution volume. Following completion of the program, DNA quality was evaluated by electrophoresis of 5 µL of isolated DNA in 1.0% agarose in 1×TAE buffer for 80 min at 80 VDV, visualized using SYBR safe stain (Thermo Fisher Scientific / Life Technologies Corp., Carlsbad CA) and recorded using a GelDoc XR+ Imaging system (BioRad Laboratories Inc., Hercules CA). DNA concentration was evaluated using a Quantifluor dsDNA system following the manufacturer's instructions, and adjusted to a final concentration of 5 ng/µL by the addition of 1×TE buffer.

Amplification

Bacterial 16S DNA from hypervariable region V4 was amplified by PCR using the core forward primer V4+515F (5'-TATGGTAATTGTGTGCCAGCMGCCGCGGTAA-3') and the core reverse primer V4-806R (5'-AGTCAGTCAGCCGGACTACHVGGGTWTCTAAT-3') synthesized to include Illumina adapers, primer pad and linker sequences, a Golay barcode (forward primer). PCRs were conducted in a final volume of 25 µL, consisting of 10 ng template DNA and 0.6 µM of each primer in Kapa2G Robust Hotstart Taq ready mix (KapaBiosystems, Wilmington MA) at 1× concentration. PCR conditions consisted of an initial denaturation at 94 for 3 min followed by 20 cycles of 94 C for 30 sec (denaturation), 50 C for 30 sec (annealing), and 72 C for 30 sec (extension), with a completion step at 10 C. Amplicon quality was assessed by electrophoresis following the procedure described in the previous section and quantitated based on the summed brightness of bands as determined by the GelDoc XR+ analyser. Samples yielding less 100 ng of total product were reamplified using diluted template DNA to reduce the concentration of PCR inhibitors.
PCR products were combined for multiplex sequencing in batches of 48 up to a maximum of 96. Total volume was reduced using Amicon Ultra centrifugal filter concentrators (Millipore Sigma, Burlington MA) to between 25–50 µL. Concentrated products electrophoresed on a 1.4% agarose gel in 1×TAE containing SYBR safe stain at 80 VDC for 90 min. Bands were excised and purified using a GeneClean Turbo Kit (MP Biomedicals, Santa Ana, CA) following the manufacturer's directions modified to elute the product in 40 µL DES for 10 min followed by centrifugation of final eluent for 2 min. Final DNA concentration was determined by Quantifluor, as above.

Sequencing and bioinformatics

Samples were sequenced by Illumina MiSeq (San Diego, CA) using 150 bp paired-end (x2) V2 chemistry. Data were outputted in a format consisting of two fastq files containing read 1 and read 2 datasets, and a third fastq file corresponding to barcodes. Following decompression and concatenation of data files, forward and reverse assembled and the resultant contigs binned by barcode using the Qiime (version 1.9.1) open-source bioinformatics pipeline. The analysis pipeline, in brief, consisted of the following; Non-bacterial sequences were excluded as those that failed to cluster against the Greengenes reference database (version May 2013) at 60% similarity. The resulting filtered dataset was subjected to closed reference picking against the Greengenes reference database at 97% similarity using USEARCH6.1. Sequences that failed to cluster were aggregated over the entire dataset (singletons removed) and subjected to de novo clustering using USEARCH10 (64 bit). Taxonomies were assigned according to the Greengenes reference database, and closed- and de novo picked datasets were merged.

Pipeline command sequence summary

#MiSeq FASTQ FROM SEQUENCER - UNZIPPING AND CONCATENATING FILES
#note used all files instead of just the 'undefined' files and allowed the quality parameters to cull for consistency across runs
gunzip *I1*.gz
gunzip *R1*.gz
gunzip *R2*.gz
cat *S1_L001_I1_001.fastq *S2_L001_I1_001.fastq *S3_L001_I1_001.fastq *S4_L001_I1_001.fastq *S5_L001_I1_001.fastq *S6_L001_I1_001.fastq *S7_L001_I1_001.fastq *S8_L001_I1_001.fastq *S9_L001_I1_001.fastq *S10_L001_I1_001.fastq *S11_L001_I1_001.fastq *S12_L001_I1_001.fastq Undetermined_S0_L001_I1_001.fastq > cat_index.fastq
cat *S1_L001_R1_001.fastq *S2_L001_R1_001.fastq *S3_L001_R1_001.fastq *S4_L001_R1_001.fastq *S5_L001_R1_001.fastq *S6_L001_R1_001.fastq *S7_L001_R1_001.fastq *S8_L001_R1_001.fastq *S9_L001_R1_001.fastq *S10_L001_R1_001.fastq *S11_L001_R1_001.fastq *S12_L001_R1_001.fastq Undetermined_S0_L001_R1_001.fastq > cat_R1.fastq

cat *S1_L001_R2_001.fastq *S2_L001_R2_001.fastq *S3_L001_R2_001.fastq *S4_L001_R2_001.fastq *S5_L001_R2_001.fastq *S6_L001_R2_001.fastq *S7_L001_R2_001.fastq *S8_L001_R2_001.fastq *S9_L001_R2_001.fastq *S10_L001_R2_001.fastq *S11_L001_R2_001.fastq *S12_L001_R2_001.fastq Undetermined_S0_L001_R2_001.fastq > cat_R2.fastq

#_____

#MiSeq RUN CLOSED REFERENCE OTU-PICKING PIPELINE

#reference database = Greengenes version May 2013

#count the number of reads for input from MiSeq R1 fastq file

count_seqs.py -i cat_R1.fastq -o cat_R1_seq_count.txt

#assembling forward and reverse reads

join_paired_ends.py -m fastq-join -j 25 -p 5 -b cat_index.fastq -f cat_R1.fastq -r cat_R2.fastq -o fastq_join/

#binning sequences by bar code

split_libraries_fastq.py -i fastq_join/fastqjoin.join.fastq -b fastq_join/fastqjoin.join_barcodes.fastq -m mapping_file.txt -r 3 -p 0.00 -n 0 --rev_comp_mapping_barcodes --barcode_type 12 --max_barcode_errors 1.5 -o split_seqs/

#identify non-bacterial reads to be filtered

parallel_pick_otus_usearch61_ref.py -i split_seqs/seqs.fna -r /home/james/qiime_software/gg_13_8_otus/rep_set/97_otus.fasta --usearch61_sort_method abundance --sizeorder --similarity 0.6 --jobs_to_start 16 -o prefilter/

#count the number of sequences that failed to cluster with reference data set at 60%

wc -l prefilter/seqs_failures.txt > prefilter/seqs_failures_count.txt

#create filtered data set

filter_fasta.py -f split_seqs/seqs.fna -s prefilter/seqs_failures.txt -n -o prefilter/prefiltered_seqs.fna

#parallel closed reference pick OTUs using usearch61

parallel_pick_otus_usearch61_ref.py -i prefilter/prefiltered_seqs.fna -r /home/james/qiime_software/gg_13_8_otus/rep_set/97_otus.fasta --usearch61_sort_method abundance --sizeorder --similarity 0.97 --jobs_to_start 16 -o closed_ref_OTUs/

#count the number of sequences that failed to cluster with reference data set at 97%

wc -l closed_ref_OTUs/prefiltered_seqs_failures.txt > closed_ref_OTUs/prefiltered_seqs_failures_count.txt

#pick representative sequence from each OTU cluster

pick_rep_set.py -i closed_ref_OTUs/prefiltered_seqs_otus.txt -f prefilter/prefiltered_seqs.fna -o closed_ref_OTUs/rep_set.fasta

#Assign taxonomy using uclust

assign_taxonomy.py -i closed_ref_OTUs/rep_set.fasta -m uclust -r /home/james/qiime_software/gg_13_8_otus/rep_set/97_otus.fasta -t /home/james/qiime_software/gg_13_8_otus/taxonomy/97_otu_taxonomy.txt --uclust_min_consensus_fraction 0.51 --uclust_similarity 0.9 --uclust_max_accepts 3 -o closed_ref_OTUs/uclust_closed_ref_tax_assign/

#make biom formatted OTU table

make_otu_table.py -i closed_ref_OTUs/prefiltered_seqs_otus.txt -t closed_ref_OTUs/uclust_closed_ref_tax_assign/rep_set_tax_assignments.txt -o closed_ref_OTUs/uclust_closed_ref_tax_assign/uclust_closed_ref_picked_OTUs.biom

#summarize biom table

biom summarize-table -i closed_ref_OTUs/uclust_closed_ref_tax_assign/uclust_closed_ref_picked_OTUs.biom -o closed_ref_OTUs/uclust_closed_ref_tax_assign/uclust_closed_ref_picked_OTUs_biom_table_summary.txt

#count the number of OTUs per sample

alpha_diversity.py -i closed_ref_OTUs/uclust_closed_ref_tax_assign/uclust_closed_ref_picked_OTUs.biom -m observed_species -o closed_ref_OTUs/uclust_closed_ref_tax_assign/uclust_closed_ref_picked_OTU_count_per_sample.txt

#convert biom-formatted table to tab-delimited text format

#note that this table is not used in downstream pipeline

biom convert -i closed_ref_OTUs/uclust_closed_ref_tax_assign/uclust_closed_ref_picked_OTUs.biom -o closed_ref_OTUs/uclust_closed_ref_tax_assign/uclust_closed_ref_picked_OTUs.txt -b --header-key taxonomy

#make directory for de novo picking files

mkdir denovo_OTUs

#pull reads that failed to cluster with reference database for de novo OTU picking

filter_fasta.py -f prefilter/prefiltered_seqs.fna -s closed_ref_OTUs/prefiltered_seqs_failures.txt -o denovo_OTUs/seqs_for_denovo_pick.fna

#count the number of reads for input into denovo-picking step (note will/should be the same number as output in prefiltered_seqs_failures_count.txt)

count_seqs.py -i denovo_OTUs/seqs_for_denovo_pick.fna -o denovo_OTUs/seq_count.txt

#_____

 # merge biom tables from sym closed OTU picking and characterize 1) full table 2) table with singletons removed

merge_otu_tables.py -i sym1_130211_uclust_closed_ref_picked_OTUs.biom,sym2_130212_uclust_closed_ref_picked_OTUs.biom,sym3_130214_uclust_closed_ref_picked_OTUs.biom,sym4_130221_uclust_closed_ref_picked_OTUs.biom,sym5_130226_uclust_closed_ref_picked_OTUs.biom,sym6_130227_uclust_closed_ref_picked_OTUs.biom,sym7_130304_uclust_closed_ref_picked_OTUs.biom,sym8_130306_uclust_closed_ref_picked_OTUs.biom,sym9_130416_uclust_closed_ref_picked_OTUs.biom,sym10_130418_uclust_closed_ref_picked_OTUs.biom,sym11_130423_uclust_closed_ref_picked_OTUs.biom,sym12_130424_uclust_closed_ref_picked_OTUs.biom,sym13_130429_uclust_closed_ref_picked_OTUs.biom,sym14_130502_uclust_closed_ref_picked_OTUs.biom,sym15_130506_uclust_closed_ref_picked_OTUs.biom,s

ym16_130417_uclust_closed_ref_picked_OTUs.biom,sym20_130711_uclust_closed_ref_picked_OTUs.biom,sym21_130716_uclust_closed_ref_picked_OTUs.biom,sym22_130717_uclust_closed_ref_picked_OTUs.biom,sym23_130718_uclust_closed_ref_picked_OTUs.biom,sym24_130924_uclust_closed_ref_picked_OTUs.biom,sym25_130925_uclust_closed_ref_picked_OTUs.biom,sym26_130930_uclust_closed_ref_picked_OTUs.biom,sym27_131001_uclust_closed_ref_picked_OTUs.biom,sym28_131009_uclust_closed_ref_picked_OTUs_corrected.biom,sym29_131106_uclust_closed_ref_picked_OTUs.biom,sym30_131111_uclust_closed_ref_picked_OTUs_corrected.biom,sym31_140225_uclust_closed_ref_picked_OTUs.biom,sym33_140219_uclust_closed_ref_picked_OTUs.biom,sym34_140220_uclust_closed_ref_picked_OTUs.biom,sym35_140227_uclust_closed_ref_picked_OTUs.biom,sym36_140916_uclust_closed_ref_picked_OTUs.biom,sym37_140918_uclust_closed_ref_picked_OTUs.biom,sym38_140818_uclust_closed_ref_picked_OTUs.biom,sym39_140826_uclust_closed_ref_picked_OTUs.biom,sym40_140826_uclust_closed_ref_picked_OTUs_corrected.biom,sym41_140818_uclust_closed_ref_picked_OTUs.biom,sym42_141103_uclust_closed_ref_picked_OTUs.biom,sym43_141103_uclust_closed_ref_picked_OTUs.biom,sym44_140916_uclust_closed_ref_picked_OTUs.biom,sym45_150113_uclust_closed_ref_picked_OTUs.biom,sym46_150114_uclust_closed_ref_picked_OTUs.biom,sym47_150224_uclust_closed_ref_picked_OTUs.biom,sym48_150225_uclust_closed_ref_picked_OTUs.biom,sym49_150225_uclust_closed_ref_picked_OTUs.biom,sym50_150504_uclust_closed_ref_picked_OTUs.biom,sym51_150611_uclust_closed_ref_picked_OTUs.biom,sym52_150506_uclust_closed_ref_picked_OTUs.biom,sym53_150507_uclust_closed_ref_picked_OTUs.biom,sym54_150602_uclust_closed_ref_picked_OTUs.biom,sym55_150603_uclust_closed_ref_picked_OTUs.biom,sym56_150604_uclust_closed_ref_picked_OTUs.biom,sym57_150604_uclust_closed_ref_picked_OTUs.biom,sym58_150707_uclust_closed_ref_picked_OTUs.biom,sym59_150708_uclust_closed_ref_picked_OTUs.biom,sym60_150709_uclust_closed_ref_picked_OTUs.biom,sym61_150715_uclust_closed_ref_picked_OTUs.biom,sym62_150817_uclust_closed_ref_picked_OTUs.biom,sym63_150812_uclust_closed_ref_picked_OTUs.biom,sym64_150818_uclust_closed_ref_picked_OTUs.biom,sym65_150819_uclust_closed_ref_picked_OTUs.biom,sym66_151001_uclust_closed_ref_picked_OTUs.biom,sym67_151006_uclust_closed_ref_picked_OTUs.biom,sym68_160105_uclust_closed_ref_picked_OTUs.biom,sym69_151007_uclust_closed_ref_picked_OTUs.biom,sym70_151109_uclust_closed_ref_picked_OTUs.biom,sym71_151110_uclust_closed_ref_picked_OTUs.biom,sym72_151111_uclust_closed_ref_picked_OTUs.biom,sym73_151125_uclust_closed_ref_picked_OTUs.biom,sym74_151209_uclust_closed_ref_picked_OTUs.biom,sym75_151210_uclust_closed_ref_picked_OTUs.biom,sym76_151214_uclust_closed_ref_picked_OTUs.biom,sym77_151215_uclust_closed_ref_picked_OTUs.biom,sym78_161114_uclust_closed_ref_picked_OTUs.biom,sym79_160216_uclust_closed_ref_picked_OTUs.biom,sym80_160217_uclust_closed_ref_picked_OTUs.biom,sym81_160217_uclust_closed_ref_picked_OTUs.biom,sym82_160329_uclust_closed_ref_picked_OTUs.biom,sym83_160329_uclust_closed_ref_picked_OTUs.biom,sym84_160404_uclust_closed_ref_picked_OTUs.biom,sym85_160404_uclust_closed_ref_picked_OTUs.biom,sym86_160629_uclust_closed_ref_picked_OTUs.biom,sym87_160718_uclust_closed_ref_picked_OTUs.biom,sym88_160705_uclust_closed_ref_picked_OTUs.biom,sym89_160707_uclust_closed_ref_picked_OTUs.biom,sym90_160809_uclust_closed_ref_picked_OTUs.biom,sym91_160809_uclust_closed_ref_picked_OTUs.biom,sym92_160825_uclust_closed_ref_picked_OTUs.biom,sym93_160825_uclust_closed_ref_picked_OTUs.biom,sym94_160928_uclust_closed_ref_picked_OTUs.biom,sym95_160929_uclust_closed_ref_picked_OTUs.biom,sym96_160929_uclust_closed_ref_picked_OTUs.biom,sym97_160929_uclust_closed_ref_picked_OTUs.biom,sym98_161012_uclust_closed_ref_picked_OTUs.biom,sym99_161012_uclust_closed_ref_picked_OTUs.biom,sym100_161013_uclust_closed_ref_picked_OTUs.biom -o merged_symbiota_closed_picked_table_1to100.biom

 # characterize full biom table

biom summarize-table -i merged_symbiota_closed_picked_table_1to100.biom -o
merged_symbiota_closed_picked_table_1to100_biom_table_summary.txt
alpha_diversity.py -i merged_symbiota_closed_picked_table_1to100.biom -m observed_species -o
merged_symbiota_closed_picked_table_1to100_OTU_count_per_sample.txt

#_____

#COMBINED MiSeq RUN DENOVO OTU-PICKING PIPELINE USING USEARCH10 64 bit

 # concatenate reads that failed to cluster with reference database into single fasta file for denovo picking

cat sym1_130211_seqs_for_denovo_pick.fna sym2_130212_seqs_for_denovo_pick.fna sym3_130214_seqs_for_denovo_pick.fna
sym4_130221_seqs_for_denovo_pick.fna sym5_130226_seqs_for_denovo_pick.fna sym6_130227_seqs_for_denovo_pick.fna
sym7_130304_seqs_for_denovo_pick.fna sym8_130306_seqs_for_denovo_pick.fna sym9_130416_seqs_for_denovo_pick.fna
sym10_130418_seqs_for_denovo_pick.fna sym11_130423_seqs_for_denovo_pick.fna sym12_130424_seqs_for_denovo_pick.fna
sym13_130429_seqs_for_denovo_pick.fna sym14_130502_seqs_for_denovo_pick.fna sym15_130506_seqs_for_denovo_pick.fna
sym16_130417_seqs_for_denovo_pick.fna sym20_130711_seqs_for_denovo_pick.fna sym21_130716_seqs_for_denovo_pick.fna
sym22_130717_seqs_for_denovo_pick.fna sym23_130718_seqs_for_denovo_pick.fna sym24_130924_seqs_for_denovo_pick.fna
sym25_130925_seqs_for_denovo_pick.fna sym26_130930_seqs_for_denovo_pick.fna sym27_131001_seqs_for_denovo_pick.fna
sym28_131009_corrected_seqs_for_denovo_pick.fna sym29_131106_seqs_for_denovo_pick.fna
sym30_131111_corrected_seqs_for_denovo_pick.fna sym31_140225_seqs_for_denovo_pick.fna
sym33_140219_seqs_for_denovo_pick.fna sym34_140220_seqs_for_denovo_pick.fna sym35_140227_seqs_for_denovo_pick.fna
sym36_140916_seqs_for_denovo_pick.fna sym37_140918_seqs_for_denovo_pick.fna sym38_140818_seqs_for_denovo_pick.fna
sym39_140826_seqs_for_denovo_pick.fna sym40_140826_corrected_seqs_for_denovo_pick.fna
sym41_140818_seqs_for_denovo_pick.fna sym42_141103_seqs_for_denovo_pick.fna sym43_141103_seqs_for_denovo_pick.fna
sym44_140916_seqs_for_denovo_pick.fna sym45_150113_seqs_for_denovo_pick.fna sym46_150114_seqs_for_denovo_pick.fna
sym47_150224_seqs_for_denovo_pick.fna sym48_150225_seqs_for_denovo_pick.fna sym49_150225_seqs_for_denovo_pick.fna
sym50_150504_seqs_for_denovo_pick.fna sym51_150611_seqs_for_denovo_pick.fna sym52_150506_seqs_for_denovo_pick.fna
sym53_150507_seqs_for_denovo_pick.fna sym54_150602_seqs_for_denovo_pick.fna sym55_150603_seqs_for_denovo_pick.fna
sym56_150604_seqs_for_denovo_pick.fna sym57_150604_seqs_for_denovo_pick.fna sym58_150707_seqs_for_denovo_pick.fna
sym59_150708_seqs_for_denovo_pick.fna sym60_150709_seqs_for_denovo_pick.fna sym61_150715_seqs_for_denovo_pick.fna

sym62_150817_seqs_for_denovo_pick.fna sym63_150812_seqs_for_denovo_pick.fna sym64_150818_seqs_for_denovo_pick.fna
sym65_150819_seqs_for_denovo_pick.fna sym66_151001_seqs_for_denovo_pick.fna sym67_151006_seqs_for_denovo_pick.fna
sym68_160105_seqs_for_denovo_pick.fna sym69_151007_seqs_for_denovo_pick.fna sym70_151109_seqs_for_denovo_pick.fna
sym71_151110_seqs_for_denovo_pick.fna sym72_151111_seqs_for_denovo_pick.fna sym73_151125_seqs_for_denovo_pick.fna
sym74_151209_seqs_for_denovo_pick.fna sym75_151210_seqs_for_denovo_pick.fna sym76_151214_seqs_for_denovo_pick.fna
sym77_151215_seqs_for_denovo_pick.fna sym78_xxxxxx_seqs_for_denovo_pick.fna sym79_151215_seqs_for_denovo_pick.fna
sym80_160217_seqs_for_denovo_pick.fna sym81_160217_seqs_for_denovo_pick.fna sym82_160329_seqs_for_denovo_pick.fna
sym83_160329_seqs_for_denovo_pick.fna sym84_160404_seqs_for_denovo_pick.fna sym85_160404_seqs_for_denovo_pick.fna
sym86_160629_seqs_for_denovo_pick.fna sym87_160718_seqs_for_denovo_pick.fna sym88_160705_seqs_for_denovo_pick.fna
sym89_160707_seqs_for_denovo_pick.fna sym90_160809_seqs_for_denovo_pick.fna sym91_160809_seqs_for_denovo_pick.fna
sym92_160825_seqs_for_denovo_pick.fna sym93_160825_seqs_for_denovo_pick.fna sym94_160928_seqs_for_denovo_pick.fna
sym95_160929_seqs_for_denovo_pick.fna sym96_160929_seqs_for_denovo_pick.fna sym97_160929_seqs_for_denovo_pick.fna
sym98_161012_seqs_for_denovo_pick.fna sym99_161012_seqs_for_denovo_pick.fna sym100_161013_seqs_for_denovo_pick.fna >
concat_sym_1to100_seqs_for_denovo_pick.fna
count_seqs.py -i concat_sym_1to100_seqs_for_denovo_pick.fna -o concat_sym_1to100_seqs_for_denovo_pick_fna_seq_count.txt
 # use filter command to relabel with sample number
 # get unique sequences for clustering
usearch10 -fastx_uniques cat_sym_1to100_reads_for_denovo_pick.fna -sizeout -fastaout
cat_sym_1to100_reads_for_denovo_pick_Uniqs.fa
00:44 4.0Gb 100.0% Reading cat_sym_1to100_reads_for_denovo_pick.fna
00:44 3.9Gb CPU has 16 cores, defaulting to 10 threads
00:50 6.2Gb 100.0% DF
00:51 6.3Gb 9382875 seqs, 7658500 uniques, 7331721 singletons (95.7%)
00:51 6.3Gb Min size 1, median 1, max 105358, avg 1.23
01:56 4.8Gb 100.0% Writing cat_sym_1to100_reads_for_denovo_pick_Uniqs.fa
 # cluster using usearch10 with n=1 - ONLY FOR TAXON ASSIGNMENT -- TOO LARGE TO MAKE TABLE
usearch10 -cluster_otus cat_sym_1to100_reads_for_denovo_pick_Uniqs.fa -minsize 2 -relabel OTU -otus
cat_sym_1to100_reads_for_denovo_pick_Uniqs_OTUrepSet.fa
05:13 89Mb 100.0% 5869 OTUs, 46132 chimeras
 # assign names with Qiime against Greengenes for consistency with closed picked OTUs (note: need to attach these to OTUs in table using biom command)
assign_taxonomy.py -i cat_sym_1to100_reads_for_denovo_pick_Uniqs_OTUrepSet.fa -m uclust -r
/home/james/qiime_software/gg_13_8_otus/rep_set/97_otus.fasta -t
/home/james/qiime_software/gg_13_8_otus/taxonomy/97_otu_taxonomy.txt --uclust_min_consensus_fraction 0.51 --uclust_similarity 0.9
--uclust_max_accepts 3 -o uclust_tax_assign/
 # cluster using usearch10 with n=100
usearch10 -cluster_otus cat_sym_1to100_reads_for_denovo_pick_Uniqs.fa -minsize 100 -relabel OTU -otus
cat_sym_1to100_reads_for_denovo_pick_Uniqs_OTUrepSet_n10.fa
usearch10 -otutab cat_sym_1to100_reads_for_denovo_pick.fna -otus
cat_sym_1to100_reads_for_denovo_pick_Uniqs_OTUrepSet_n100.fa -biomout cat_sym_1to100_denovoOTUtable.json -mapout
denovoOTUs_map.txt -notmatched denovo_unmapped_reads.fa -dbmatched denovo_matched_reads_with_sizes.fa -sizeout
usearch10 -otutab cat_sym_1to100_reads_for_denovo_pick.fna -otus
cat_sym_1to100_reads_for_denovo_pick_Uniqs_OTUrepSet_n100.fa -otutabout cat_sym_1to100_denovoOTUtable_n100.txt
_____
biom convert -i cat_sym_1to100_denovoOTUtable_n100.txt -o cat_sym_1to100_denovoOTUtable_n100.biom --table-type='OTU table' --
process-obs-metadata taxonomy
biom add-metadata --sc-separated taxonomy --observation-header OTUID,taxonomy --observation-metadata-fp
cat_sym_1to100_reads_for_denovo_pick_Uniqs_OTUrepSet_tax_assignments.txt -i cat_sym_1to100_denovoOTUtable_n100.biom -o
cat_sym_1to100_denovoOTUtable_n100_wTaxa.biom
biom summarize-table -i cat_sym_1to100_denovoOTUtable_n100_wTaxa.biom -o
cat_sym_1to100_denovoOTUtable_n100_wTaxa_biom_summary.txt
 # make .txt version of table to check labels etc
biom convert -i cat_sym_1to100_denovoOTUtable_n100_wTaxa.biom -o cat_sym_1to100_denovoOTUtable_n100_wTaxa_biom.txt -b --
header-key taxonomy
_____
 # filter closed pick table at n100 to match
filter_otus_from_otu_table.py -i merged_symbiota_closed_picked_table_1to100.biom -n 100 -o
merged_symbiota_closed_picked_table_1to100_n100.biom
 # make text version of table to quality check
biom convert -i merged_symbiota_closed_picked_table_1to100_n100.biom -o
merged_symbiota_closed_picked_table_1to100_n100_biom_.txt -b --header-key taxonomy

biom summarize-table -i merged_symbiota_closed_picked_table_1to100_n100.biom -o merged_symbiota_closed_picked_table_1to100_n100_biom_summary.txt

_____

 # merge closed picked and denovo picked tables

merge_otu_tables.py -i cat_sym_1to100_denovoOTUtable_n100_wTaxa.biom,merged_symbiota_closed_picked_table_1to100_n100.biom -o FINAL_merged_closed_and_denovo_picked_OTUtable_filter100.biom

biom summarize-table -i FINAL_merged_closed_and_denovo_picked_OTUtable_filter100.biom -o FINAL_merged_closed_and_denovo_picked_OTUtable_filter100_biom_summary.txt

biom convert -i merged_closed_denovo_sym1to100_Feb12_2017.biom -o merged_closed_denovo_sym1to100_Feb12_2017_biom.txt -b --header-key taxonomy