## Supplementary Online Content

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*. Published online May 8, 2017. doi:10.1001/jamapediatrics.2017.0378

**eMethods.** Detailed Methodology
**eFigure 1.** Quality Control
**eFigure 2.** Right vs Left Breast
**eFigure 3.** Modeling of OTU Networks at Early Development (1st Month of Life) and Between 4 and 6 Months of Life
**eFigure 4.** Breast Milk Influences Composition in a Dose-Dependent Manner
**eFigure 5.** Bacterial Taxa That Discriminate Microbial Communities in Infant Stool Using a Random Forest Algorithm Based on Feeding Status
**eFigure 6.** Introduction of Solid Foods Into the Infant Diet Results in Changes in the Microbial Community
**eFigure 7.** Differences in the Infant Stool Microbiome Arising From the Amount of Daily Breastfeeding Persist Even After Solid Foods Are Introduced
**eFigure 8.** Functional Capabilities of Breast Milk Microbial Community Predicted Using PICRUSt
**eFigure 9.** Differences in Predicted Metagenomic Functional Analysis of Infant Stool Dependent on Feeding Method
**eTable 1.** Characteristics of the Infant's Feeding Type at the Time of Each Unique Sample Collection
**eTable 2.** Factors Contributing To The Variation In The Microbial Community Of Infant Stool From Adonis Multivariate Analysis of Variance Using the Bray-Curtis and Unweighted UniFrac Distance Matrices
**eTable 3.** Distance Comparison of Microbial Communities Between True Compared With Random Mother-Infant Pairs (Wilcoxon Rank-Sum Test)
**eTable 4.** OTUs With Significant Difference in Sharing Rate Between True Mother-Infant Pairs Compared to Random Mother-Infant Pairs When Comparing Bacterial Communities in Breast Milk and Infant Stool
**eTable 5.** Oligotyping Analysis Performed Using Sequences Mapped to Otus Shared Within Mother-Infant Dyads

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

## **eMethods.** Detailed Methodology

### DNA extraction, library preparation, and sequencing

Frozen breast milk samples and areolar skin samples were extracted using the BiOstic Bacteremia DNA isolation kit (MOBIO Laboratories, Carlsbad, California, USA). We homogenized the frozen raw stool samples in DNA stabilizer (Stratec Molecular, Berlin, Germany) and followed the manufacturer's instructions for the PSP Spin Stool DNA Kit (Stratec Molecular, Berlin, Germany), substituting Lysing Matrix E tubes (MP Biomedicals, Burlingame, California, USA) for the provided zirconia beads. The 16S rDNA was amplified in triplicate and barcoded using a previously published protocol.[1] The protocol utilized the V4 region of the 16S rRNA gene. Illumina (San Diego, California, USA) flow cell adapter sequences and a twelve base pair barcode region were incorporated into the PCR primers. DNA amplicon concentrations were then quantified on the 2100 Bioanalyzer and 2200 TapeStation (Agilent Technologies, Santa Clara, California, USA).

We followed the sequencing protocol as presented previously by Caporaso, et al.[1] Briefly, we pooled the amplicons and diluted to 2nM. Amplicons were then denatured and loaded onto a (Illumina, San Diego, California, USA) using 2x150bp version 2 chemistry following the manufacturer's instructions. Reagent controls including DNA stabilization buffer and isolation kit reagents were sequenced alongside the samples. The reagent controls had significantly lower read counts compared to breast milk, areolar skin or infant stool samples (eFigure 1).

### Quantitative PCR

Samples, reagent controls, and a standard were preamplified prior to RT-PCR as per Fluidigm's Biomark protocol using the universal primers targeting the V3/V4 region of the 16S ribosomal RNA gene (340F- TCCTACGGGAGGCAGCAGT, 806R-GGACTACCAGGGTATCTAATCCTGTT)[2,3] obtained from Eurofins Scientific, Luxembourg. The same primers used for preamplification and TaqMan probe (V3/V4 16S probe- 6-Fam-5'-CGTATTACCGCGGCTGCTGGCAC-3') were used for RT-PCR on Fluidigm's Biomark HX system. The standard used was a full length 16S gene cloned in TOPO Cloning Vector [4] and tenfold serial dilutions of the standard were used to set up a calibration curve-based analysis. All samples and standards were run in triplicate and data are presented as means. Data analysis was performed using Fluidigm BioMark software, following instructions provided in the RT-PCR Analysis document for calibration curve (e.g. "standard curve") based analysis. Assays that failed the BioMark quality filtering were excluded from further analysis. For reagent controls, almost all assays were below the limit of detection. Therefore, we assigned censored values of 50% of the lower limit of detection according to the standard curve (10 copies per 5uL) to all reagent control assays that did not pass the quality checks. Results were exported as .csv files and additional analyses were performed using R statistical software version 3.0.3.

**Sequence data processing and analysis**

Sequence data (median 166,199 reads per specimen) was processed using Quantitative Insights Into Microbial Ecology (QIIME) 1.8.0.[5] Paired-end reads were joined using fastq-join.[6] Sequences were demultiplexed and quality filtered using default QIIME parameters. 16S rRNA Operational Taxonomic Units (OTUs) were picked using an open reference OTU picking procedure (QIIME script, pick_open_reference_otus.py) such that sequences were clustered against the Greengenes 13.8 database at 97% identity; those failing to match were clustered de novo. Chimera checking was performed using ChimeraSlayer with standard options as implemented in QIIME, followed by filtering of OTUs representing fewer than 0.005% of all sequences. These standard quality filtering methods[5] reduced 238,035 OTUs to 478 OTUs for analysis.

Of 863 samples sequenced, we excluded 84 duplicate samples, 15 samples with fewer than 200 reads, and 4 samples from infants who received antibiotics. We analyzed the remaining 760 samples including 326 milk, 114 areolar skin, 232 stool samples and 58 controls (20 breast milk, 5 stool, and 33 reagent blanks). Quantitative bacterial PCR results are shown in Supplementary Figure 1. The core microbiome was defined as the OTUs present in at least 50% of the samples and identified using QIIME (compute_core_microbiome.py). Faith's phylogenetic diversity (PD) was computed and rarefaction analyses were conducted in QIIME (multiple_rarefaction.py, alpha_diversity.py and collate_ alpha.py). Beta-diversity analyses of community similarity were performed by calculating pairwise distances using the phylogenetic metric UniFrac distance. The resulting distance matrices were used for principle coordinates analyses (PCoA).

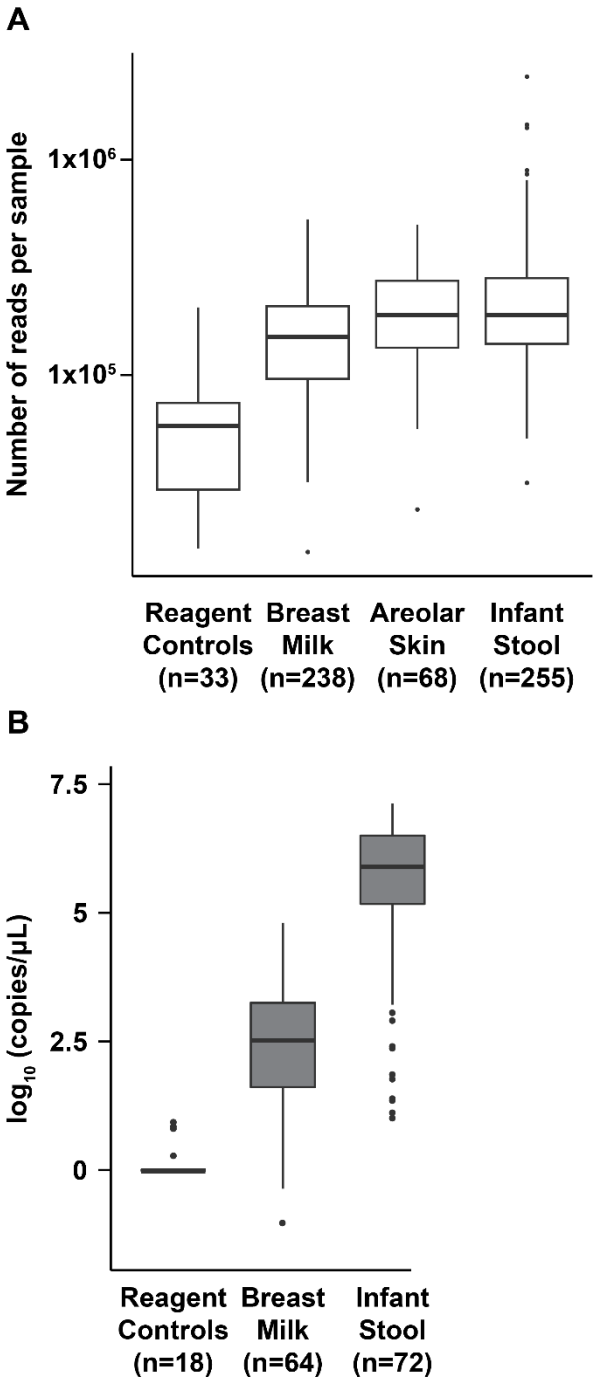
 

**eReferences**


1. Caporaso JG, Lauber CL, Walters WA, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 2012;6(8):1621-1624.
2. Nadkarni MA, Martin FE, Jacques NA, Hunter N. Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology.* 2002;148(Pt 1):257-266.
3. Stiverson J, Williams T, Chen J, et al. Prebiotic Oligosaccharides: Comparative Evaluation Using In Vitro Cultures of Infants' Fecal Microbiomes. *Appl Environ Microbiol.* 2014;80(23):7388-7397.
4. Chehoud C, Dryga A, Hwang Y, et al. Transfer of Viral Communities between Human Individuals during Fecal Microbiota Transplantation. *MBio.* 2016;7(2).
5. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods.* 2010;7(5):335-336.
6. Aronesty E. ea-utils : Command-line tools for processing biological sequencing data. 2011; http://code.google.com/p/ea-utils.

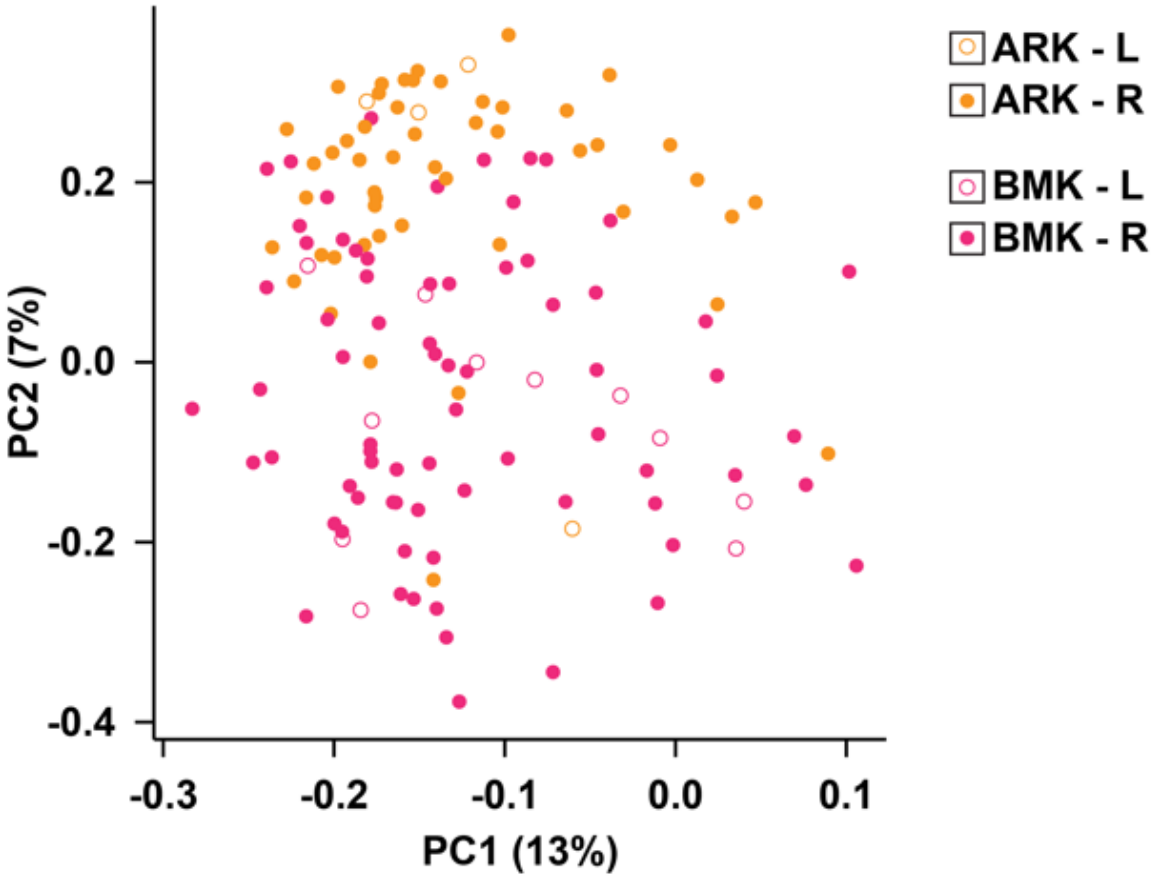

 

**eFigure 1. Quality Control**. (a) Read counts were significantly lower in the reagent controls compared to milk (p=1.3e-11), areolar skin (p=2.7e-13), and stool samples (p=9.8e-18) (Wilcoxon rank-sum test). (b) Quantitative PCR with primers targeting the 16S ribosomal RNA gene identified bacteria in all sample types with significantly higher bacterial cell counts in stool versus breast milk samples (p=2.7e-15, Wilcoxon rank-sum test). Reagent controls had significantly lower bacterial cell counts compared to breast milk or stool (p=6.1e-11 and p=2.3e-16, respectively, Wilcoxon rank-sum test).



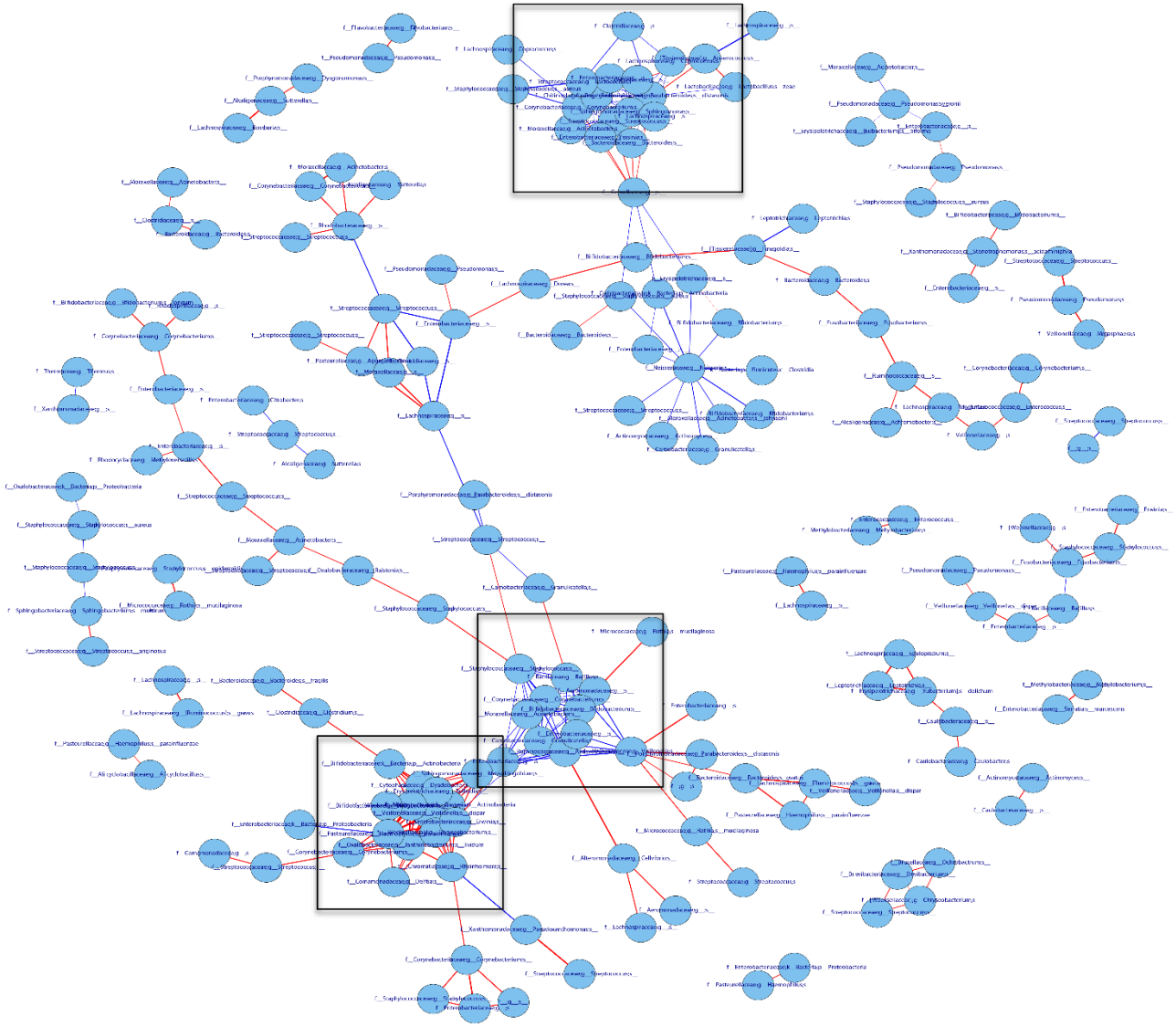

 

**eFigure 2**. **Right vs Left Breast.** No differences were detected between samples from the right compared to left breast milk samples or areolar skin swabs.



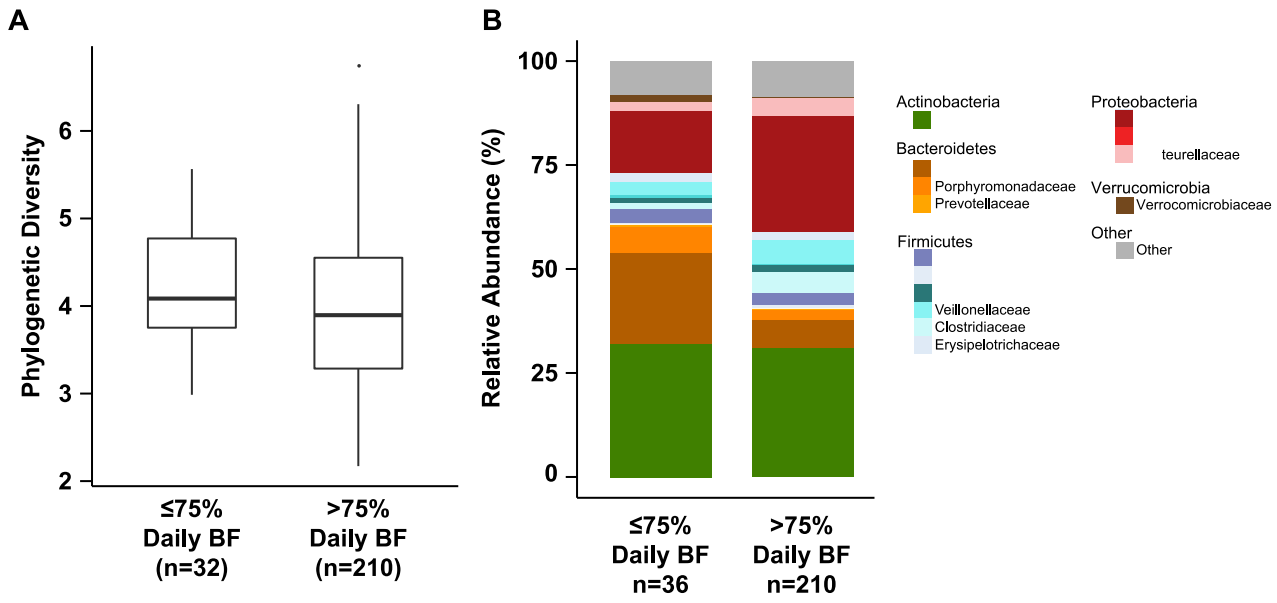

 

**eFigure 3. Modeling of OTU Networks at Early Development (1st Month of Life) and Between 4 and 6 Months of Life.** Spearman's correlation coefficients (cc) > 0.6 or < -0.6 and $p < 0.05$ were compared; the significance of the difference between two stages was tested by permutation. Edges were colored according to the Spearman's cc. Blue edges are correlations that decrease with age (i.e. higher correlation in 0-30 days vs. 120-180 days). Red represents correlations that increase with age. Multiple hubs of bacteria have abundance changes that are directly or inversely correlated.



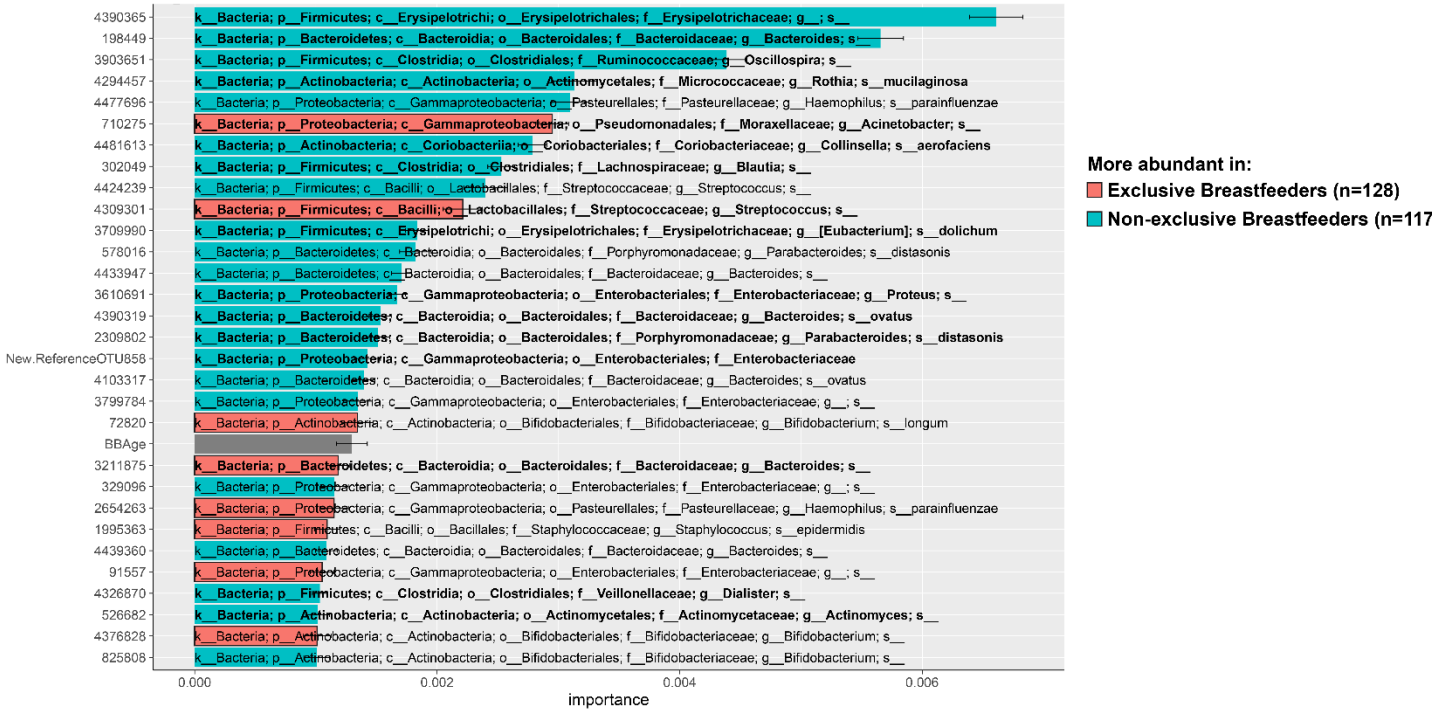
**eFigure 4. Breast Milk Influences Composition in a Dose-Dependent Manner.** (a) Alpha diversity as measured by Faith's phylogenetic diversity shows no difference between primarily breastfed infants and those who breastfed for ≤75% of their daily milk intake (p=0.063) (b) Infants who breastfeed for ≤75% of their daily milk intake have significantly more Bacteroidaceae (FDR_p=0.016) and Erysipelotrichaceae (FDR_p=0.019) compared to primarily breastfed infants.



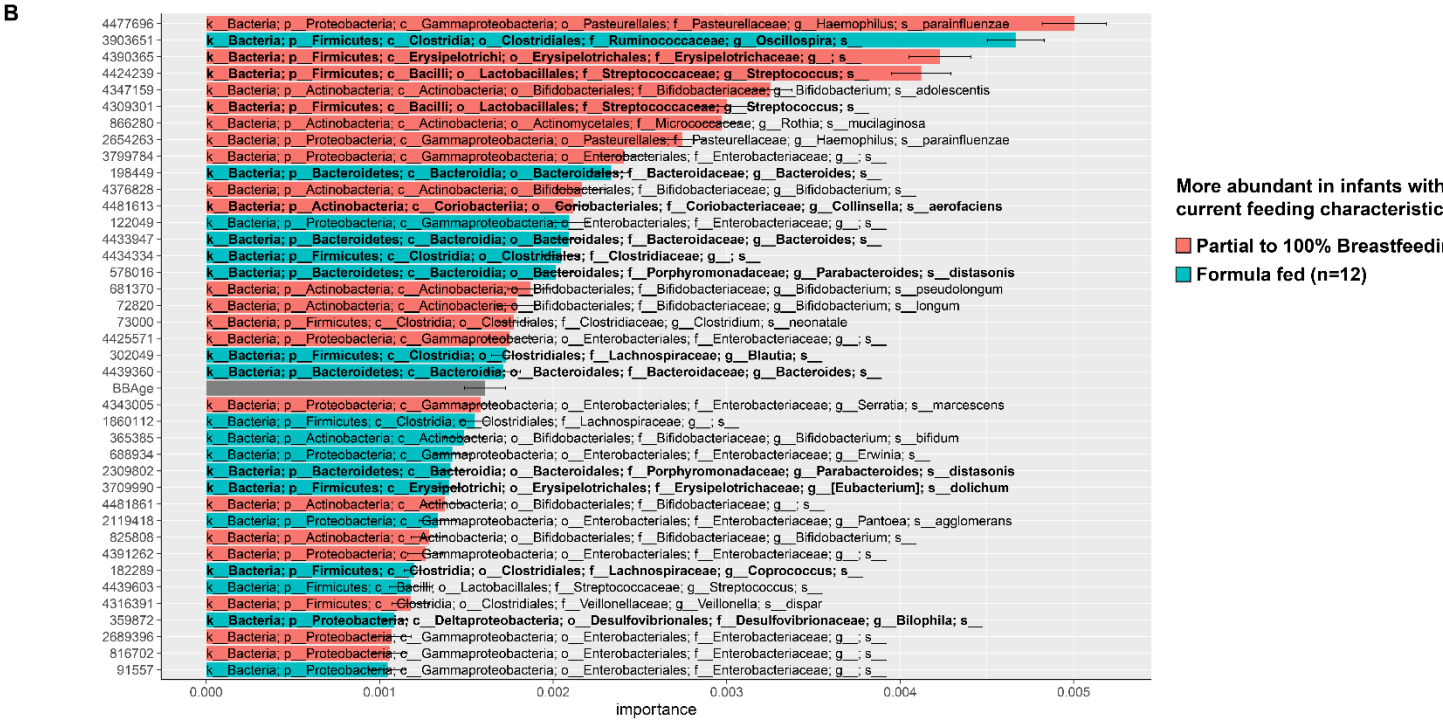

 

**eFigure 5. Bacterial Taxa That Discriminate Microbial Communities in Infant Stool Using a Random Forest Algorithm Based on Feeding Status** in (a) Exclusive versus non-exclusive breastfeeders since birth and (b) infants receiving any amount of breast milk versus only formula feeding at time of sample collection. Taxa reaching statistical significance (FDR p-value <0.1) are bolded.

A

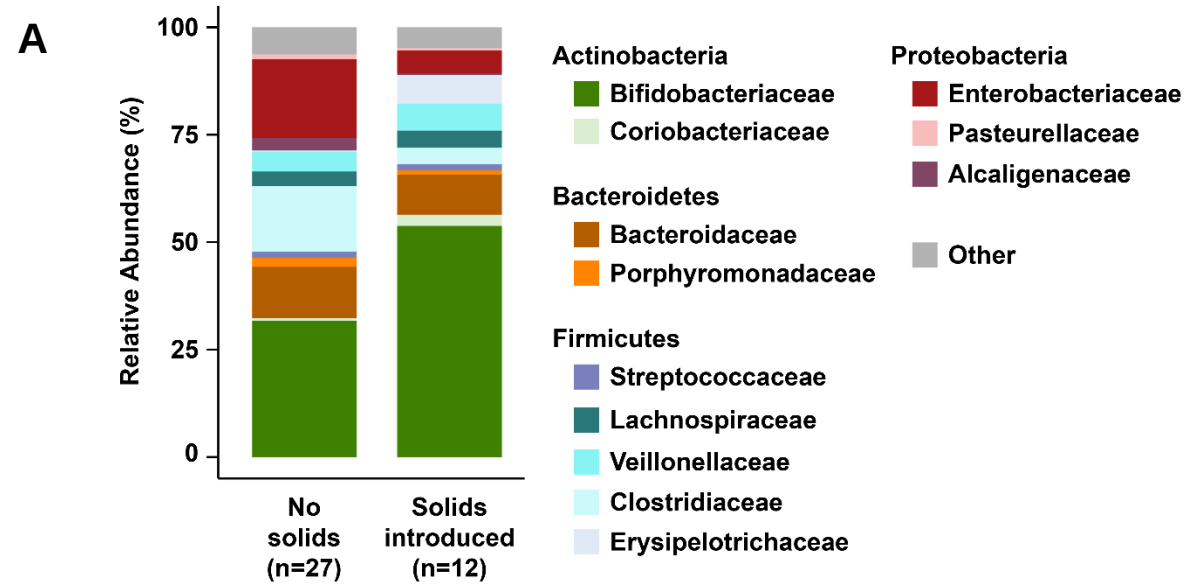


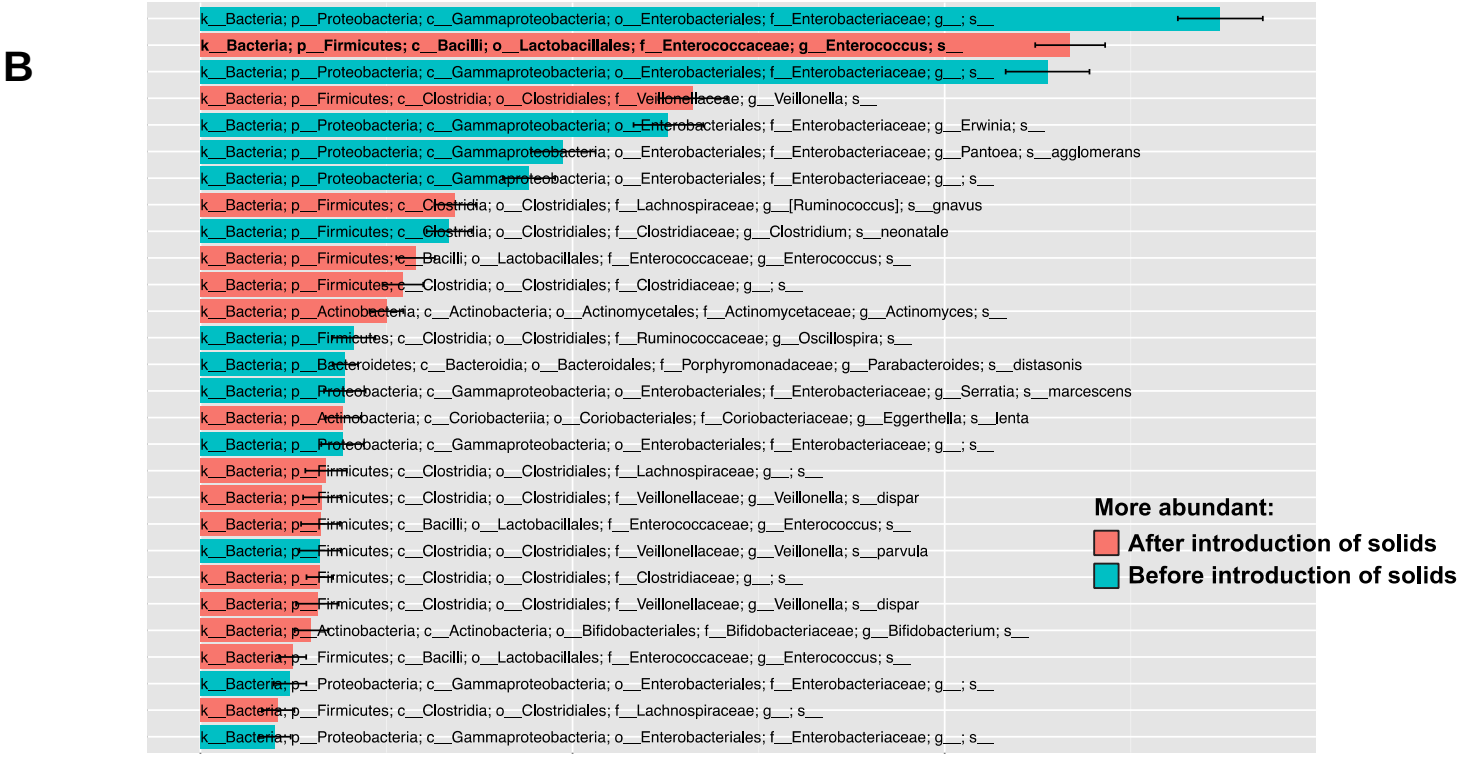

B



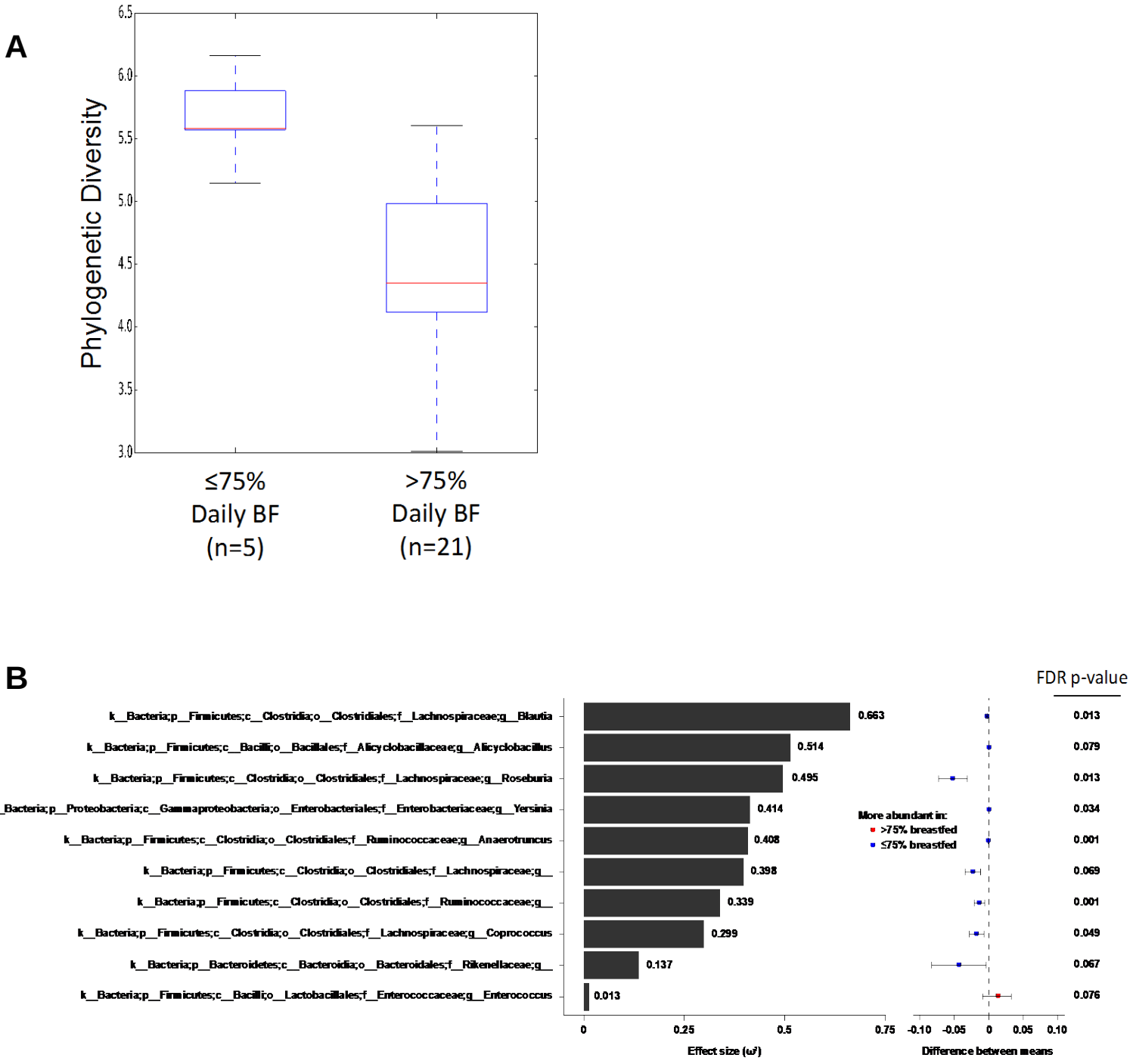
**eFigure 6. Introduction of Solid Foods Into the Infant Diet Results in Changes in the Microbial Community.** Analysis was controlled for age by evaluating infants 4 to 6 months of age. (a) Introduction of solid foods changes the relative abundance of multiple taxa at the family level, especially Enterobacteriaceae (FDR_p=0.083), Erysipelotrichaceae (FDR_p=0.083) and Verrucomicrobiaceae (FDR_p=0.083) (b) Random forest model displays taxa that discriminate the microbiota of infants before and after solids introduction.



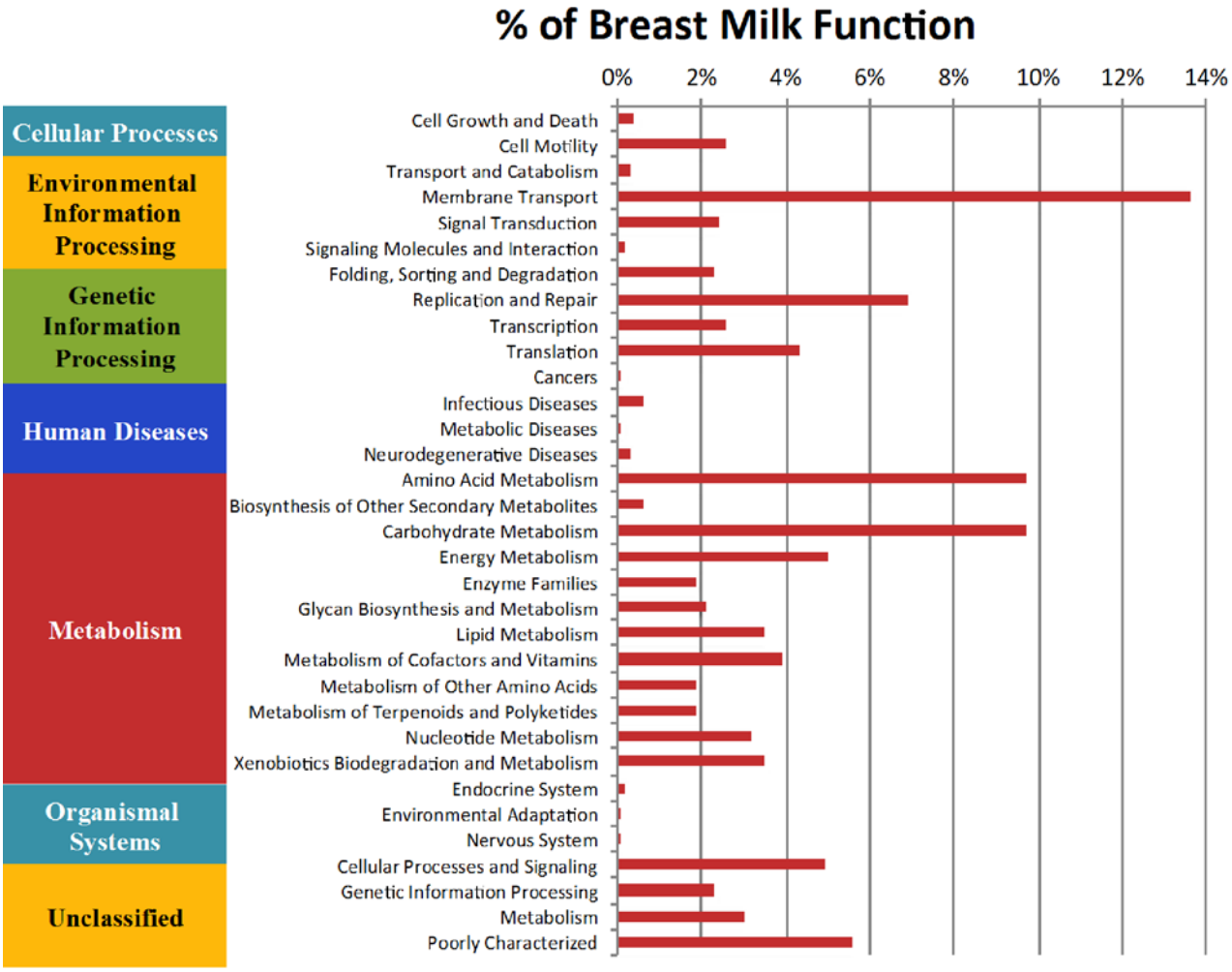

 

**eFigure 7. Differences in the Infant Stool Microbiome Arising From the Amount of Daily Breastfeeding Persist Even After Solid Foods Are Introduced.** Infants who primarily breastfeed have (a) lower alpha diversity (Faith's Phylogenetic diversity, p=0.003) and (b) different community membership.



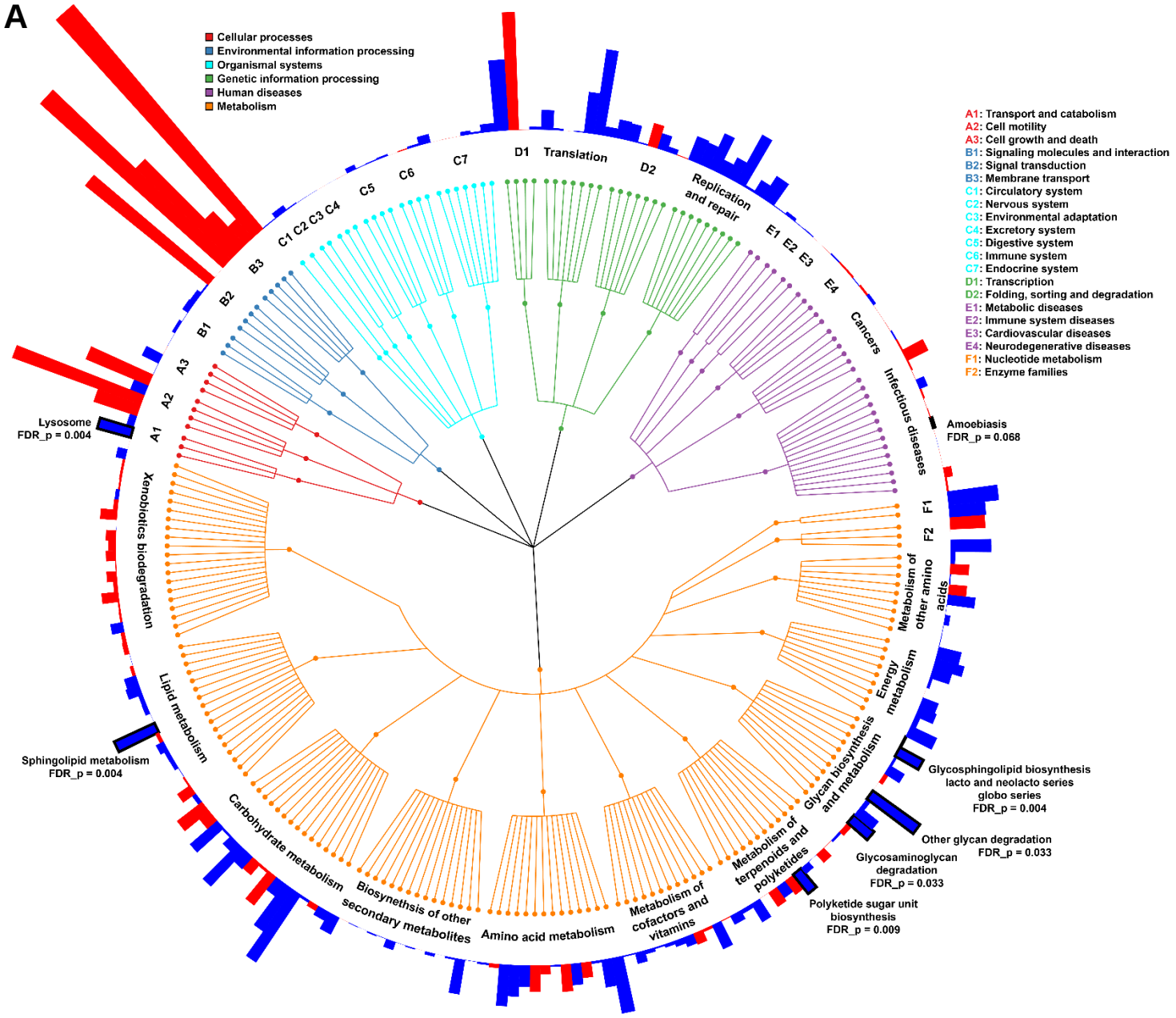

 

**eFigure 8. Functional Capabilities of Breast Milk Microbial Community Predicted Using PICRUSt.** Human breast milk contains bacteria with high abundance in gene families associated with membrane transport and carbohydrate, amino acid and energy metabolism functions.



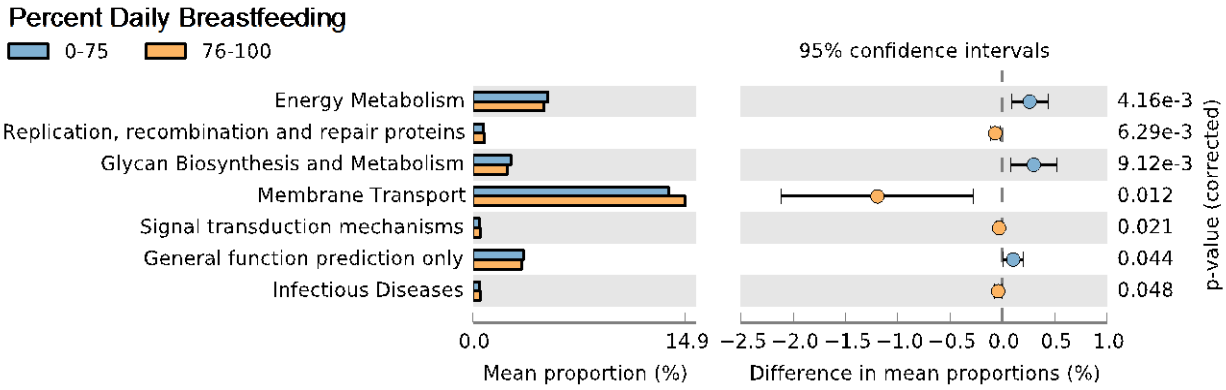

 

**eFigure 9. Differences in Predicted Metagenomic Functional Analysis of Infant Stool Dependent on Feeding Method.** (a) Multiple levels of functional differences among infants with daily breast milk intake of ≤75% compared to >75% of daily milk intake using PICRUSt are displayed here. Blue bars indicate higher average counts of functional genes across samples in infants who breastfeed ≤75%; Red bars indicate higher average counts of functional genes across samples in infants who breastfeed >75%. (b) Predicted mean proportion of infant stool microbe function with 95% confidence intervals by percent daily breastfeeding. (c) Differences in function based on age (days) of solid introduction.

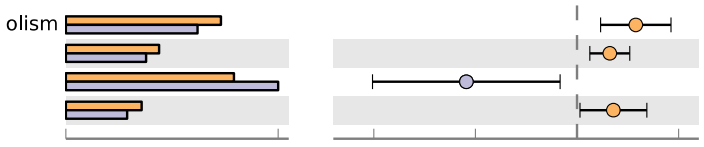


**B**




**C** Age of Solids Introduction



**eTable 1**. Characteristics of the Infant's Feeding Type at the Time of Each Unique Sample Collection (N=569)

| | |
|---|---|
| **Current milk feeding type** | |
| Breast milk only | 370 (65.0) |
| Mixed feeds (milk + formula) | 166 (29.2) |
| 76-99% breast milk | 114 (68.7) |
| 1-75% breast milk | 45 (27.1) |
| Unknown | 7 (4.2) |
| Formula only | 14 (2.5) |
| Unknown | 19 (3.3) |
| **Exclusive Breastfeeding*** | |
| Yes | 298 (52.4) |
| No | 252 (44.3) |
| Unknown | 19 (3.3) |
| **Solids Introduced** | |
| Yes | 51 (9.0) |
| No | 506 (88.9) |
| Unknown | 12 (2.1) |

*Exclusive breastfeeding is defined as having never received formula.

**eTable 2.** Factors Contributing To The Variation In The Microbial Community Of Infant Stool From Adonis Multivariate Analysis of Variance Using the Bray-Curtis and Unweighted UniFrac Distance Matrices.

| Characteristic | Unweighted UniFrac | | Bray-Curtis | |
|---|---|---|---|---|
| | $R^2$ | p-value | $R^2$ | p-value |
| Infant's Age (days) | 0.23 | 0.001* | 0.22 | 0.001* |
| Age of Formula Introduction | 0.023 | 0.001* | 0.030 | 0.001* |
| Age of Solids Introduction | 0.029 | 0.001* | 0.042 | 0.001* |
| Percent Daily Breastfeeding[1] | 0.008 | 0.002* | 0.007 | 0.025* |
| Delivery Method | 0.009 | 0.001* | 0.011 | 0.002* |
| Mother's Age | 0.005 | 0.087 | 0.005 | 0.11 |
| Race | 0.018 | 0.018* | 0.016 | 0.21 |
| Geographical location | 0.004 | 0.22 | 0.005 | 0.13 |
| Run | 0.12 | 0.001* | 0.077 | 0.001* |
| Residuals | 0.55 | | 0.59 | |

*statistically significant

[1]Percentage Daily Breastfeeding was calculated by the number of breastfeeding events divided by the sum of breastfeeding and formula feeding events per day.

**eTable 3.** Distance Comparison of Microbial Communities Between True Compared With Random Mother-Infant Pairs (Wilcoxon Rank-Sum Test).

| Distance Type | Breast milk – Infant stool | | | Areolar – Infant stool | | |
|---|---|---|---|---|---|---|
| | True pair Distance (mean ± SD) | Random pair distance (mean ± SD) | p-value | True pair Distance (mean ± SD) | Random pair distance (mean ± SD) | p-value |
| Bray-Curtis | 0.93 ± 0.12 | 0.95 ± 0.10 | <0.001 | 0.95 ± 0.08 | 0.96 ± 0.06 | <0.001 |
| Chao-Jaccard | 0.62 ± 0.26 | 0.66 ± 0.25 | 0.001 | 0.52 ± 0.23 | 0.65 ± 0.25 | <0.001 |
| Jaccard | 0.80 ± 0.06 | 0.81 ± 0.06 | <0.001 | 0.82 ± 0.06 | 0.83 ± 0.06 | 0.006 |
| Unweighted UniFrac | 0.62 ± 0.07 | 0.63 ± 0.07 | <0.001 | 0.65 ± 0.06 | 0.66 ±0.05 | 0.015 |
| Weighted UniFrac | 0.44 ± 0.11 | 0.44 ± 0.11 | 0.20 | 0.49 ± 0.09 | 0.50 ± 0.08 | 0.12 |

**eTable 4**. OTUs With Significant Difference in Sharing Rate Between True Mother-Infant Pairs Compared to Random Mother-Infant Pairs When Comparing Bacterial Communities in Breast Milk and Infant Stool.

| OTU[1] | Bacteria | | | | Percent shared in mom-infant pairs | Percent shared in random pair permutations (± SD[2]) | FDR p-value |
|---|---|---|---|---|---|---|---|
| | Phylum | Family | Genus | Species | | | |
| 4439603 | Firmicutes | Streptococcaceae | Streptococcus | | 44.4 | 34.7 ± 2.2 | 0.008 |
| 4411138 | Actinobacteria | Micrococcaceae | Rothia | mucilaginosa | 22.2 | 14.7 ± 2.2 | 0.038 |
| 4451251 | Actinobacteria | Coriobacteriaceae | Atopobium | | 11.1 | 3.8 ± 1.4 | 0.008 |
| 4329518 | Actinobacteria | Weeksellaceae | Chryseobacterium | | 12.8 | 5.6 ± 1.7 | 0.008 |
| 4294457 | Actinobacteria | Micrococcaceae | Rothia | mucilaginosa | 17.9 | 11.3 ± 2.1 | 0.040 |
| 4316391 | Firmicutes | Veillonellaceae | Veillonella | dispar | 12.0 | 5.8 ± 1.6 | 0.011 |
| 12574 | Actinobacteria | Actinomycetaceae | Actinomyces | | 10.3 | 4.2 ± 1.5 | 0.011 |
| 866280 | Actinobacteria | Micrococcaceae | Rothia | mucilaginosa | 9.4 | 3.4 ± 1.4 | 0.008 |
| 740317 | Proteobacteria | Bradyrhizobiaceae | Bradyrhizobium | | 7.7 | 1.7 ± 1.0 | 0.008 |
| 4410401 | Firmicutes | Veillonellaceae | Veillonella | dispar | 16.2 | 10.4 ± 2.0 | 0.058 |
| 4453501 | Firmicutes | Veillonellaceae | Veillonella | dispar | 46.2 | 40.3 ± 1.9 | 0.048 |
| 72820 | Actinobacteria | Bifidobacteriaceae | Bifidobacterium | longum | 57.3 | 51.5 ± 1.7 | 0.031 |
| 4471251 | Proteobacteria | Pasteurellaceae | Haemophilus | | 16.2 | 10.5 ± 2.0 | 0.096 |
| 4390365 | Firmicutes | Erysipelotrichaceae | | | 9.4 | 4.2 ± 1.5 | 0.041 |
| 73000 | Firmicutes | Clostridiaceae | Clostridium | | 6.8 | 2.0 ± 1.1 | 0.011 |
| 109413 | Proteobacteria | Pasteurellaceae | Actinobacillus | | 6.0 | 1.4 ± 1.0 | 0.008 |
| 538000 | Proteobacteria | Enterobacteriales | Enterobacteriaceae | | 6.0 | 1.6 ± 1.0 | 0.019 |
| 4363066 | Proteobacteria | Pasteurellaceae | Aggregatibacter | | 6.8 | 2.5 ± 1.2 | 0.040 |
| 4473129 | Proteobacteria | Pasteurellaceae | Haemophilus | parainfluenzae | 4.3 | 0.9 ± 0.8 | 0.022 |
| 526682 | Actinobacteria | Actinomycetaceae | Actinomyces | | 4.3 | 1.0 ± 0.8 | 0.040 |
| 1109251 | Proteobacteria | Pseudomonadaceae | Pseudomonas | | 4.3 | 1.5 ± 0.9 | 0.086 |
| 2438948 | Actinobacteria | Bifidobacteriaceae | Scardovia | | 2.6 | 0.2 ± 0.4 | 0.024 |
| 2530636 | Firmicutes | Veillonellaceae | Megamonas | | 2.6 | 0.3 ± 0.5 | 0.040 |
| 4315658 | Firmicutes | Lactobacillaceae | Lactobacillus | | 1.7 | 0.1 ± 0.2 | 0.029 |
| 4440670 | Firmicutes | Veillonellaceae | Veillonella | | 1.7 | 0.1 ± 0.3 | 0.050 |
| 851938 | Firmicutes | Erysipelotrichaceae | Bulleidia | | 1.7 | 0.1 ± 0.3 | 0.084 |

[1]OTU identity from Greengenes
[2]SD, Standard deviation

**eTable 5.** Oligotyping Analysis Performed Using Sequences Mapped to Otus Shared Within Mother-Infant Dyads. Oligotypes with significant sharing rate differences between true mother-infant pairs compared to random mother-infant pairs when comparing bacterial communities in breast milk and infant stool are shown.

| Oligotype | OTU | Family | Genus | Species | Percent shared in mom-infant pairs | Percent shared in random pair permutations (± SD) | FDR p-value |
|---|---|---|---|---|---|---|---|
| TCCAAGTGGG | 4439603 | Streptococcaceae | Streptococcus | | 14.3 | 4.6 ± 1.9 | 0.023 |
| TTGGTGCAGG | 4453501 | Veillonellaceae | Veillonella | dispar | 33.7 | 22.9 ± 2.6 | 0.023 |
| TTGGTACAGA | 4453501 | Veillonellaceae | Veillonella | dispar | 31.5 | 20.6 ± 2.6 | 0.031 |
| TCCGAATTGG | 4439603 | Streptococcaceae | Streptococcus | | 13.0 | 4.7 ± 1.9 | 0.039 |
| TCCACGTTGG | 4439603 | Streptococcaceae | Streptococcus | | 5.2 | 0.5 ± 0.8 | 0.039 |
| TTGGTGGAGG | 4453501 | Veillonellaceae | Veillonella | dispar | 23.6 | 14.1 ± 2.4 | 0.039 |
| TTGGTACGGG | 4453501 | Veillonellaceae | Veillonella | dispar | 19.1 | 10.0 ± 2.3 | 0.039 |
| TCCACATTGG | 4439603 | Streptococcaceae | Streptococcus | | 24.7 | 14.2 ± 2.7 | 0.046 |
| GTGAAATTGG | 4439603 | Streptococcaceae | Streptococcus | | 16.9 | 8.6 ± 2.3 | 0.046 |
| TCCCAATTGG | 4439603 | Streptococcaceae | Streptococcus | | 13.0 | 5.1 ± 2.0 | 0.046 |
| TTGGTAGAGA | 4453501 | Veillonellaceae | Veillonella | dispar | 13.5 | 6.2 ± 1.9 | 0.046 |
| TCTAAATTGG | 4439603 | Streptococcaceae | Streptococcus | | 11.7 | 4.3 ± 1.9 | 0.053 |
| GCGCTAAAGG | 4294457 | Micrococcaceae | Rothia | mucilaginosa | 43.4 | 31.0 ± 3.4 | 0.053 |
| TTGGTACACG | 4453501 | Veillonellaceae | Veillonella | dispar | 3.4 | 0.3 ± 0.5 | 0.053 |
| GACTTAAGGG | 4411138 | Micrococcaceae | Rothia | mucilaginosa | 13.3 | 4.1 ± 2.4 | 0.064 |