Supplemental Methods

Sample Sequencing

The V4-V5 hypervariable region of 16S ribosomal DNA were fusion-primer amplified from purified genomic DNA from each sample and sequenced using the Illumina MiSeq platform at the Josephine Bay Paul Marine Biological Laboratory in Woods Hole, MA, as described previously (1,2). All samples were routinely amplified in triplicate 33 uL reaction volumes and there was a no-template control for each primer set (the rest of the master mix was the same). For a PCR positive control, a bacterial mock community DNA preparation was used. Forward fusion primers contained one of eight 5-nucleotide barcodes between the Illumina-specific bridge and sequencing primer regions and the 16S-specific region. The single reverse fusion primer contained one of 12 Illumina indices. The combination of forward and reverse primers allows for 96 unique barcode-index combinations, and thus multiplexing of up to 96 samples per lane. The amplification cocktail contained Platinum Taq Hi-Fidelity Polymerase (Life Technologies), 1X Hi-Fidelity buffer, 200 uM dNTP PurePeak DNA polymerase mix (ThermoFisher), 2 mM MgSO4 and 0.2 uM of each primer. Each library pool was quantitated by qPCR (Kapa Biosystems) and then sequenced on the Illumina Miseq platform with paired end reads of 250nt each. Datasets were demultiplexed by index by on-instrument software and separated by barcode using a custom script. Illumina-utils (https://github.com/meren/illumina-utils) programs merged paired end reads into consensus full-length v4v5 sequences, trimmed the adapters, and filtered for quality. 66% of the nucleotides in the non-overlapping region were required to have a quality score greater than Q30 score and no more than 3 mismatches in the area of overlap are accepted. Chimeric reads were identified by vsearch (3) and removed from the dataset. Samples resulting in fewer than 10,000 reads were not included in this study.

Taxonomic assignment was performed using the GAST algorithm (4) as described previously (5). All new data generated have been deposited into the National Center for Biotechnology Information Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) with accession numbers SRP007679 and SRP064159. Additionally, sequences and taxonomy data were deposited in the Marine Biological Laboratory Visualization and Analysis of Microbial Population Structures (https://vamps.mbl.edu) (6).

Data Analysis

Bacterial alpha-diversity was evaluated using Simpson's diversity index (SDI), which is a measure of both the number (richness) and the relative abundance of genera in a sample (7). We evaluated differences in bacterial alpha-diversity and abundance by building linear mixed effects models that accounted for repeated measures, age (day of life at time of sample collection), and exposures. Where feasible, corrected gestational age (gestational age at birth + day of life) was also adjusted for. We considered unadjusted models as well as models adjusted for exposures. Rarefaction was performed prior to alpha-diversity analysis. We removed taxa that occurred in fewer than 10 samples from the dataset for analyses of bacterial abundance. If groups existed within the dataset (for example, preterm and term infants), each taxon was required to have at least one sample with non-zero counts within each group. If the *nlme* function in R still produced an error, then taxa for which *nlme* could not analyse bacterial abundance differences were removed. To avoid type I errors when evaluating differences or changes in abundance of a large number of bacterial genera, we controlled the false discovery rate by adjusting p-values for multiple comparisons using the R function *p.adjust*.

To perform hierarchical clustering, we used the R function *pheatmap* with the Euclidean distance metric and complete linkage method. The program *pvclust*

(http://www.is.titech.ac.jp/~shimo/prog/pvclust/) was used in R to calculate the statistical significance of clusters using multiscale bootstrap resampling. Statistical significance was set at alpha = 0.05.

Bacterial beta-diversity, a measure of bacterial phylogenetic relatedness between pairs of samples, was characterized by generalized UniFrac distances (8). FASTA files containing bacterial 16S DNA sequence data were downloaded from VAMPS (vamps.mbl.edu). Open reference OTU picking was then performed using the command pick_open_reference_otus.py in the MacQIIME program using default settings to yield the OTU table and phylogenetic tree (9). The phylogenetic tree was midpoint rooted, and the R function *GUniFrac* from the package "GUniFrac" was used to compute generalized UniFrac distances between samples using the OTU table and phylogenetic tree file.

A permutational multivariate analysis of variance using distance matrices was used to assess differences in bacterial phylogenetic distances (phylogenetic relatedness) between groups. This was performed using the R function *adonis* (from the R package "vegan"). Analysis of dispersion between groups found no evidence of presence of significant dispersion difference by permutation testing (Pr(>F)=0.68). Therefore, *adonis* results were not due to differences in group dispersions. An ANOVA analysis of the difference in dispersion between groups was also not significant (Pr(>F)=0.665), so there was not significant variation between the two groups. Thus, it is unlikely that dispersion heterogeneity confounds our analysis.

To account for the effects of exposures on bacterial alpha-diversity and bacterial phylogenetic relatedness, we included the exposures in linear mixed effects models and when evaluating bacterial phylogenetic relatedness differences. When adjusting for exposures, samples missing data on at least one exposure were removed. We were not able to account for both the

effects of exposures and repeated measures specifically when using the R function *adonis* due to limitations on equation structure; therefore, infant stool samples were considered independent samples when analyzing bacterial phylogenetic relatedness. Only the 6 week timepoint was used in *adonis* analyses as it had the largest overlap between the premature and term infants. Day of life 0-5 was also considered as a timepoint, but there would only be n=4 premature infant stool samples to compare to, which would be too large a loss of power to be useful.

References:

1. Newton RJ, McLellan SL, Dila DK, et al. Sewage reflects the microbiomes of human populations. MBio 2015;6:1–9.
2. Eren AM, Vineis JH, Morrison HG, Sogin ML. A Filtering Method to Generate High Quality Short Reads Using Illumina Paired-End Technology. PLoS One 2013;8:6–11.
3. Rognes T, Mahé F, Flouri T, McDonald D, Schloss P. vsearch: VSEARCH 1.4.0 [Internet]. 2015;Available from: http://dx.doi.org/10.5281/zenodo.31443
4. Huse SM, Dethlefsen L, Huber JA, Mark Welch D, Relman DA, Sogin ML. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. PLoS Genet 2008;4:e1000255.
5. Chernikova DA, Koestler DC, Hoen AG, et al. Fetal exposures and perinatal influences on the stool microbiota of premature infants. J Matern Fetal Neonatal Med [Internet] 2016 [cited 2014 Dec 17];29:99–105. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25394613
6. Huse SM, Mark Welch DB, Voorhis A, et al. VAMPS: a website for visualization and analysis of microbial population structures. BMC Bioinformatics 2014;15:41.

7. SIMPSON EH. Measurement of Diversity. Nature. 1949;163:688–688.

8. Chen J, Bittinger K, Charlson ES, et al. Associating microbiome composition with environmental covariates using generalized UniFrac distances. Bioinformatics 2012;28:2106–13.

9. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods [Internet] 2010;7:335–6. Available from: http://dx.doi.org/10.1038/nmeth0510-335