









# Nutrition and the Gut Microbiota in 10- to 18-Month-Old Children Living in Urban Slums of Mumbai, India

 Samantha L. Huey,<sup>a</sup>  Lingjing Jiang,<sup>k</sup>  Marcus W. Fedarko,<sup>c,d</sup>  Daniel McDonald,<sup>e</sup>  Cameron Martino,<sup>c,f</sup>  Farhana Ali,<sup>b</sup>  David G. Russell,<sup>g</sup>  Shobha A. Udipi,<sup>h</sup>  Aparna Thorat,<sup>h</sup>  Varsha Thakker,<sup>h</sup>  Padmini Ghugre,<sup>h</sup>  R. D. Potdar,<sup>i</sup>  Harsha Chopra,<sup>i</sup>  Kripa Rajagopalan,<sup>a</sup>  Jere D. Haas,<sup>a</sup>  Julia L. Finkelstein,<sup>a,j</sup>  Rob Knight,<sup>b,c,d,e</sup>  Saurabh Mehta<sup>a,j</sup>

<sup>a</sup>Division of Nutritional Sciences, Cornell University, Ithaca, New York, USA

<sup>b</sup>Department of Bioengineering, University of California, San Diego, California, USA

<sup>c</sup>Center for Microbiome Innovation, University of California San Diego, La Jolla, California, USA

<sup>d</sup>Department of Computer Science and Engineering, University of California San Diego, La Jolla, California, USA

<sup>e</sup>Department of Pediatrics, School of Medicine, University of California San Diego, La Jolla, California, USA

<sup>f</sup>Bioinformatics and Systems Biology Program, University of California San Diego, La Jolla, California, USA

<sup>g</sup>Department of Microbiology and Immunology, Cornell University, Ithaca, New York, USA

<sup>h</sup>Department of Nutrition and Food Science, SNDT Women's University, Mumbai, India

<sup>i</sup>Centre for the Study of Social Change, Mumbai, India

<sup>j</sup>Institute for Nutritional Sciences, Global Health, and Technology (INSIGHT), Cornell University, Ithaca, New York, USA

<sup>k</sup>Division of Biostatistics, University of California, San Diego, California, USA

**ABSTRACT** In this cross-sectional study, we describe the composition and diversity of the gut microbiota among undernourished children living in urban slums of Mumbai, India, and determine how nutritional status, including anthropometric measurements, dietary intakes from complementary foods, feeding practices, and micronutrient concentrations, is associated with their gut microbiota. We collected rectal swabs from children aged 10 to 18 months living in urban slums of Mumbai participating in a randomized controlled feeding trial and conducted 16S rRNA sequencing to determine the composition of the gut microbiota. Across the study cohort, *Proteobacteria* dominated the gut microbiota at over 80% relative abundance, with *Actinobacteria* representation at <4%, suggesting immaturity of the gut. Increased microbial  $\alpha$ -diversity was associated with current breastfeeding, greater head circumference, higher fat intake, and lower hemoglobin concentration and weight-for-length Z-score. In redundancy analyses, 47% of the variation in Faith's phylogenetic diversity (Faith's PD) could be accounted for by age and by iron and polyunsaturated fatty acid intakes. Differences in community structure ( $\beta$ -diversity) of the microbiota were observed among those consuming fats and oils the previous day compared to those not consuming fats and oils the previous day. Our findings suggest that growth, diet, and feeding practices are associated with gut microbiota metrics in undernourished children, whose gut microbiota were comprised mainly of *Proteobacteria*, a phylum containing many potentially pathogenic taxa.

**IMPORTANCE** The impact of comprehensive nutritional status, defined as growth, nutritional blood biomarkers, dietary intakes, and feeding practices, on the gut microbiome in children living in low-resource settings has remained underreported in microbiome research. Among undernourished children living in urban slums of Mumbai, India, we observed a high relative abundance of *Proteobacteria*, a phylum including many potentially pathogenic species similar to the composition in preterm infants, suggesting immaturity of the gut, or potentially a high inflammatory burden. We found head circumference, fat and iron intake, and current breastfeeding were positively associated with microbial diversity, while hemoglobin and weight for length were associated with lower diversity. Findings suggest that examining com-

**Citation** Huey SL, Jiang L, Fedarko MW, McDonald D, Martino C, Ali F, Russell DG, Udipi SA, Thorat A, Thakker V, Ghugre P, Potdar RD, Chopra H, Rajagopalan K, Haas JD, Finkelstein JL, Knight R, Mehta S. 2020. Nutrition and the gut microbiota in 10- to 18-month-old children living in urban slums of Mumbai, India. *mSphere* 5:e00731-20. <https://doi.org/10.1128/mSphere.00731-20>.

**Editor** Vincent B. Young, University of Michigan—Ann Arbor

**Copyright** © 2020 Huey et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Saurabh Mehta, [smehta@cornell.edu](mailto:smehta@cornell.edu).

 New paper from @MehtaRG shows diet, feeding practices and growth associated with microbial diversity and composition in children living in urban slums of Mumbai, with @slhuey @KnightLabNews in @mSphereJ @CornellNutri @Cornell @HarvestPlus

**Received** 17 July 2020

**Accepted** 3 September 2020

**Published** 23 September 2020

prehensive nutrition is critical to gain more understanding of how nutrition and the gut microbiota are linked, particularly in vulnerable populations such as children in urban slum settings.

**KEYWORDS** infant, child, growth, diet, nutrition, feeding practices, microbiome, fat intake, anthropometry

The dynamics of the developing gut microbiota in later infancy and early childhood have not been fully explored, compared to early infancy, and particularly in low-resource settings. Perturbations in the microbiota of infants and young children have been associated with many factors, including nutrition status, diet, delivery mode, infections, and immune function, including antibiotic or other medication administration (1). It is accepted that the infant gut microbiota are first represented by facultative anaerobes in the *Proteobacteria* phylum and succeed in favor of *Actinobacteria* and in particular, *Bifidobacteria* spp. with the introduction of breast milk (2). By the time the child is 2 to 3 years of age, the microbiota in early life reach an adult-like balance mainly consisting of *Firmicutes* and *Bacteroidetes* phyla with depleted *Proteobacteria* and even lower *Actinobacteria* (2). However, the factors associated with this transition are less understood.

Diet plays a direct role in the gut microbiota composition and function (3). Food components indigestible by humans (such as fiber [4]) are broken down by commensal microbiota in the gut to later serve as signaling molecules such as short-chain fatty acids (SCFAs) for other human cells such as immune cells as well as nutrients for commensal bacteria (5, 6). In infants, both indirect and direct dietary intakes are thought to be a major driver of gut composition, from maternal diet during pregnancy and lactation (7), to consumption by infants. Breast milk and formula have also been shown to be associated with differences in the composition of the gut microbiota; for example, breastfed infants tend to have lower bacterial diversity (8) and a higher relative abundance of beneficial *Bifidobacterium* species, than formula-fed infants who tend to have higher abundances of *Escherichia coli* and *Clostridium difficile* (9). However, few studies have examined changes in the gut microbiota during weaning, with the introduction to complementary foods.

It is well established that immune function and health outcomes differ by diet and nutritional status during the first few years of life, particularly in low-resource settings (10). Ascertaining differences in the gut microbiota by nutritional status and dietary intake may represent a mediating factor in immune functional capacity, as gut microbiota and immune function develop in tandem (1), involving commensal microbiota-host immune cell cross talk, signaling, and education (11, 12).

Most studies examining early life development of the gut microbiota have been performed in higher-income settings (13–29) or in low-income settings such as Bangladesh (30) and Malawi (31). However, an analysis of the gut microbiota during the second year of life in young children living in a low-resource urban slum setting of Mumbai, India—where poor growth (as measured by anthropometric length-for-age Z-score [LAZ], weight-for-age Z-score [WAZ], and weight-for-length Z-score [WLZ] [32]), infections, and poor sanitation are common (33–35)—has yet to be described. Therefore, the objective of this cross-sectional analysis was to (i) characterize the gut microbiota among 10- to 18-month-old children in Mumbai's urban slums and (ii) determine the association between comprehensive nutritional status (as determined by anthropometric measurements, blood nutritional biomarkers, dietary intakes, and feeding practices) and relative abundance,  $\alpha$ -diversity, and  $\beta$ -diversity of the gut microbiota.

## RESULTS

**Demographic characteristics.** Participant characteristics of the samples ( $n = 53$ ) are described in Table 1. Children were sampled from five urban slum communities in Mumbai, India. Nearly 30% of children were stunted (LAZ <  $-2$ ), 25% underweight (WAZ <  $-2$ ), and 12% wasted (WLZ <  $-2$ ). From a subset of participants with

**TABLE 1** Participant characteristics

Parameter	<i>n</i> <sup>a</sup>	Median (IQR) or <i>n</i> (%)
<b>Sociodemographic</b>		
Age (mo)	53	14.8 (13.1, 16.7)
Female	53	25 (47.2)
Vaginally delivered (versus Caesarean)	51	29 (56.9)
<b>Anthropometry</b>		
Birth weight (kg)	53	2.7 (2.5, 3.0)
Low birth weight (<2.5 kg)	53	10 (18.9)
Current weight (kg)	53	8.7 (8.0, 9.8)
Mid-upper arm circumference (cm)	53	14.8 (14.1, 15.1)
Head circumference (cm)	53	44.2 (43.0, 45.4)
Head circumference-for-age Z-score	52	-1.58 (-2.12, -0.73)
Length (cm)	51	74.0 (72.1, 77.7)
Length-for-age Z-score (LAZ)	51	-1.29 (-2.38, -0.46)
Stunting	51	15 (29.4)
Weight-for-age Z-score (WAZ)	53	-0.96 (-1.99, -0.34)
Underweight	53	13 (24.5)
Weight-for-length Z-score (WLZ)	51	-0.68 (-1.34, 0.08)
Wasting	51	6 (11.8)
<b>Blood biomarkers and illness history</b>		
Ferritin (ng/ml)	44	7.55 (3.20, 15.70)
Iron deficiency (<12 ng/ml)	44	27 (61.4)
Zinc ( $\mu$ mol/liter)	38	13.18 (11.08, 15.28)
Zinc deficiency (<9 $\mu$ mol/liter)	38	0 (0)
Hemoglobin (g/dl)	43	10.10 (9.10, 10.80)
Anemia (hemoglobin < 11 g/dl)	43	33 (76.7)
C-reactive protein (CRP) $\geq$ 5 mg/liter	39	3 (7.7)
Diarrhea today or within past 4 weeks	51	9 (17.7)
Fever today or within past 4 weeks	51	19 (37.3)
Cough today or within past 4 weeks	51	7 (13.7)
<b>Dietary intakes<sup>b</sup></b>		
Calories (kcal)	52	393.0 (270.0, 645.5)
Protein (g)	52	13.6 (8.1, 21.6)
Fat (g)	52	12.9 (8.9, 20.8)
Saturated fat (g)	52	4.5 (0.8, 5.8)
Monounsaturated fat (g)	52	2.2 (0.4, 3.0)
Polyunsaturated fat (g)	52	0.4 (0.3, 0.7)
Carbohydrate (g)	52	65.3 (33.6, 85.9)
Fiber (g)	52	2.0 (0.5, 5.5)
Calcium (mg)	52	216.5 (62.0, 349.5)
Iron (mg)	52	2.0 (1.2, 3.3)
Zinc (mg)	52	1.3 (0.8, 2.0)
Vitamin A ( $\mu$ g RAE)	52	99.0 (17.0, 141.0)
Cobalamin (vitamin B <sub>12</sub> ) ( $\mu$ g)	52	0 (0, 0)
<b>Feeding practices</b>		
Diet: Vegetarian (including eggs)	51	7 (13.7)
Diet: Vegetarian (no eggs)	51	15 (29.4)
Diet: Nonvegetarian	51	29 (56.9)
Ever breastfed	52	48 (92.3)
Breastfed yesterday (current breastfeeding)	48	40 (83.3)
Exclusive breastfeeding duration (months)	51	7.0 (6.0, 7.0)
Exclusively breastfed > 6 months	51	33 (64.7)
Consumed grains (bread, rice, noodles, porridge) yesterday <sup>c</sup>	51	47 (92.2)
Consumed any fruits and vegetables yesterday <sup>d</sup>	51	23 (45.1)
Liver, kidney, heart, or other organ meats <sup>e</sup>	50	2 (4.0)
Consumed beef, pork, poultry yesterday <sup>f</sup>	51	7 (13.7)
Consumed eggs yesterday <sup>g</sup>	51	10 (19.6)
Consumed dried fish or seafood yesterday <sup>h</sup>	51	2 (3.9)
Consumed beans, peas, lentils, nuts, or seeds yesterday <sup>i</sup>	51	22 (43.1)
Consumed dairy yesterday <sup>j</sup>	50	26 (52.0)
Consumed oil or butter yesterday <sup>k</sup>	51	13 (25.5)
Consumed sugary foods yesterday <sup>l</sup>	51	30 (58.8)
Consumed condiments yesterday <sup>m</sup>	50	6 (12.0)

(Continued on next page)

**TABLE 1** (Continued)

<sup>a</sup>*n* is the number of participants in the study.

<sup>b</sup>Dietary intakes as absolute (unadjusted) values.

<sup>c</sup>Infant and Young Child Feeding (IYCF) (World Health Organization) food group A.

<sup>d</sup>IYCF food groups B to F combined (orange and starchy root vegetables, dark leafy green vegetables, ripe mangoes or papayas, any other fruits and vegetables).

<sup>e</sup>IYCF food group G.

<sup>f</sup>IYCF food group H.

<sup>g</sup>IYCF food group I.

<sup>h</sup>IYCF food group J.

<sup>i</sup>IYCF food group K.

<sup>j</sup>IYCF food group L.

<sup>k</sup>IYCF food group M.

<sup>l</sup>IYCF food group N.

<sup>m</sup>IYCF food group O.

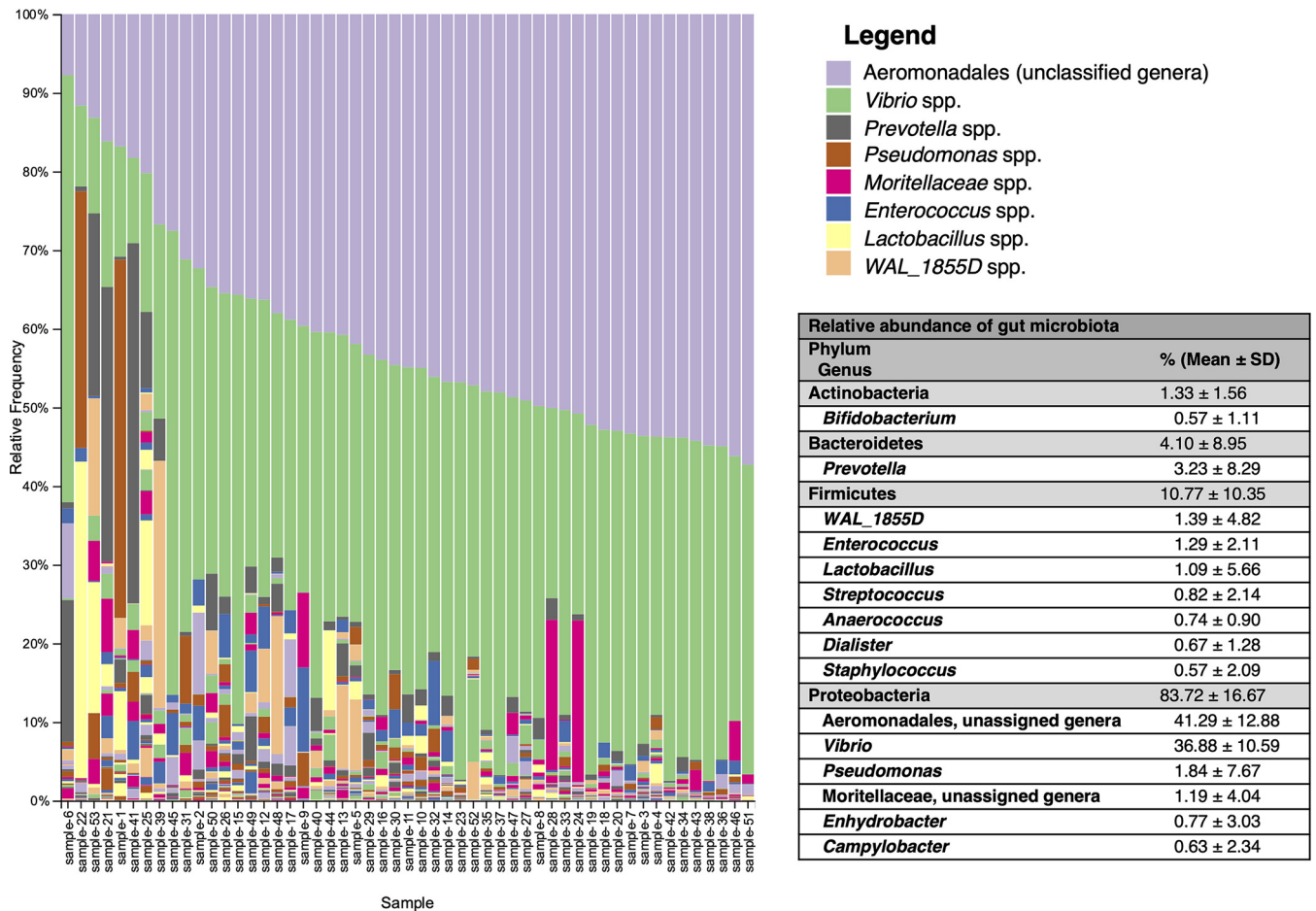
hemoglobin data, 77% were anemic, and from a subset of participants with nutritional blood biomarker data, 74% were iron deficient. Most children consumed a nonvegetarian diet and were reported to be currently breastfeeding. Comparison of characteristics between the included cohort in this report ( $n = 53$ ) and the rest of the screened population ( $n = 312$ ) from the parent trial is in Table S1 in the supplemental material. Participants did not differ in most characteristics; however, more children were born by Caesarean section (compared to vaginal birth) in the current study.

**Gut microbiota 16S sequencing.** From 53 participants, a total of 16,778,710 forward and reverse reads were processed. After initial quality filtering and trimming (see Materials and Methods for more details) (36), the number of joined paired-end reads totaled 7,554,901, with 9,251 unique sequences across the study population (median 138,928 [interquartile range {IQR}, 94,704, 175,232] reads per sample). Filtering out mitochondria and chloroplasts removed 18,119 sequences for a total frequency of 7,536,782 sequence variants, including 2,248 unique sequences (median 138,834 {IQR, 94,357, 174,909}) per sample, with a minimum sequence frequency per sample of 38,012.

**Relative abundance.** Rarefaction to 38,012 reads per sample retained 2,014,636 sequences (26.73% of total sequences). Genera of the phylum *Proteobacteria* (Fig. 1; see also Fig. S1 in the supplemental material) dominantly represented the gut microbiota. Approximately 128 identified and unidentified genera were found across the population. The *Aeromonadales* order (unspecified genera) and genus *Vibrio* spp. combined represented over 75% of all sequences across all participants (Fig. 1). Additional genera represented at greater abundance (>1%) included *Prevotella*, *Pseudomonas*, and *Enterococcus*, with genera at minimal representation (mean abundance of <1%) including *Streptococcus*, *Enhydrobacter*, *Anaerococcus*, *Dialister*, *Campylobacter*, *Bifidobacterium*, and *Staphylococcus* (Fig. 1). A full legend of all genera may be found in Fig. S1.

After correction for multiple comparisons using the Benjamini-Hochberg false discovery rate (FDR) (37), multivariate linear regressions showed no nutritional exposures associated with percent relative abundance of the four main phyla present in gut microbiota: *Proteobacteria*, *Firmicutes*, *Bacteroidetes*, or *Actinobacteria* (Table 2). All linear regressions examining associations between nutritional exposures and phylum relative abundance are included in Table S2.

**$\alpha$ -Diversity.** Across the study population,  $\alpha$ -diversity metrics included Shannon diversity index (SDI), a measure of taxon diversity and evenness (38), and Faith's phylogenetic diversity (Faith's PD), which accounts for the phylogenetic distance between taxa within each sample (39). The median (IQR) SDI and Faith's PD were 3.77 (3.39, 4.33) and 13.41 (11.15, 15.32), respectively. In multivariate linear regression, greater head circumference was positively associated with a 0.23 (95% confidence interval [95% CI], 0.09, 0.37) unit increase in SDI, while higher weight-for-length Z-score was associated with a 0.31 (95% CI, 0.13, 0.49) unit decrease in SDI (Table 3). In sex-stratified analysis, WLZ remained inversely associated with SDI only in male children (−0.35 [95% CI, −0.57, −0.13];  $P = 0.002$ ), while head circumference remained positively associated with SDI only in female children (0.28 [95% CI, 0.16, 0.40];  $P < 0.0001$ ).



**FIG 1** Relative abundance of genera across participants. Each stacked bar plot corresponds to one infant subject. Figure legend colors repeat for additional identified taxa. Please see Fig. S1B in the supplemental material for full legend.

Head circumference was also associated with a 0.77 (95% CI, 0.37, 1.20) unit increase in Faith’s PD. Other nutritional exposures positively associated with Faith’s PD included fat intake, current breastfeeding, and child’s age in multivariate regression, while increasing hemoglobin concentration was inversely associated with Faith’s PD, and female children had lower Faith’s PD than male children. All associations remained significant after correction for multiple comparisons (Table 3), and all linear regressions examining  $\alpha$ -diversity as an outcome are shown in Table S3.

Redundancy analysis (RDA) of sociodemographic, clinical, dietary intakes as nutrient residuals adjusted for energy, and feeding practices revealed that after removing colinear variables, the age of the child had the largest explanatory power (27%) on variation in Faith’s PD, followed by iron intake (15%) and intake of polyunsaturated fatty acids (5%) (Fig. 2; details may be found in Table S4). No correlates were associated with SDI in redundancy analysis.

**$\beta$ -Diversity.** Differences in community structure between groups were measured using unweighted and weighted UniFrac distances and tested for significant differences by permutational multivariate analysis of variance (PERMANOVA) (40) as well as PERMDISP (41) to test for dispersion, in addition to DEICODE robust Aitchison distances (42) and Qurro (43) to determine specific taxa driving clustering (Table S5). No nutritional exposures were associated with differences in community structures measured by either unweighted or weighted UniFrac distances (Table 4). Robust Aitchison distances produced by DEICODE assessed by PERMANOVA were significantly different between samples from subjects who did and did not consume oils and fats the previous day

**TABLE 2** Correlates of phylum relative abundance in multivariate analyses<sup>a</sup>

Taxon and parameter	n	$\beta$ (%) (95% CI)	P value <sup>b</sup>	P value adj. <sup>c</sup>
<i>Actinobacteria</i> (%)				
Head circumference (cm)	52	0.30 (0.02, 0.57)	<b>0.0327</b>	0.0719
Age (mo)	52	0.14 (−0.05, 0.33)	0.1439	0.1759
Female	52	−0.39 (−1.23, 0.45)	0.3660	0.4026
<i>Bacteroidetes</i> (%)				
Serum zinc ( $\mu$ mol/liter)	38	1.41 (0.21, 2.61)	<b>0.0211</b>	0.0719
Age (mo)	38	0.51 (−0.83, 1.85)	0.4570	0.4570
Female	38	4.63 (−1.40, 10.66)	0.1320	0.1759
<i>Firmicutes</i> (%)				
Iron (mg) <sup>d</sup>	48	7.50 (1.98, 13.03)	<b>0.0078</b>	0.0572
Log energy (kcal) <sup>d</sup>	48	−3.81 (−8.27, 0.65)	0.0944	0.1731
Age (mo)	48	1.34 (0.16, 2.53)	<b>0.0264</b>	0.0719
Female	48	4.05 (−1.17, 9.27)	0.1280	0.1759
<i>Proteobacteria</i> (%)				
Weight-for-length Z-score (WLZ) <sup>e</sup>	51	5.33 (1.25, 9.41)	<b>0.0104</b>	0.0572

<sup>a</sup>Analyses were performed using complete case analysis; similar results were found using missing indicators with median imputation (not shown).

<sup>b</sup>Boldface P values are statistically significant ( $P < 0.05$ ).

<sup>c</sup>P value adj., adjusted P value corrected for multiple comparisons using the Benjamini-Hochberg false discovery rate (FDR).

<sup>d</sup>Iron intake residual adjusted for energy.

<sup>e</sup>Not adjusted for age and sex, as calculation of WLZ incorporates age and sex.

(Infant and Young Child Feeding [IYCF] food group M; Fig. 3A,  $P = 0.04$ ) (Table 4; Table S6). The specially separated arrows in the compositional biplot revealed ratios of features classified of the phylum *Firmicutes* (lowest classified taxonomic level; *Enterococcaceae*, *Enterococcus*, *Lactococcus*, *Anaerococcus*, *WAL\_1855D*) and *Proteobacteria* (lowest classified taxonomic level; *Aeromonadales*, *Moritellaceae*, *Vibrio*) that drove sample separation (Fig. 3A; Fig. S2A; see Table S5 for full list of features). By inspecting the biplot, we selected features spatially separating oil and fat consumption and visualized the corresponding log ratios using Qurro, which were assessed by *t* tests (Fig. 3A). We repeated this for 17 separate log ratios according to the cluster in which they were identified. We found that the log ratio of *Lactococcus* to *Anaerococcus* was significantly higher in the group consuming oils and fats (median [IQR], 2.62 [1.05, 6.15])

**TABLE 3** Correlates of  $\alpha$ -diversity in multivariate analyses<sup>a</sup>

Correlate and parameter	n	$\beta$ (95% CI)	P value <sup>b</sup>	P value adj. <sup>c</sup>
Shannon diversity index (SDI)				
Head circumference (cm)	51	0.21 (0.08, 0.33)	<b>0.001</b>	<b>0.0023</b>
Weight-for-length Z-score (WLZ)	51	−0.31 (−0.49, −0.13)	<b>0.0007</b>	<b>0.0021</b>
Faith's phylogenetic diversity (PD)				
Head circumference (cm)	37	0.77 (0.34, 1.20)	<b>0.0004</b>	<b>0.0018</b>
Hemoglobin (g/dl)	37	−0.58 (−1.06, −0.10)	<b>0.0183</b>	<b>0.0206</b>
Fat (g) <sup>d</sup>	37	2.91 (1.33, 4.48)	<b>0.0003</b>	<b>0.0018</b>
Log energy (kcal) <sup>e</sup>	37	−1.58 (−2.67, −0.49)	<b>0.0045</b>	<b>0.0081</b>
Breastfed yesterday (current breastfeeding)	37	2.27 (0.53, 4.02)	<b>0.0106</b>	<b>0.0153</b>
Age (mo)	37	0.31 (0.03, 0.60)	<b>0.0298</b>	<b>0.0298</b>
Female	37	−1.58 (−2.81, −0.35)	<b>0.0119</b>	<b>0.0153</b>

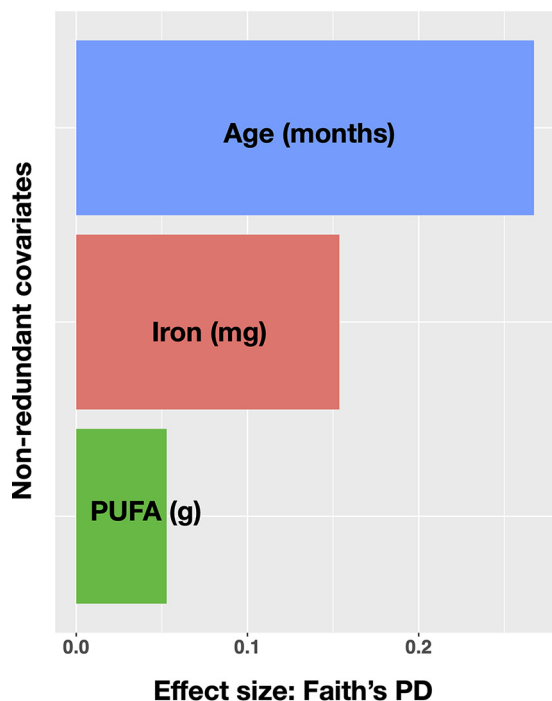
<sup>a</sup>Analyses were performed using complete case analysis; similar results were found using missing indicators with median imputation (not shown).

<sup>b</sup>Boldface P values are statistically significant ( $P < 0.05$ ).

<sup>c</sup>P value adj., adjusted P value corrected for multiple comparisons using the Benjamini-Hochberg false discovery rate (FDR).

<sup>d</sup>Not adjusted for age and sex, as calculation of WLZ incorporates age and sex.

<sup>e</sup>Fat residual adjusted for energy.



**FIG 2** Redundancy analysis (RDA) for Faith's phylogenetic diversity (Faith's PD) of sociodemographic, clinical, feeding practices, and dietary correlates (dietary intakes are nutrient residuals adjusted for energy after removing colinear variables). Factors are sorted according to their effect size in the sample population and colored for distinguishability. PUFA, polyunsaturated fatty acids.

versus 1.05 [−2.25, 3.57];  $P = 0.01$ ) (Fig. 3B). Figure S2B shows the Qurro rank plot with the *Lactococcus/Anaerococcus* log ratio highlighted.

**DISCUSSION**

In this cross-sectional study, we examined the gut microbiota of 10- to 18-month-old children living in urban slums of Mumbai, India, and determined associations with comprehensive nutritional status. Overall, children were undernourished with high proportions of poor growth and nutrient deficiency, with the *Proteobacteria* phylum representing the majority of taxa in their gut microbiota. Multivariate analyses revealed differences in gut microbiota composition and measures of diversity in association with nutritional markers, including anthropometry indicators such as head circumference and weight for length; blood nutritional biomarkers, including hemoglobin; dietary fat and iron intakes; and feeding practices, such as current breastfeeding and consumption of fats and oils. Taxa from the *Firmicutes* and *Proteobacteria* phyla were identified as driving the gut microbiome sample separation globally in this cohort. Among these

**TABLE 4** Correlates of  $\beta$ -diversity

Characteristic	PERMANOVA <sup>a</sup>			PERMDISP <sup>b</sup>			PERMANOVA			DEICODE <sup>c</sup>					
	Unweighted UniFrac		Weighted UniFrac	Unweighted UniFrac		Weighted UniFrac	Unweighted UniFrac		Weighted UniFrac	Unweighted UniFrac		Weighted UniFrac			
	Test statistic	<i>P</i> value	q-value <sup>d</sup>	Test statistic	<i>P</i> value	q-value	Test statistic	<i>P</i> value	q-value	Test statistic	<i>P</i> value	q-value			
Consumed oil or butter yesterday <sup>e</sup>	1.39	0.05	0.04	0.81	0.45	0.44	0.19	0.75	0.77	0.53	0.50	0.47	3.54	<b>0.04</b>	<b>0.04</b>

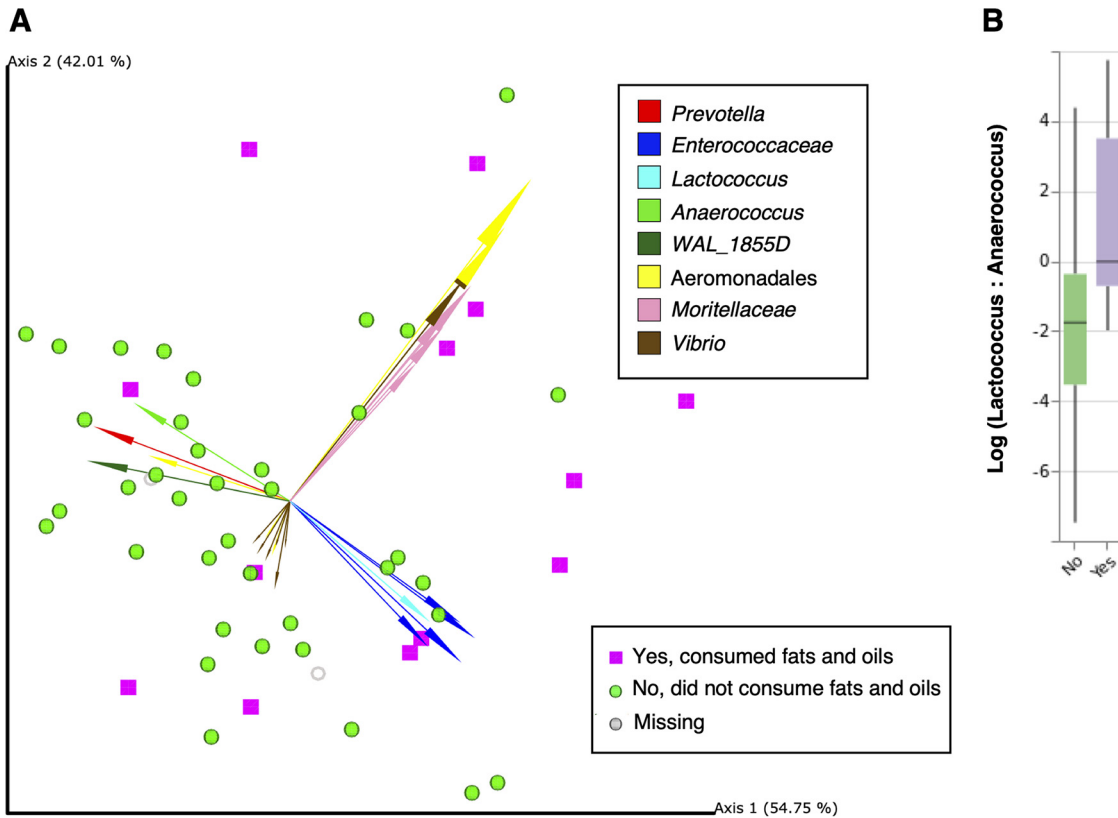
<sup>a</sup>PERMANOVA, permutational multivariate analysis of variance.

<sup>b</sup>PERMDISP, test for homogeneity of multivariate dispersions.

<sup>c</sup>DEICODE, robust Aitchison principal component analysis (RPCA) to determine which taxa strongly influence clustering. A boldface *P* value or q-value is statistically significant ( $P < 0.05$ ).

<sup>d</sup>q-values derived from pairwise testing and represent the false discovery rate (FDR) analog of a *P* value.

<sup>e</sup>Infant and Young Child Feeding (IYCF) (World Health Organization) food group M.



**FIG 3** (A) DEICODE biplot showing distances among samples from children consuming fats and oils the previous day (IYCF food group M) (purple squares) compared to samples from children not consuming fats and oils the previous day (PERMANOVA,  $P = 0.04$ ). Samples with missing data ( $n = 2$ ) are represented by gray rings. (B) *Lactococcus/Anaerococcus* log ratio between groups reporting “yes” versus “no” showing the 37 (69.81%) samples containing a valid log ratio (not containing zero). The difference in the *Lactococcus/Anaerococcus* log ratio was statistically significant ( $P = 0.01$ ).

phyla, species in the *Enterococcus* genus, *Anaerococcus* genus, *Aeromonadales* order, and *Vibrio* genus were of high relative abundance; these have been previously shown to cause infection and harbor antibiotic resistance (44).

**Dominance of Proteobacteria.** Members of the *Proteobacteria* phylum dominated the gut microbiota among infants and children in this study, contrary to our expectation that *Bifidobacteria* would be in abundance due to the report of current breastfeeding in over 80% of participants. A high abundance of *Proteobacteria* has been considered a “marker for dysbiosis” or gut microbial imbalance and associated with negative health outcomes (45). For example, preterm newborns tend to have greater *Proteobacteria* abundance in their gut microbiota compared to full-term newborns (46–48), which has been shown to be associated with necrotizing enterocolitis (NEC), a devastating and potentially fatal disease in which the underdeveloped intestinal wall is invaded by bacteria with subsequent inflammation (49). In particular, higher abundance of the *Gammaproteobacteria* class has been observed in premature infants with NEC (50); interestingly, we found that most of the bacterial sequences in the data set in urban slums of Mumbai, India, were classified as *Gammaproteobacteria*. Previous studies have also found associations between *Proteobacteria* abundance and poor health states in other age groups, such as inflammatory bowel disorders (49), irritable bowel syndrome (49), gastric bypass surgery (51), metabolic disorders (52), and intestinal inflammation (45, 53, 54), perhaps due to many *Proteobacteria* species having highly immunogenic lipopolysaccharide in the cell wall in comparison to other Gram-negative bacteria (55, 56). From these findings, *Proteobacteria* has been considered to reflect the “unstable structure of the gut microbial community” (45), and the abundance of *Proteobacteria* in this population may be a sign of an imbalance of the gut



microbiota or dysbiosis, suggesting gut microbial immaturity compared to children who are healthy.

Further, though this study did not find an association between undernutrition and *Proteobacteria*, poor nutrition status has been linked with higher *Proteobacteria* abundance and other aberrations in gut microbiota (31, 57–59). In a study of 20 children from 0 to 2 years of age living in slums in southern India, healthy children had a higher prevalence of *Bacteroidetes*, *Bifidobacterium longum*, and *Lactobacillus mucosae*, compared to stunted children, who harbored more potentially pathogenic organisms such as *Desulfovibrio* and *Campylobacterales* (60). Similarly, in a case study of two children (one healthy, one malnourished) living in an urban slum of Kolkata, India, a higher prevalence of *Campylobacterales* and *Clostridiales* was observed in the malnourished child's gut microbiome (59). Another study in a Bangladesh slum among healthy and malnourished children found a high prevalence of *Klebsiella* and *Escherichia* with a decrease in *Bacteroidetes* and other anaerobes as well as *Lactobacillus* in the malnourished children (57). Similarly, a study in a rural community of West Bengal, India, found a significant clustering of potentially pathogenic groups such as *Escherichia*, *Streptococcus*, and *Shigella* in severely malnourished children compared to healthy children 0 to 60 months of age (61). In the gut microbiomes of Malawian twins discordant for kwashiorkor, gut microbiota were causal in the development of kwashiorkor after performing mechanistic studies in mice; upon examining their gut microbiome, the mice that had developed kwashiorkor had more members of *Proteobacteria*, particularly *Bilophila wadsworthia*, which caused systemic inflammation in specific-pathogen-free mice (62) as well as *Clostridium innocuum*, a member of *Firmicutes* and associated with sudden infant death syndrome (63).

**Anthropometry.** The prevalence of poor growth was high in this study, and anthropometric measurements were found to be differentially associated with the gut microbiota, which parallels a recent review (2). In this study, head circumference was positively associated with  $\alpha$ -diversity. To our knowledge, studies examining head circumference in relation to the gut microbiome in children have only been done in neonatal and/or premature infants. One study found that administration of a synbiotic, which included members of *Firmicutes* and *Actinobacteria* as well as fructo-oligosaccharides, resulted in a lower odds of head circumference below the 10th percentile, compared to a control group after 1 year of supplementation (64). In another study, certain genera of the maternal microbiota and maternal SDI were positively correlated with neonatal male head circumference (65). In newborn preterm infants, receiving an intervention of 10 to 15 g of medically graded bee honey (a source of oligosaccharides) daily was associated with increased head circumference after 2 weeks, in addition to increased colonization with *Bifidobacterium bifidum* compared to control receiving no intervention (66). However, to our knowledge, head circumference has not been examined in concert with other nutritional metrics in relation to the gut microbiota diversity or composition in older infants and children.

Increasing weight-for-length Z-score (WLZ) was inversely associated with SDI. When stratified by sex, these associations were present only in male children. Poor growth outcomes in male children compared to female children have been previously observed (67–70). Earlier studies in undernourished infants and their microbiota have found lower microbial diversity with stunting (71), underweight and wasting (57), or severe acute malnutrition (30). Other studies have found no differences in diversity in mice which received microbiota from either severely stunted or nonstunted infants after 30 days postcolonization (72), or in Malawian infant weight-for-age Z-scores (WAZ) between 12 and 18 months of age (73). A potential explanation for our results is our small sample size, or some unmeasured environmental factor(s) may have influenced the association between WLZ and the gut microbiota in this cross-sectional analysis. However, our data suggest that both age and sex should be considered when examining growth and the gut microbiota of infants and children, particularly in lower-resource settings.

**Biomarkers.** In this study, hemoglobin concentration was associated with decreases in Faith's phylogenetic diversity (Faith's PD). A previous study of rural Kenyan infants and children found that hemoglobin concentration was not associated with microbial diversity but was positively correlated with numbers of *Escherichia coli*, a member of the *Proteobacteria* phylum (74, 75). Further, in anemic infants, a positive correlation was found between hemoglobin concentration and *Actinomycetales*, an order of *Actinobacteria* (74, 75). These findings together suggest that further examination of hemoglobin and the gut microbiota would be informative.

**Dietary intakes and feeding practices.** Consumption of fat was consistently associated with  $\alpha$ - and  $\beta$ -diversity in this study. Specifically, the log ratio between two genera of the *Firmicutes* phylum, the *Lactococcus/Anaerococcus* log ratio, was found to be significantly higher among children who consumed oils and fats compared to children who did not consume oils and fats the previous day. While these particular genera have not been specifically explored in association with dietary fat intake, a recent meta-analysis of 27 studies done with mice and humans found that high-fat diets reproducibly changed gut microbial community structure, including increased *Firmicutes* relative abundance, but had no consistent association with diversity (76). In children 1 to 6 years of age, European children with fat constituting 44 to 47% of their diet, *Bacteroidetes* and *Firmicutes* were more abundant than in African children whose diets were made up of 25 to 28% fat (77). Some studies of children and infants do suggest an association between fat intake and microbial diversity. In a study in premature infants, supplemental polyunsaturated fatty acids increased bacterial diversity (78), similar to our findings from the Faith's PD redundancy analysis. Increased fat intake from complementary foods becomes an important source of energy as breast milk consumption decreases (79), particularly in populations at higher risk of under-nourishment; our finding of higher fat intake from complementary foods may reflect greater dietary quality, translating to an increase in *Firmicutes* (80) and greater microbial diversity.

Current breastfeeding was also associated with greater  $\alpha$ -diversity. Previous studies have found lower  $\alpha$ -diversity in breastfed individuals, as breast milk selects for microbiota capable of digesting particular human milk oligosaccharides (HMO) present in breast milk (81–83), such as *Bifidobacterium*, which may suppress the expansion of other microbiota incapable of digesting HMO, resulting in lower diversity (8). In these studies, *Bifidobacterium* spp. dominated the gut microbiota as a result of breast milk consumption, in contrast to our study where *Proteobacteria* represented over 80% of the taxonomic composition despite the report of current breastfeeding in over 80% of participants. We observed higher  $\alpha$ -diversity when *Proteobacteria* abundance was relatively low (data not shown), which could allow the expansion of members from other phyla; indeed, we found an inverse association between *Actinobacteria* and Faith's PD (data not shown). Understanding associations between diet and gut microbiota in populations living in environments with higher risk for undernutrition and poor sanitation may require additional scrutiny compared to populations in higher-income countries.

**Strengths.** There are several strengths of this study. This is the first study to examine the gut microbiota among Indian children living in urban slums of Mumbai, and children from five slums were sampled and analyzed, improving generalizability of findings to children in this age group from other urban slums of Mumbai. We used rectal swabs to sample the microbiota, which has multiple benefits. They are easily obtained and convenient (84). They are stored immediately after sampling, resulting in a lower risk of contamination (84). Short-term storage at room temperature has been shown to have had no impact on composition of gut microbiota (84). They are appropriate for sampling from this age group (85–87). Skin bacterial contamination has been none or low (84). Importantly, the microbiome profile has been shown to be comparable to bulk stool samples in previous studies (48, 84, 85, 87). As another strength, we followed protocols from the Earth Microbiome Project (88, 89), such as

DNA extraction and sequencing protocols and 16S rRNA hypervariable region selected, to facilitate interstudy comparability. Further, we analyzed data using state-of-the-art bioinformatics methods (90, 91), including the fragment insertion method to acquire the taxonomic assignments of sequences, which provides advantages over *de novo* phylogenies, including accurate branch lengths, multistudy meta-analyses, and mixed region meta-analyses (36). Use of the V4 hypervariable 16S rRNA region has been shown to be especially appropriate for infant microbiota investigations, as this region tends to allow better recovery of *Bifidobacterium* (92), making our observation of very low bifidobacteria in this population even more stark.

**Limitations.** This study has some limitations. First, this pilot study sample was comprised of the first 53 participants from whom a rectal swab was able to be collected between June and July 2017 (consecutive sampling); as a result, we did not include samples from infants from the full screening period of the parent trial (93): specifically, 5 out of 20 communities sampled are represented in this cross-sectional analysis, which may present selection bias. In comparison to the rest of participants screened, we found that there were no dissimilarities in age, sex, other sociodemographic characteristics, but we did find differences in other measures, including a lower proportion of vaginal births in the microbiome subset (see Table S1 in the supplemental material). The difference in the higher proportion of vaginal births is unexpected, but it may be dependent on the subset of communities included and specific cultural practices. Another limitation in this study is the possibility of error due to the inherent drawbacks of a 24-h dietary recall questionnaire. Recall bias due to the mother's inability to accurately recall the child's dietary intake, as well as information bias due to the limitations in food composition databases to convert the reported food consumption to energy and nutrient intakes, may limit the validity of the dietary information collected (94). However, the diet in early childhood is relatively simple with lower dietary diversity than an adult's diet; in addition, we used the recently updated Indian food composition database (95), suggesting that this is a likely relatively minor weakness of the study. Another limitation is that few children consumed adequate dietary intakes from complementary foods, limiting power for statistical analysis. However, this finding parallels the results from India's 2016–2018 Comprehensive National Nutrition Survey in children under 2 years of age, suggesting increased generalizability of the cohort (96). Finally, as we conducted 16S rRNA sequencing, we were unable to include an investigation of the functional potential of the gut microbiota, given that algorithms used to predict function, such as Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) (97), have been validated only on adult populations and therefore are not appropriate for infant populations.

**Conclusions.** This sample of 10- to 18-month-old children living in urban slums of Mumbai, India, had high prevalence of poor growth and nutrient deficiencies, as well as a dominance of *Proteobacteria* in the gut. Anthropometry (head circumference, weight for age), nutritional biomarkers (hemoglobin), and diet (fat intake, iron intake, current breastfeeding) were associated with gut microbiota composition and diversity. Further longitudinal research examining comprehensive nutritional status and the gut microbiota in similar populations is warranted, given that multiple markers of nutrition—growth, biomarkers, diet, and feeding practices—were associated with the gut microbiota.

## MATERIALS AND METHODS

**Study population, setting, and design.** Participants were children between 10 and 18 completed months of age living in urban slums of Mumbai, India (including the eastern wards of Khar, Santacruz, and Bandra) who provided informed caregiver consent to be screened for enrollment into the parent study, a randomized controlled nutrition intervention trial (Clinicaltrials.gov ID: NCT02233764) (93). This exploratory cross-sectional study examined a subset of participants at screening, prior to enrollment (67).

The protocol was reviewed and approved by the Inter Systems Biomedical Ethical Committee (ISBEC) (Mumbai, Maharashtra, India), St. John's Research Institute (SJRI) Institutional Ethics Committee (IEC), and the Institutional Review Board (IRB) of Cornell University. In addition, permissions to conduct the study were obtained from the Health Ministry Screening Committee of India (Indian Council of Medical

Research). Informed consent was obtained from all caregivers in an audio/visual format per Indian Government guidelines (98).

Screening data for this cross-sectional study were collected from June to July 2017; screening data for the parent trial were collected between March and November 2017. Caregivers who had at least one 10- to 18-month-old child as identified during a census survey were invited to come to the study center, the Centre for the Study of Social Change (CSSC) (Bandra East, Mumbai, India), with their child to be screened for eligibility in the randomized trial. Inclusion criteria for enrollment into the parent trial have been described previously (93). The sample size constituted the first 53 stool samples to be collected and next-generation-sequenced and is therefore a convenient sample for this exploratory study. All children were provided 400 mg albendazole as recommended by the World Health Organization (99) during screening under supervision by the study physician.

**Anthropometry.** Trained research assistants collected anthropometric measurements using standardized procedures (100). The average of duplicate (recumbent length, mid-upper-arm circumference, head circumference) measurements was used as the final measurement. The weight of each child was measured using Rice Lake and Seca 703 body weight scales to the nearest 0.01 kg and calculated as the difference in weight of the child's caregiver alone compared to the weight of the caregiver holding the child, both wearing standard attire (without shoes) (GmbH & Co. KG, Hamburg, Germany). Child recumbent length was measured to the nearest 0.1 cm using an infant length board (ShorrBoard; Weigh and Measure LLC, Olney, MD, USA). Infant anthropometric Z-scores were computed using WHO International Growth References (version 3.2.2, 2011).

**Demographic data, dietary intakes, feeding practices, and health history.** Research assistants collected maternal and child demographic and health history data through interviews with caregivers. These variables included the age of the child, sex of the child, child's birth weight (low birthweight was defined as less than 2.5 kg [101, 102] and determined by caregiver's recall and confirmed by maternal/child health card), dietary information, and birth/delivery mode. The child's health history data were reported to the study physician by the mother, a physical examination was conducted by the study clinician, and morbidity data were recorded as a report of the child having had any occurrence of diarrhea, fever, or cough within the past month. In addition to breastfeeding status information, dietary food group consumption data from the Infant and Young Child Feeding (IYCF) questionnaire were collected (103). These IYCF dietary food groups A to O include the following: A, porridge, bread, rice, noodles, or other foods made from grains; B, pumpkin, carrots, squash, or sweet potatoes that are yellow or orange inside; C, white potatoes, white yams, manioc, cassava, or any other foods made from roots; D, any dark green leafy vegetables; E, ripe mangoes, ripe papayas; F, any other fruits or vegetables; G, liver, kidney, heart, or other organ meats; H, any meat, such as beef, pork, mutton, lamb, goat, chicken, or duck; I, eggs; J, fresh or dried fish, shellfish, or seafood; K, any foods made from beans, peas, lentils, nuts, or seeds; L, cheese, yogurt, paneer, butter, milk, or other milk products; M, any oil, fat, palmolein, butter, or foods made with any of these; N, any sugary foods such as chocolate, sweets, candies, pastries, cakes, or biscuits; and O, condiments for flavor such as chilies, spices, herbs, or fish powder. IYCF food groups B, C, D, E, and F were combined to reflect consumption of all fruits and vegetables in analysis. Dietary intakes of the child were estimated using 24-h dietary recall administered to their mother or caregiver; nutrient intakes were calculated using the updated Indian Food Composition Tables (95) via CS Dietary System software (CS Dietary System, version 1.1). These intakes represent a conservative estimate of macro- and micronutrients consumed from only complementary foods, and no nutrient contributions from breast milk. Dietary intakes were adjusted for energy using multivariate nutrient residual models which included log calories as a constant, i.e., microbiota outcome =  $b_1$ Nutrient residual +  $b_2$ Calories (104).

**Biological specimen collection.** At the study center (CSSC), a pediatric phlebotomist applied topical anesthetic (Pirilox Cream [lidocaine with prilocaine]; Neon Laboratories Limited, Mumbai, India) and performed topical antiseptics before collecting blood from the antecubital vein. After centrifugation to separate serum from whole blood, blood was divided into aliquots and immediately transported (within a range of 1 to 6 h after collection) to SRL Diagnostics (Goregaon, Mumbai, India) for immediate analysis as well as storage at  $-80^{\circ}\text{C}$  for future batch analyses of nutrition status and immune function. Complete blood counts, including hemoglobin were immediately analyzed (DXH 600 Coulter Counter) (intra-assay coefficient of variation [CV], 0.43%). Serum ferritin was measured using electrochemiluminescence (Cobas8000) (intra-assay CV, 4.5%) (limit of detection [LOD]  $<0.5$  ng/ml). Serum zinc was measured using FAAS with D2 correction (Aanalyst800) (intra-assay CV, 4.99%). C-reactive protein (CRP) was measured using nephelometry (BN II nephelometer) (intra-assay CV, 5.24%).

Stool samples were collected using Copan FecalSwab Regular Flocked Collection kit (Nylon FLO-QSwab and tube containing 2 ml Cary-Blair medium) (ThermoFisher, catalog no. R723487) by inserting the swab gently 2 to 3 cm into the rectum and rotating 360 degrees until fecal material was visible on the swab. The rectal swab was then stored at  $4^{\circ}\text{C}$  for a maximum of 48 h and subsequently stored at  $-20^{\circ}\text{C}$  for a maximum of 1 month until DNA extraction (per the manufacturer's instructions). Two separate swabs per participant were collected to ensure maximal DNA recovery.

**DNA extraction and 16S rRNA gene next-generation sequencing.** Samples were shipped on ice packs to Genotypic Technology in Bangalore, India, for DNA extraction (modified from the MoBio DNeasy PowerSoil HTP 96 kit Instruction Manual, per instructions for DNA extraction proposed by the Earth Microbiome Project 2018 instructions [105, 106]) and sequencing. The concentration and purity of samples were estimated using the Nanodrop spectrophotometer and Qubit fluorometer. Genomic DNA (25 ng) was amplified for 26 cycles using KAPA HiFi HotStart PCR kit (Kapa Biosystems Inc., Boston, MA, USA). The V3-V4 region was targeted for library construction using 341F/806RB primers modified as

described previously (107–110). The forward PCR oligonucleotide (50 bp) contained a 5' Illumina sequencing adapter, a 10-nucleotide (nt) pad sequence, and the 341 16S specific linker-primer sequence (5'-CCTACGGGNGGCWGCAG-3'). The reverse PCR oligonucleotide (55 bp) contained the 3' reverse complement of an Illumina sequencing adapter, the 12-nt Golay barcode, a 10-nt pad sequence, and the 16S specific 806R (modified) reverse linker-primer sequence (5'-CCGGACTACNVGGGTWTCTAAT-3').

The V4 region was chosen to correspond with the reverse primer of the Earth Microbiome Project (EMP), and the V3 region was included in the single fragment to facilitate better alignment during the subsequent analysis. The forward and reverse primer concentrations (0.2  $\mu$ M each) were analyzed on a 1.2% agarose gel. Round 1 PCR amplicons (1  $\mu$ L, 1:10) were used for round 2 indexing PCR by amplifying round 1 PCR amplicons for 10 cycles to add Nextera adapters (Nextera XT v2 index kit; Illumina, USA). Round 2 PCR amplicons were analyzed on 1.2% agarose gel. Amplicons were sequenced on Illumina MiSeq, using 275 bp x 2 paired-end sequences by Genotypic Laboratory (Bangalore, India).

**Bioinformatics processing.** Demultiplexing of paired-end reads was performed using `bcl2fastq v1.8.4` (111). Further processing was performed using the open-source bioinformatics pipeline, QIIME 2 version 2019.7, installed in a conda environment in Linux CentOS (90). Sequence primers were quality filtered and trimmed using a length of 100 bp via the `Deblur` workflow, using a minimum read number of 1 and trim length of 100 bp (36, 112, 113). The resulting quality-filtered feature table of sequence variants (equivalent to 100% operational taxonomic units [OTUs]) was visually summarized using the `qiime feature-table summarize` command to generate descriptive statistics. For phylogenetic diversity analysis, we used the fragment insertion method (114–117) using the Greengenes 13.8 reference database at 99% (command `qiime fragment-insertion sepp`) as detailed earlier (118). Chloroplast and mitochondrial sequences were filtered out of the resulting table using the `qiime taxa filter-table` command.

The `qiime feature-table rarefy` command was used to rarefy the data (by random subsampling) to a sampling depth of 38,012, the deepest sampling depth possible that included all 53 samples (119). The `qiime diversity alpha` and `qiime diversity beta-phylogenetic` plug-ins were used to compute  $\alpha$ -diversity and  $\beta$ -diversity metrics on the rarefied sequence variant table. Measures of  $\alpha$ -diversity analyzed included the Shannon diversity index (SDI) (38), and Faith's phylogenetic diversity (Faith's PD) (39).  $\beta$ -Diversity was measured using unweighted and weighted UniFrac to consider the relative abundance of taxa in addition to presence or absence information (120, 121) as well as DEICODE (42).

**Statistical analysis.** We first analyzed continuous variables for normality using the Shapiro-Wilk test. If data were not normally distributed, median and interquartile range (IQR) values were reported. Continuous data were assessed for correlational relationships using Spearman correlation, and medians (IQRs) were compared between groups using the Hodges-Lehmann-Sen test. Categorical data were compared between groups using the chi square test; Fisher's exact test was used in analyses where at least 25% of expected counts were less than five.

Linear regression was performed to examine the association between exposures of poor nutrition status and gut microbiota outcomes, percent relative abundance, and  $\alpha$ -diversity metrics, SDI and Faith's PD. To identify potentially confounding factors, nutritional exposure variables associated with the outcome of interest at  $P < 0.20$  in univariate analysis were included in the multivariate model; only those correlates were retained in the model that were associated with the outcome with a  $P$  value of  $\leq 0.05$  (122). All analyses were adjusted for age and sex.

We also identified nonredundant covariates using a forward stepwise redundancy analysis (RDA) with the `vegan` package in R. This analysis estimates the linear cumulative and independent effect size of each nonredundant covariate on microbiome diversity variation (123). For RDA analysis, after removing colinear variables, we included dietary, sociodemographic, and clinical correlates in the model.

$\beta$ -Diversity (unweighted and weighted UniFrac metrics) was visualized by principal coordinate analysis (PCoA) using the Emperor software through QIIME 2, using abundance data to determine clustering patterns between the participants (124). To determine whether UniFrac distances clustered among participants with particular characteristics, we used the `qiime diversity beta-group-significance` command to run a permutational multivariate analysis of variance (PERMANOVA) (40, 125) test to determine whether (weighted or unweighted) UniFrac distances between participants within a group were more similar to each other than they were to participants representing the other (reference) group. To test for homogeneity of multivariate dispersions, `PERMDISP` (41) was run to compare within-group spread among groups using the average distance between individual observations to their group centroid to identify the relative spreads of data cloud shapes among groups (126).

To address sparsity in our data, we also examined which taxa strongly influenced clustering using default parameters of the robust Aitchison principal component analysis (RPCA) DEICODE (version 0.2.4) QIIME 2 plug-in (`qiime deicode rpca`) (42). DEICODE is robust to high levels of sparsity, such that zero values do not influence the resulting ordination. Any taxon identified to the species level was considered misidentified, as 16S rRNA sequencing is susceptible to species-level misidentifications. The resulting compositional biplots (127) were visualized in QIIME 2 using Emperor to assess the 25 features with the highest magnitudes, i.e., those expected to be important in causing separation in the data set (<https://forum.qiime2.org/t/how-to-make-pcoa-biplot-in-r-using-q2-deicode-ordination/8377/6>) (`--p-number-of-features, 25`). We performed PERMANOVA on the DEICODE results.

The feature loadings in a compositional biplot produced by DEICODE output were then visualized in the `Qurro` version 0.4.0 QIIME 2 plug-in (command `qiime qurro loading-plot`) to display a plot of feature loadings for a given axis in the biplot alongside a plot of the log ratios of selected features' abundances within samples (43). The rank plot field was unadjusted, and therefore, the feature loadings

from axis 1 of the biplot were assessed. The log ratios of taxa at the extremes of the Qurro rank plot were selected to compare (using Student's *t* test assuming unequal variances and no multiple comparisons correction applied) between group characteristics that were statistically significantly different from the DEICODE results.

Exposures of interest included anthropometric indices, including birth weight, current weight, mid-upper arm circumference, length, head circumference, weight-for-age Z-score (WAZ), length-for-age Z-score (LAZ), and weight-for-length Z-score (WHZ) (128); blood biomarkers, including ferritin, zinc, and C-reactive protein, hemoglobin concentration; report of diarrhea, cough, or fever within the past 4 weeks; and dietary and feeding practices, including dietary intakes of macro- and micronutrients, current breastfeeding, IYCF indicators (103); as well as data on birth mode (vaginal or Caesarean). All analyses were adjusted for age and sex where appropriate.

Selection bias was examined by comparing characteristics (such as age, sex, and other clinical and dietary characteristics) of the sample in this cross-sectional study with the rest of the screened participants. All exposure/outcome combinations were tested and are reported in supplemental tables; we reported those of interest and those that were statistically significant.

After statistical tests, we utilized false discovery rates (FDR) per the Benjamini and Hochberg approach (37), as *post hoc* multiple testing corrections. All analyses were two sided, and differences between groups were considered significant at  $P < 0.05$ . Data were analyzed using SAS version 9.4 (SAS Institute, Cary, NC, USA), R Studio (R Foundation), and QIIME 2 version 2019.7.

**Data availability.** These data were subject to cross-checking and confirmation by the Cornell Institute for Social and Economic Research (CISER) (member of DataCite, <https://doi.org/10.5281/zenodo.556235>) to ensure reproducibility; data sets and code are available upon request at <https://doi.org/10.6077/zrv-cpc31>. The data that support the findings of this study are openly available in NCBI BioProject at <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA657036>. The DNA sequences corresponding to the 16S rRNA gene data in this study have been submitted as raw fastq files to the SRA at <https://www.ncbi.nlm.nih.gov/sra/PRJNA657036>. Various QIIME 2 files, including Qurro plots, are available at <https://github.com/knightlab-analyses/nutrition-gut-microbiota-mumbai>.

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**FIG S1**, TIF file, 2.3 MB.

**FIG S2**, TIFF file, 0.5 MB.

**TABLE S1**, DOCX file, 0.02 MB.

**TABLE S2**, DOCX file, 0.02 MB.

**TABLE S3**, DOCX file, 0.02 MB.

**TABLE S4**, DOCX file, 0.01 MB.

**TABLE S5**, DOCX file, 0.02 MB.

**TABLE S6**, DOCX file, 0.02 MB.

## ACKNOWLEDGMENTS

Financial support was awarded by HarvestPlus, grant number 2014H8302, awarded to Cornell University. The publication costs of this manuscript were supported by the Bill and Melinda Gates Foundation through their open access fund. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

We gratefully acknowledge the children and their mothers for participating in this study. We also thank our study coordinators, Laura Powis and Yadurshini Raveendran; nutrition coordinator, Harsha C. Chopra; local community health workers and project clerks, Ashwini Pendurkar, Nazma Sheikh, Qamrunisa Shah, Yasmin Shaikh, Tasleem Shaikh, Sheetal Surve, Bharti Kumbhar, Regina Fernandes, Steffie Fernandes, Suvarna Santosh Parab, Trupti Mohite, Deepa Tandel, Neelam Jadhav, Bashirunnisa Shaikh, Afroz Farooqie, Kiran Mestry, Shanta Pundalik Hiddujikar, Mayuri Ghag, Neelam Jalankar, Subhada Hardas, Prachi Mathkar, Yojina Dhanawade, Mangal Khade, Suvarna Ganekar, Sangeetha Ghadi, Ankita Devlekar, Bhagyasri Babrekar, Aditi Tamble, Lakshmi Martal, Aswhini Warang, Satyabhama Matondkar, Aswini Warang, Dipika Dipak Maingade, Varsha Ambre, Namrata Narkar, Suneetha Kadam, Jyothi Lokhande, Ankita Pendurkar, Shivani Jaiswal, Shraddha Tivrekar, Ragini Arde, and Seetamaya Saw; research assistants, Shobaha R. Iyer, Janhavi Sonawane, Tejashree Thorat, Pratiksha Bagal, Pranali Pangerkar, Swati J. Kathar, Priyanka Kadam, Rinki Saw, Chinki Saw, Lalita Saw, Samrutthi Surve, Sona Majethiya, Sonu Mishra, Bansari Rao, Anupriya Pawar, Namrata Shinde, Monthie Fernandes, and Ashwini Salvi; local Cornell study personnel, Sudha Venkatramanan; study physicians, Sirazul Ameen Sahariah and Sanjay Ojha from the Centre for the Study of Social Change in Mumbai, India, who

facilitated the collection of these data. We are also grateful to Harpreet Singh and the team at Genotypic Technology Pvt Ltd., Bangalore, India, for sample processing and sequencing of the gut microbiota.

S.M. conceived the project. S.L.H. developed the protocol, performed bioinformatic and statistical analysis, and drafted the manuscript. L.J. performed bioinformatic and statistical analysis. L.J., D.M., C.M., M.F., F.A., and R.K. assisted with interpretation of the microbiota data. S.A.U., A.T., V.T., P.G., and K.R. assisted with dietary data acquisition and analysis. R.D.P. and H.C. assisted with all data acquisition. S.M., S.A.U., D.G.R., J.D.H., and J.L.F. assisted with data analysis and interpretation. All authors substantively revised the work. S.M. and R.K. jointly supervised the project.

S.M. is an unpaid board member for and has an equity stake in a diagnostic start-up focused on developing point-of-care assays for nutritional status informed by his research as a faculty member at Cornell University. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## REFERENCES

- Tamburini S, Shen N, Wu HC, Clemente JC. 2016. The microbiome in early life: implications for health outcomes. *Nat Med* 22:713–722. <https://doi.org/10.1038/nm.4142>.
- Million M, Diallo A, Raoult D. 2017. Gut microbiota and malnutrition. *Microb Pathog* 106:127–138. <https://doi.org/10.1016/j.micpath.2016.02.003>.
- Donovan SM. 2017. Introduction to the special focus issue on the impact of diet on gut microbiota composition and function and future opportunities for nutritional modulation of the gut microbiome to improve human health. *Gut Microbes* 8:75–81. <https://doi.org/10.1080/19490976.2017.1299309>.
- Holscher HD. 2017. Dietary fiber and prebiotics and the gastrointestinal microbiota. *Gut Microbes* 8:172–184. <https://doi.org/10.1080/19490976.2017.1290756>.
- Arpaia N, Campbell C, Fan X, Dikiy S, van der Veeken J, deRoos P, Liu H, Cross JR, Pfeffer K, Coffey PJ, Rudenski AY. 2013. Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* 504:451–455. <https://doi.org/10.1038/nature12726>.
- Smith PM, Howitt MR, Panikov N, Michaud M, Gallini CA, Bohlooly YM, Glickman JN, Garrett WS. 2013. The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. *Science* 341:569–573. <https://doi.org/10.1126/science.1241165>.
- Chu DM, Meyer KM, Prince AL, Aagaard KM. 2016. Impact of maternal nutrition in pregnancy and lactation on offspring gut microbial composition and function. *Gut Microbes* 7:459–470. <https://doi.org/10.1080/19490976.2016.1241357>.
- Laursen MF, Bahl MI, Michaelsen KF, Licht TR. 2017. First foods and gut microbes. *Front Microbiol* 8:356. <https://doi.org/10.3389/fmicb.2017.00356>.
- Penders J, Thijs C, Vink C, Stelma FF, Snijders B, Kummeling I, van den Brandt PA, Stobberingh EE. 2006. Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics* 118:511–521. <https://doi.org/10.1542/peds.2005-2824>.
- United Nations Children's Fund (UNICEF), World Health Organization, International Bank for Reconstruction and Development/The World Bank. 2019. Levels and trends in child malnutrition: key findings of the 2019 Edition of the Joint Child Malnutrition Estimate. World Health Organization, Geneva, Switzerland.
- Clemente JC, Manasson J, Scher JU. 2018. The role of the gut microbiome in systemic inflammatory disease. *BMJ* 360:j5145. <https://doi.org/10.1136/bmj.j5145>.
- Round JL, Palm NW. 2018. Causal effects of the microbiota on immune-mediated diseases. *Sci Immunol* 3:eaa01603. <https://doi.org/10.1126/sciimmunol.aao1603>.
- Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, Angenent LT, Ley RE. 2011. Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A* 108(Suppl 1):4578–4585. <https://doi.org/10.1073/pnas.1000081107>.
- Thompson AL, Monteagudo-Mera A, Cadenas MB, Lampl ML, Azcarate-Peril MA. 2015. Milk- and solid-feeding practices and daycare attendance are associated with differences in bacterial diversity, predominant communities, and metabolic and immune function of the infant gut microbiome. *Front Cell Infect Microbiol* 5:3. <https://doi.org/10.3389/fcimb.2015.00003>.
- Sordillo J, Zhou Y, McGeachie M, Ziniti J, Lange N, Laranjo N, Savage J, Carey V, O'Connor G, Sandel M, Strunk R, Bacharier L, Zeiger R, Weiss S, Weinstock G, Gold D, Litonjua A. 2017. Factors influencing the infant gut microbiome at age 3–6 months: findings from the ethnically diverse Vitamin D Antenatal Asthma Reduction Trial (VDAART). *J Allergy Clin Immunol* 139:482–491.e14. <https://doi.org/10.1016/j.jaci.2016.08.045>.
- Madan JC, Hoen AG, Lundgren SN, Farzan SF, Cottingham KL, Morrison HG, Sogin ML, Li H, Moore JH, Karagas MR. 2016. Association of Cesarean delivery and formula supplementation with the intestinal microbiome of 6-week-old infants. *JAMA Pediatr* 170:212–219. <https://doi.org/10.1001/jamapediatrics.2015.3732>.
- Costello EK, Carlisle EM, Bik EM, Morowitz MJ, Relman DA. 2013. Microbiome assembly across multiple body sites in low-birthweight infants. *mBio* 4:e00782-13. <https://doi.org/10.1128/mBio.00782-13>.
- Zhou Y, Shan G, Sodergren E, Weinstock G, Walker WA, Gregory KE. 2015. Longitudinal analysis of the premature infant intestinal microbiome prior to necrotizing enterocolitis: a case-control study. *PLoS One* 10:e0118632. <https://doi.org/10.1371/journal.pone.0118632>.
- Butcher J, Unger S, Li J, Bando N, Romain G, Francis J, Mottawea W, Mack D, Stintzi A, O'Connor DL. 2018. Independent of birth mode or gestational age, very-low-birth-weight infants fed their mothers' milk rapidly develop personalized microbiotas low in Bifidobacterium. *J Nutr* 148:326–335. <https://doi.org/10.1093/jn/nxx071>.
- Tang M, Frank DN, Sherlock L, Ir D, Robertson CE, Krebs NF. 2016. Effect of vitamin E with therapeutic iron supplementation on iron repletion and gut microbiome in U.S. iron deficient infants and toddlers. *J Pediatr Gastroenterol Nutr* 63:379–385. <https://doi.org/10.1097/MPG.0000000000001154>.
- Krebs NF, Sherlock LG, Westcott J, Culbertson D, Hambidge KM, Feazel LM, Robertson CE, Frank DN. 2013. Effects of different complementary feeding regimens on iron status and enteric microbiota in breastfed infants. *J Pediatr* 163:416–423. <https://doi.org/10.1016/j.jpeds.2013.01.024>.
- Azad MB, Konya T, Maughan H, Guttman DS, Field CJ, Chari RS, Sears MR, Becker AB, Scott JA, Kozyrskyj AL, CHILD Study Investigators. 2013. Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. *CMAJ* 185:385–394. <https://doi.org/10.1503/cmaj.121189>.
- Koleva PT, Tun HM, Konya T, Guttman DS, Becker AB, Mandhane PJ, Turvey SE, Subbarao P, Sears MR, Scott JA, Kozyrskyj AL, CHILD Study Investigators. 2017. Sex-specific impact of asthma during pregnancy on infant gut microbiota. *Eur Respir J* 50:1700280. <https://doi.org/10.1183/13993003.00280-2017>.
- Murphy K, Curley D, O'Callaghan TF, O'Shea CA, Dempsey EM, O'Toole PW, Ross RP, Ryan CA, Stanton C. 2017. The composition of human milk and infant faecal microbiota over the first three months of life: a pilot study. *Sci Rep* 7:40597. <https://doi.org/10.1038/srep40597>.

25. Lee MJ, Kang MJ, Lee SY, Lee E, Kim K, Won S, Suh DJ, Kim KW, Sheen YH, Ahn K, Kim BS, Hong SJ. 2018. Perturbations of gut microbiome genes in infants with atopic dermatitis according to feeding type. *J Allergy Clin Immunol* 141:1310–1319. <https://doi.org/10.1016/j.jaci.2017.11.045>.
26. Kostic AD, Gevers D, Siljander H, Vatanen T, Hyötyläinen T, Hamalainen AM, Peet A, Tillmann V, Poho P, Mattila I, Lahdesmaki H, Franzosa EA, Vaarala O, de Goffau M, Harmsen H, Ilonen J, Virtanen SM, Clish CB, Oresic M, Huttenhower C, Knip M, DIABIMMUNE Study Group, Xavier RJ. 2015. The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe* 17:260–273. <https://doi.org/10.1016/j.chom.2015.01.001>.
27. Wampach L, Heintz-Buschart A, Hogan A, Muller EEL, Narayanasamy S, Laczny CC, Hugerth LW, Bindl L, Bottu J, Andersson AF, de Beaufort C, Wilmes P. 2017. Colonization and succession within the human gut microbiome by archaea, bacteria, and microeukaryotes during the first year of life. *Front Microbiol* 8:738. <https://doi.org/10.3389/fmicb.2017.00738>.
28. Talsness CE, Penders J, Jansen E, Damoiseaux J, Thijs C, Mommers M. 2017. Influence of vitamin D on key bacterial taxa in infant microbiota in the KOALA Birth Cohort Study. *PLoS One* 12:e0188011. <https://doi.org/10.1371/journal.pone.0188011>.
29. Dogra S, Sakwinska O, Soh SE, Ngom-Bru C, Bruck WM, Berger B, Brussow H, Lee YS, Yap F, Chong YS, Godfrey KM, Holbrook JD, GUSTO Study Group. 2015. Dynamics of infant gut microbiota are influenced by delivery mode and gestational duration and are associated with subsequent adiposity. *mBio* 6:e02419-14. <https://doi.org/10.1128/mBio.02419-14>.
30. Subramanian S, Huq S, Yatsunenkov T, Haque R, Mahfuz M, Alam MA, Benezra A, DeStefano J, Meier MF, Muegge BD, Barratt MJ, VanArendonk LG, Zhang Q, Province MA, Petri WA, Jr, Ahmed T, Gordon JL. 2014. Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature* 510:417–421. <https://doi.org/10.1038/nature13421>.
31. Smith MI, Yatsunenkov T, Manary MJ, Trehan I, Mkakosya R, Cheng J, Kau AL, Rich SS, Concannon P, Mychaleckyj JC, Liu J, Houpt E, Li JV, Holmes E, Nicholson J, Knights D, Ursell LK, Knight R, Gordon JL. 2013. Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. *Science* 339:548–554. <https://doi.org/10.1126/science.1229000>.
32. WHO. 2009. WHO child growth standards and the identification of severe acute malnutrition in infants and children. WHO, UNICEF, Geneva, Switzerland.
33. UN-HABITAT. 2003. The challenge of slums: global report on human settlements. Earthscan Publications Ltd, London, United Kingdom.
34. Nolan LB. 2015. Slum definitions in urban India: implications for the measurement of health inequalities. *Popul Dev Rev* 41:59–84. <https://doi.org/10.1111/j.1728-4457.2015.00026.x>.
35. Chandramouli C. 2011. Housing stock, amenities & assets in slums – Census 2011 (presentation). Census of India. Ministry of Home Affairs, Government of India, India. [censusindia.gov.in](http://censusindia.gov.in).
36. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, Kightley EP, Thompson LR, Hyde ER, Gonzalez A, Knight R. 2017. Deblur rapidly resolves single-nucleotide community sequence patterns. *mSystems* 2:e00191-16. <https://doi.org/10.1128/mSystems.00191-16>.
37. Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B (Methodological)* 57:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
38. Shannon CE. 1948. A mathematical theory of communication. *Bell Syst Tech J* 27:379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb00917.x>.
39. Faith D. 1992. Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61:1–10. [https://doi.org/10.1016/0006-3207\(92\)91201-3](https://doi.org/10.1016/0006-3207(92)91201-3).
40. Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. *Aust Ecol* 26:32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>.
41. Anderson MJ. 2006. Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* 62:245–253. <https://doi.org/10.1111/j.1541-0420.2005.00440.x>.
42. Martino C, Morton JT, Marotz CA, Thompson LR, Tripathi A, Knight R, Zengler K. 2019. A novel sparse compositional technique reveals microbial perturbations. *mSystems* 4:e00016-19. <https://doi.org/10.1128/mSystems.00016-19>.
43. Fedarko MW, Martino C, Morton JT, González A, Rahman G, Marotz CA, Minich JJ, Allen EE, Knight R. 2020. Visualizing ‘omic feature rankings and log-ratios using Qurro. *NAR Genom Bioinform* 2:lqaa023. <https://doi.org/10.1093/nargab/lqaa023>.
44. Hoque SS, Faruque AS, Mahalanabis D, Hasnat A. 1994. Infectious agents causing acute watery diarrhoea in infants and young children in Bangladesh and their public health implications. *J Trop Pediatr* 40: 351–354.
45. Shin NR, Whon TW, Bae JW. 2015. Proteobacteria: microbial signature of dysbiosis in gut microbiota. *Trends Biotechnol* 33:496–503. <https://doi.org/10.1016/j.tibtech.2015.06.011>.
46. Hill CJ, Lynch DB, Murphy K, Ulaszewska M, Jeffery IB, O’Shea CA, Watkins C, Dempsey E, Mattivi F, Tuohy K, Ross RP, Ryan CA, O’Toole PW, Stanton C. 2017. Erratum to: Evolution of gut microbiota composition from birth to 24 weeks in the INFANTMET Cohort. *Microbiome* 5:21. <https://doi.org/10.1186/s40168-017-0240-3>.
47. Hill CJ, Lynch DB, Murphy K, Ulaszewska M, Jeffery IB, O’Shea CA, Watkins C, Dempsey E, Mattivi F, Tuohy K, Ross RP, Ryan CA, O’Toole PW, Stanton C. 2017. Evolution of gut microbiota composition from birth to 24 weeks in the INFANTMET Cohort. *Microbiome* 5:4. <https://doi.org/10.1186/s40168-016-0213-y>.
48. Grier A, Qiu X, Bandyopadhyay S, Holden-Wiltse J, Kessler HA, Gill AL, Hamilton B, Huyc H, Misra S, Mariani TJ, Ryan RM, Scholer L, Scheible KM, Lee YH, Caserta MT, Pryhuber GS, Gill SR. 2017. Impact of prematurity and nutrition on the developing gut microbiome and preterm infant growth. *Microbiome* 5:158. <https://doi.org/10.1186/s40168-017-0377-0>.
49. Li M, Wang M, Donovan SM. 2014. Early development of the gut microbiome and immune-mediated childhood disorders. *Semin Reprod Med* 32:74–86. <https://doi.org/10.1055/s-0033-1361825>.
50. Lindberg TP, Caimano MJ, Hagadorn JI, Bennett EM, Maas K, Brownell EA, Matson AP. 2020. Preterm infant gut microbial patterns related to the development of necrotizing enterocolitis. *J Matern Fetal Neonatal Med* 33:349–358. <https://doi.org/10.1080/14767058.2018.1490719>.
51. Liou AP, Paziuk M, Luevano JM, Jr, Machineni S, Turnbaugh PJ, Kaplan LM. 2013. Conserved shifts in the gut microbiota due to gastric bypass reduce host weight and adiposity. *Sci Transl Med* 5:178ra41. <https://doi.org/10.1126/scitranslmed.3005687>.
52. Fei N, Zhao L. 2013. An opportunistic pathogen isolated from the gut of an obese human causes obesity in germfree mice. *ISME J* 7:880–884. <https://doi.org/10.1038/ismej.2012.153>.
53. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, Reyes JA, Shah SA, LeLeiko N, Snapper SB, Bousvaros A, Korzenik J, Sands BE, Xavier RJ, Huttenhower C. 2012. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* 13:R79. <https://doi.org/10.1186/gb-2012-13-9-r79>.
54. Mirpuri J, Raetz M, Sturge CR, Wilhelm CL, Benson A, Savani RC, Hooper LV, Yarovsky F. 2014. Proteobacteria-specific IgA regulates maturation of the intestinal microbiota. *Gut Microbes* 5:28–39. <https://doi.org/10.4161/gmic.26489>.
55. Vatanen T, Kostic AD, d’Hennezel E, Siljander H, Franzosa EA, Yassour M, Kolde R, Vlamakis H, Arthur TD, Hamalainen AM, Peet A, Tillmann V, Uibo R, Mokurov S, Dorshakova N, Ilonen J, Virtanen SM, Szabo SJ, Porter JA, Lahdesmaki H, Huttenhower C, Gevers D, Cullen TW, Knip M, DIABIMMUNE Study Group, Xavier RJ. 2016. Variation in microbiome LPS immunogenicity contributes to autoimmunity in humans. *Cell* 165:842–853. <https://doi.org/10.1016/j.cell.2016.04.007>.
56. Korpela K. 2018. Diet, microbiota, and metabolic health: trade-off between saccharolytic and proteolytic fermentation. *Annu Rev Food Sci Technol* 9:65–84. <https://doi.org/10.1146/annurev-food-030117-012830>.
57. Monira S, Nakamura S, Gotoh K, Izutsu K, Watanabe H, Alam NH, Endtz HP, Cravioto A, Ali SI, Nakaya T, Horii T, Iida T, Alam M. 2011. Gut microbiota of healthy and malnourished children in Bangladesh. *Front Microbiol* 2:228. <https://doi.org/10.3389/fmicb.2011.00228>.
58. Mata LJ, Jimenez F, Cordon M, Rosales R, Prera E, Schneider RE, Viteri F. 1972. Gastrointestinal flora of children with protein-calorie malnutrition. *Am J Clin Nutr* 25:118–126. <https://doi.org/10.1093/ajcn/25.10.1118>.
59. Gupta SS, Mohammed MH, Ghosh TS, Kanungo S, Nair GB, Mande SS. 2011. Metagenome of the gut of a malnourished child. *Gut Pathog* 3:7. <https://doi.org/10.1186/1757-4749-3-7>.
60. Dinh DM, Ramadass B, Kattula D, Sarkar R, Braunstein P, Tai A, Wanke CA, Hassoun S, Kane AV, Naumova EN, Kang G, Ward HD. 2016. Longitudinal analysis of the intestinal microbiota in persistently stunted young children in south India. *PLoS One* 11:e0155405. <https://doi.org/10.1371/journal.pone.0155405>.
61. Ghosh TS, Gupta SS, Bhattacharya T, Yadav D, Barik A, Chowdhury A, Das B, Mande SS, Nair GB. 2014. Gut microbiomes of Indian children of



- varying nutritional status. *PLoS One* 9:e95547. <https://doi.org/10.1371/journal.pone.0095547>.
62. Feng Z, Long W, Hao B, Ding D, Ma X, Zhao L, Pang X. 2017. A human stool-derived *Bifidobacterium wadsworthii* strain caused systemic inflammation in specific-pathogen-free mice. *Gut Pathog* 9:59. <https://doi.org/10.1186/s13099-017-0208-7>.
  63. Highet AR, Berry AM, Bettelheim KA, Goldwater PN. 2014. Gut microbiome in sudden infant death syndrome (SIDS) differs from that in healthy comparison babies and offers an explanation for the risk factor of prone position. *Int J Med Microbiol* 304:735–741. <https://doi.org/10.1016/j.ijmm.2014.05.007>.
  64. Guney Varal I, Koksak N, Ozkan H, Bagci O, Dogan P. 2018. Potential use of multi-strain synbiotics for improving postnatal head circumference. *Pak J Med Sci* 34:1502–1506. <https://doi.org/10.12669/pjms.346.16107>.
  65. Sato Y, Sakurai K, Tanabe H, Kato T, Nakanishi Y, Ohno H, Mori C. 2019. Maternal gut microbiota is associated with newborn anthropometrics in a sex-specific manner. *J Dev Orig Health Dis* 10:659–666. <https://doi.org/10.1017/S2040174419000138>.
  66. Aly H, Said RN, Wali IE, Elwakkad A, Soliman Y, Awad AR, Shawky MA, Alam MSA, Mohamed MA. 2017. Medically graded honey supplementation formula to preterm infants as a prebiotic: a randomized controlled trial. *J Pediatr Gastroenterol Nutr* 64:966–970. <https://doi.org/10.1097/MPG.0000000000001597>.
  67. Huey SL, Finkelstein JL, Venkatraman S, Udipi SA, Ghugre P, Thakker V, Thorat A, Potdar RD, Chopra HV, Kurpad AV, Haas JD, Mehta S. 2019. Prevalence and correlates of undernutrition in young children living in urban slums of Mumbai, India: a cross sectional study. *Front Public Health* 7:191. <https://doi.org/10.3389/fpubh.2019.00191>.
  68. Bork KA, Diallo A. 2017. Boys are more stunted than girls from early infancy to 3 years of age in rural Senegal. *J Nutr* 147:940–947. <https://doi.org/10.3945/jn.116.243246>.
  69. Wamani H, Astrom AN, Peterson S, Tumwine JK, Tylleskar T. 2007. Boys are more stunted than girls in sub-Saharan Africa: a meta-analysis of 16 demographic and health surveys. *BMC Pediatr* 7:17. <https://doi.org/10.1186/1471-2431-7-17>.
  70. Mehta S, Mugusi FM, Bosch RJ, Chatterjee A, Finkelstein JL, Fataki M, Kisenge R, Fawzi WW. 2011. Author's response to reviews: Title: A randomized trial of multivitamin supplementation in children with tuberculosis in Tanzania, version 2.
  71. Gough EK, Stephens DA, Moodie EE, Prendergast AJ, Stoltzfus RJ, Humphrey JH, Manges AR. 2015. Linear growth faltering in infants is associated with *Acidaminococcus* sp. and community-level changes in the gut microbiota. *Microbiome* 3:24. <https://doi.org/10.1186/s40168-015-0089-2>.
  72. Blanton LV, Charbonneau MR, Salih T, Barratt MJ, Venkatesh S, Ilkaveya O, Subramanian S, Manary MJ, Trehan I, Jorgensen JM, Fan YM, Henrissat B, Leyn SA, Rodionov DA, Osterman AL, Maleta KM, Newgard CB, Ashorn P, Dewey KG, Gordon JI. 2016. Gut bacteria that prevent growth impairments transmitted by microbiota from malnourished children. *Science* 351:aad3311. <https://doi.org/10.1126/science.aad3311>.
  73. Kamng'ona AW, Young R, Arnold CD, Kortekangas E, Patson N, Jorgensen JM, Prado EL, Chaima D, Malamba C, Ashorn U, Fan Y-M, Cheung YB, Ashorn P, Maleta K, Dewey KG. 2019. The association of gut microbiota characteristics in Malawian infants with growth and inflammation. *Sci Rep* 9:12893. <https://doi.org/10.1038/s41598-019-49274-y>.
  74. Paganini D, Jaeggi T, Cercamondi C, Kujinga P, Moretti D, Zimmermann M. 2016. Anemia and iron status are predictors of gut microbiome composition and metabolites in infants and children in rural Kenya. *FASEB J* 30:296.2.
  75. Jaeggi T, Kortman GA, Moretti D, Chassard C, Holding P, Dostal A, Boekhorst J, Timmerman HM, Swinkels DW, Tjalsma H, Njenga J, Mwangi A, Kvalsvig J, Lacroix C, Zimmermann MB. 2015. Iron fortification adversely affects the gut microbiome, increases pathogen abundance and induces intestinal inflammation in Kenyan infants. *Gut* 64:731–742. <https://doi.org/10.1136/gutjnl-2014-307720>.
  76. Bisanz JE, Upadhyay V, Turnbaugh JA, Ly K, Turnbaugh PJ. 2019. Meta-analysis reveals reproducible gut microbiome alterations in response to a high-fat diet. *Cell Host Microbe* 26:265–272.e4. <https://doi.org/10.1016/j.chom.2019.06.013>.
  77. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, Collini S, Pieraccini G, Lionetti P. 2010. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc Natl Acad Sci U S A* 107:14691–14696. <https://doi.org/10.1073/pnas.1005963107>.
  78. Kankaanpaa PE, Salminen SJ, Isolauri E, Lee YK. 2001. The influence of polyunsaturated fatty acids on probiotic growth and adhesion. *FEMS Microbiol Lett* 194:149–153. <https://doi.org/10.1111/j.1574-6968.2001.tb09460.x>.
  79. Abeshu MA, Lelisa A, Geleta B. 2016. Complementary feeding: review of recommendations, feeding practices, and adequacy of homemade complementary food preparations in developing countries – lessons from Ethiopia. *Front Nutr* 3:41. <https://doi.org/10.3389/fnut.2016.00041>.
  80. Bokulich NA, Chung J, Battaglia T, Henderson N, Jay M, Li H, Lieber AD, Wu F, Perez-Perez GI, Chen Y, Schweizer W, Zheng X, Contreras M, Dominguez-Bello MG, Blaser MJ. 2016. Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Sci Transl Med* 8:343ra82. <https://doi.org/10.1126/scitranslmed.aad7121>.
  81. Yatsunenkov T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI. 2012. Human gut microbiome viewed across age and geography. *Nature* 486:222–227. <https://doi.org/10.1038/nature11053>.
  82. O'Sullivan A, He X, McNiven EM, Haggarty NW, Lonnerdal B, Slupsky CM. 2013. Early diet impacts infant rhesus gut microbiome, immunity, and metabolism. *J Proteome Res* 12:2833–2845. <https://doi.org/10.1021/pr4001702>.
  83. Sela DA, Chapman J, Adeuya A, Kim JH, Chen F, Whitehead TR, Lapidus A, Rokhsar DS, Lebrilla CB, German JB, Price NP, Richardson PM, Mills DA. 2008. The genome sequence of *Bifidobacterium longum* subsp. infantis reveals adaptations for milk utilization within the infant microbiome. *Proc Natl Acad Sci U S A* 105:18964–18969. <https://doi.org/10.1073/pnas.0809584105>.
  84. Budding AE, Grasman ME, Eck A, Bogaards JA, Vandenbroucke-Grauls CM, van Bodegraven AA, Savelkoul PH. 2014. Rectal swabs for analysis of the intestinal microbiota. *PLoS One* 9:e101344. <https://doi.org/10.1371/journal.pone.0101344>.
  85. Bassis CM, Moore NM, Lolans K, Seekatz AM, Weinstein RA, Young VB, Hayden MK, CDC Prevention Epicenters Program. 2017. Comparison of stool versus rectal swab samples and storage conditions on bacterial community profiles. *BMC Microbiol* 17:78. <https://doi.org/10.1186/s12866-017-0983-9>.
  86. Patel AL, Mutlu EA, Sun Y, Koenig L, Green S, Jakubowicz A, Mryan J, Engen P, Fogg L, Chen AL, Pombar X, Meier PP, Keshavarzian A. 2016. Longitudinal survey of microbiota in hospitalized preterm very-low-birth-weight infants. *J Pediatr Gastroenterol Nutr* 62:292–303. <https://doi.org/10.1097/MPG.0000000000000913>.
  87. Pollock J, Glendinning L, Wisedchanwet T, Watson M. 2018. The madness of microbiome: attempting to find consensus “best practice” for 16S microbiome studies. *Appl Environ Microbiol* 84:e02627-17. <https://doi.org/10.1128/AEM.02627-17>.
  88. Gilbert JA, Jansson JK, Knight R. 2014. The Earth Microbiome project: successes and aspirations. *BMC Biol* 12:69. <https://doi.org/10.1186/s12915-014-0069-1>.
  89. Gilbert JA, Meyer F, Jansson J, Gordon J, Pace N, Tiedje J, Ley R, Fierer N, Field D, Kyrpides N, Glockner FO, Klenk HP, Wommack KE, Glass E, Docherty K, Gallery R, Stevens R, Knight R. 2010. The Earth Microbiome Project: Meeting report of the “1st EMP meeting on sample selection and acquisition” at Argonne National Laboratory October 6th 2010. *Stand Genomic Sci* 3:249–253. <https://doi.org/10.4056/aigs.1443528>.
  90. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 37:852–857. <https://doi.org/10.1038/s41587-019-0209-9>.
  91. Almeida A, Mitchell AL, Tarkowska A, Finn RD. 2018. Benchmarking taxonomic assignments based on 16S rRNA gene profiling of the microbiota from commonly sampled environments. *Gigascience* 7:gy054. <https://doi.org/10.1093/gigascience/giy054>.
  92. Arrieta MC, Stiemsma LT, Amenyogbe N, Brown EM, Finlay B. 2014. The intestinal microbiome in early life: health and disease. *Front Immunol* 5:427. <https://doi.org/10.3389/fimmu.2014.00427>.

93. Mehta S, Finkelstein JL, Venkatraman S, Huey SL, Udipi SA, Ghugre P, Ruth C, Canfield RL, Kurpad AV, Potdar RD, Haas JD. 2017. Effect of iron and zinc-biofortified pearl millet consumption on growth and immune competence in children aged 12–18 months in India: study protocol for a randomised controlled trial. *BMJ Open* 7:e017631. <https://doi.org/10.1136/bmjopen-2017-017631>.
94. Naska A, Lagiou A, Lagiou P. 2017. Dietary assessment methods in epidemiological research: current state of the art and future prospects. *F1000Res* 6:926. <https://doi.org/10.12688/f1000research.10703.1>.
95. Longvah T, Ananthan R, Bhaskarachary K, Venkaiah K. 2017. Indian food composition tables. National Institute of Nutrition (Indian Council of Medical Research), Hyderabad, Telangana State, India.
96. Ministry of Health and Family Welfare, Government of India, UNICEF and Population Council. 2019. Comprehensive National Nutrition Survey (CNNS) 2016 – 2018 National Report. Ministry of Health and Family Welfare, Government of India, UNICEF and Population Council, New Delhi, India. [https://www.popcouncil.org/uploads/pdfs/2019RH\\_CNNSreport.pdf](https://www.popcouncil.org/uploads/pdfs/2019RH_CNNSreport.pdf).
97. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkpile DE, Vega Thurber RL, Knight R, Beiko RG, Huttenhower C. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* 31: 814–821. <https://doi.org/10.1038/nbt.2676>.
98. Indian Council of Medical Research. 2006. Ethical guidelines for biomedical research on human participants. Indian Council of Medical Research, New Delhi, India.
99. WHO. 2017. Guideline: Preventive chemotherapy to control soil-transmitted helminthiasis averted more than 500,000 DALYs in 2015. World Health Organization, Geneva, Switzerland.
100. Gibson RS. 1993. Nutritional assessment: a laboratory manual. Oxford University Press, New York, NY.
101. Edmond K, Bahl R. 2006. Optimal feeding of low-birth-weight infants. Technical review. World Health Organization, Geneva, Switzerland.
102. WHO. 2002. International statistical classification of diseases and related health problems. World Health Organization, Geneva, Switzerland.
103. WHO. 2010. Indicators for assessing infant and young child feeding practices. Part 2: Measurement, indicators for assessing infant and young child feeding practices. World Health Organization, Geneva, Switzerland.
104. Willett W. 2013. Nutritional epidemiology, 3rd ed. Oxford University Press, Oxford, England.
105. Marotz C, Amir A, Humphrey G, Gaffney J, Gogul G, Knight R. 2017. DNA extraction for streamlined metagenomics of diverse environmental samples. *Biotechniques* 62:290–293. <https://doi.org/10.2144/000114559>.
106. Berg-Lyons D, Lauber CL, Humphrey G, Thompson L, Gilbert JA, Jansson JK, Knight R. 2018. EMP DNA extraction protocol. <http://press.igsb.anl.gov/earthmicrobiome/protocols-and-standards/dna-extraction-protocol/>. Accessed 13 August 2020.
107. Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, Knights D, Gajer P, Ravel J, Fierer N, Gordon JI, Knight R. 2011. Moving pictures of the human microbiome. *Genome Biol* 12:R50. <https://doi.org/10.1186/gb-2011-12-5-r50>.
108. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 6:1621–1624. <https://doi.org/10.1038/ismej.2012.8>.
109. Parada AE, Needham DM, Fuhrman JA. 2016. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol* 18:1403–1414. <https://doi.org/10.1111/1462-2920.13023>.
110. Apprill A, McNally S, Parsons R, Weber L. 2015. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat Microb Ecol* 75:129–137. <https://doi.org/10.3354/ame01753>.
111. Illumina. 2013. bcl2fastq conversion user guide. Illumina, San Diego, CA.
112. Adair LS, Fall CHD, Osmond C, Stein AD, Martorell R, Ramirez-Zea M, Sachdev HS, Dahly DL, Bas I, Norris SA, Micklefield L, Hallal P, Victora CG. 2013. Associations of linear growth and relative weight gain during early life with adult health and human capital in countries of low and middle income: findings from five birth cohort studies. *Lancet* 382: 525–534. [https://doi.org/10.1016/S0140-6736\(13\)60103-8](https://doi.org/10.1016/S0140-6736(13)60103-8).
113. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso JG. 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods* 10:57–59. <https://doi.org/10.1038/nmeth.2276>.
114. Janssen S, McDonald D, Gonzalez A, Navas-Molina JA, Jiang L, Xu ZZ, Winker K, Kado DM, Orwoll E, Manary M, Mirarab S, Knight R. 2018. Phylogenetic placement of exact amplicon sequences improves associations with clinical information. *mSystems* 3:e00021-18. <https://doi.org/10.1128/mSystems.00021-18>.
115. Matsen FA, Hoffman NG, Gallagher A, Stamatakis A. 2012. A format for phylogenetic placements. *PLoS One* 7:e31009. <https://doi.org/10.1371/journal.pone.0031009>.
116. Matsen FA, Kodner RB, Armbrust EV. 2010. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* 11:538. <https://doi.org/10.1186/1471-2105-11-538>.
117. Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* 7:e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>.
118. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. 2012. *ISME J* 6:610–618. <https://doi.org/10.1038/ismej.2011.139>.
119. McDonald D, Clemente JC, Kuczynski J, Rideout JR, Stombaugh J, Wendel D, Wilke A, Huse S, Hufnagle J, Meyer F, Knight R, Caporaso JG. 2012. The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. *Gigascience* 1:7. <https://doi.org/10.1186/2047-217X-1-7>.
120. Lozupone C, Knight R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 71: 8228–8235. <https://doi.org/10.1128/AEM.71.12.8228-8235.2005>.
121. Lozupone CA, Hamady M, Kelley ST, Knight R. 2007. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol* 73: 1576–1585. <https://doi.org/10.1128/AEM.01996-06>.
122. Greenland S. 1989. Modeling and variable selection in epidemiologic analysis. *Am J Public Health* 79:340–349. <https://doi.org/10.2105/ajph.79.3.340>.
123. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, Kurilshikov A, Bonder MJ, Valles-Colomer M, Vandeputte D, Tito RY, Chaffron S, Rymenans L, Verspecht C, De Sutter L, Lima-Mendez G, D'hoey K, Jonckheere K, Homola D, Garcia R, Tigchelaar EF, Eeckhaut L, Fu J, Henckaerts L, Zhernakova A, Wijmenga C, Raes J. 2016. Population-level analysis of gut microbiome variation. *Science* 352:560–564. <https://doi.org/10.1126/science.aad3503>.
124. Vazquez-Baeza Y, Pirrung M, Gonzalez A, Knight R. 2013. EMPeror: a tool for visualizing high-throughput microbial community data. *Gigascience* 2:16. <https://doi.org/10.1186/2047-217X-2-16>.
125. Kelly BJ, Gross R, Bittinger K, Sherrill-Mix S, Lewis JD, Collman RG, Bushman FD, Li H. 2015. Power and sample-size estimation for microbiome studies using pairwise distances and PERMANOVA. *Bioinformatics* 31:2461–2468. <https://doi.org/10.1093/bioinformatics/btv183>.
126. Anderson MJ. 2017. Permutational multivariate analysis of variance (PERMANOVA). *Wiley StatsRef: Statistics Reference Online*. <https://doi.org/10.1002/9781118445112.stat07841>.
127. Aitchison J, Greenacre M. 2002. Biplots of compositional data. *J R Stat Soc Ser C* 51:375–392. <https://doi.org/10.1111/1467-9876.00275>.
128. de Onis M, Onyango A, Borghi E, Siyam A, Blossner M, Lutter C, WHO Multicentre Growth Reference Study Group. 2012. Worldwide implementation of the WHO Child Growth Standards. *Public Health Nutr* 15:1603–1610. <https://doi.org/10.1017/S136898001200105X>.