Supporting Information

De Filippo et al. 10.1073/pnas.1005963107

SI Materials and Methods

Fecal Sample Collection. All individuals were made aware of the nature of the experiment, and all gave written informed consent in accordance with the sampling protocol approved by the Ethical Committee of Meyer Children Hospital, Florence, Italy. In both populations, fecal samples were collected by physicians from each individual in the morning, 1–2 h after the first meal, and preserved in RNAlater (Qiagen) at 4 °C for the first 48 h, and then kept at −80 °C until extraction of nucleic acids. The major obstacle to analysis of fecal samples in a distant area such as rural Burkina Faso is that standard fecal collection procedures require fresh or frozen samples, which limit its application in rural Africa.

As new technologies have become available to preserve tissue DNA and RNA for some time at room temperature, the application of such technologies to fecal samples may have great potential for epidemiological studies (1). During transport and storage, we lost one of the BF fecal samples the DNA in which did not pass quality control, reducing the BF samples from 15 to 14.

DNA Extraction. The genomic DNA extraction procedure was based on a protocol proposed by Zoetendal et al. (2). After dissolving ∼500 mg of each fecal sample in physiological solution and homogenization by vigorous hand shaking, 600 μL suspension was centrifuged (10,000 × *g* for 10 min at 4 °C) to obtain pellets. The pellets were dissolved in 1 mL ice-cold 1× PBS and centrifuged at 700 × *g* at 4 °C for 1 min. The supernatants were transferred into a 15-mL tube and were centrifuged at 9,000 × *g* at 4 °C for 5 min. Subsequently, the pellets were suspended in 2.8 mL TE buffer by repeated pipetting. Then 180 μL SDS 10% (wt/vol) and 18 μL of proteinase K (20 mg/mL) were added. The samples were incubated for 1 h at 37 °C. Afterward, 20 μL RNase (40 μg/mL) was added and incubated at RT for 5 min. An equal volume of phenol/chloroform (50:50) was added, and the samples were shaken well until the phases were completely mixed. The mixtures were centrifuged at 4500 × *g* for 2 min. The upper layers were then transferred into a new tube. This step was repeated again so that the interface of the two layers was clean. Next, 1/10 volume 3 M sodium acetate, pH 5.2, and two volumes 96% ethanol were added and mixed gently. The mixtures were stored overnight at −20 °C to precipitate the genomic DNA, and then the samples were centrifuged at 4 °C at 9,000 × *g* for 10 min. The precipitated genomic DNA was washed twice in 1 mL 70% ethanol. Finally, dried samples were suspended in 200–400 μL H_2O. DNA quality was assessed by gel electrophoresis and spectrophotometry measuring OD 260/280. Only samples with good DNA quality were processed. Extracted DNA was stored at −20 °C.

PCR Amplification of V5-V6 Region of Bacterial 16S rRNA Genes. For each sample, we amplified 16S rRNA genes using a primer set corresponding to primers 784F and 1061R described by Andersson et al. (3). These PCR primers target the V5 and V6 hypervariable 16S rRNA region. The forward primer contained the sequence of the Titanium A adaptor (5′-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3′) and a barcode sequence. For each sample, a PCR mix of 100 μL was prepared containing 1× PCR buffer, 5U of FastStart High Fidelity polymerase blend and dNTPs from the FastStart High Fidelity PCR system (Roche), 200 nM primers (Eurogentec), and 100 ng gDNA. Thermal cycling consisted of initial denaturation at 95 °C for 5 min, followed by 30 cycles of denaturation at 95 °C for 40 s, annealing at 55 °C for 40 s, and extension at 72 °C for 1 min, with a final extension of 7 min at 72 °C. Amplicons were visualized on 1.0% agarose gels using SYBR Safe DNA gel stain in 0.5× TBE (Invitrogen) and were cleaned using the HighPure Cleanup kit (Roche) according to the manufacturer's instructions.

Amplicon Quantitation, Pooling, and Pyrosequencing. Amplicon DNA concentrations were determined using the Quant-iT PicoGreen dsDNA reagent and kit (Invitrogen) following the manufacturer's instructions. Assays were carried out using 10 μL cleaned PCR product in a total reaction volume of 200 μL in black, 96-well microtiter plates. Fluorescence was measured on Perkin-Elmer Victor Plate reader using the 485/530 nm excitation/emission filter pair with measurement time 0.1 s. Following quantitation, cleaned amplicons were combined in equimolar ratios into a single tube. The final pool of DNA was precipitated on ice for 45 min following the addition of 5 M NaCl (0.2 M final concentration) and two volumes of ice-cold 100% ethanol. The precipitated DNA was centrifuged at 7,800 × *g* for 40 min at 4 °C, and the resulting pellet was washed with an equal volume of ice-cold 70% ethanol and centrifuged again at 7,800 × *g* for 20 min at 4 °C. The supernatant was removed, and the pellet was air dried for 10 min at room temperature and then resuspended in 100 μL nuclease-free water (Ambion). The final concentration of the pooled DNA was determined using a NanoDrop spectrophotometer (Thermo Fisher). Pyrosequencing was carried out using primer A on a 454 Life Sciences Genome Sequencer FLX instrument (Roche) following titanium chemistry.

Pyrosequencing Quality Control. The pyrosequencing produced a total of 438,219 reads of 16S rDNA reads. The sequences were assigned to samples according to sample-specific barcodes. This allowed us to collect FASTA formatted files containing an average (± SD) of 15,111 ± 3774 sequences per sample. Sequences were then checked for the following criteria: (*i*) almost perfect match with barcode and primers; (*ii*) length of at least 150 nucleotides (barcodes and primers excluded); (*iii*) no more than two undetermined bases (denoted by N). By "almost perfect match," we mean that one mismatch/deletion/insertion is allowed in the barcode, idem for the primer. After this quality check, most of the sequences resulted with a length of ∼260 bp. Data were submitted to the Sequence Read Archive (SRA) using ISA tools (ISAcreator and ISAconverter, http://isatab.sourceforge.net/index.html). The dataset is available at http://www.ebi.ac.uk/ena/data/view/ERP000133.

Complete Linkage Hierarchical Clustering. The clustering was performed on genera obtained from RDP Classifier by means of a complete linkage hierarchical clustering technique using the R package hclust (http://sekhon.berkeley.edu/stats/html/hclust.html). The clustering function is able to find the most similar clusters by performing a hierarchical cluster analysis using a set of dissimilarities for the *n* objects being clustered. Initially, each object is assigned to its own cluster, and then the algorithm proceeds iteratively, at each stage joining the two most similar clusters, continuing until there is just one cluster. At each stage, distances between clusters are recomputed by the Lance–Williams dissimilarity update formula according to the particular clustering method being used.

Assignment at the Species Level. Bacterial species were assigned with BLAST using a speed-optimized procedure. Briefly, the per-read genus assignment by RDP classifier was used to construct subsamples of the RDP 10, Update 18 database containing the 16S rRNAs sequences of all species assigned to a specific genus by

the project curators. After proper formatting, the corresponding sequence was searched with the BLASTn program using the first hit method (keeping only the outcomes with the lowest E-value, given a minimal E-value of 10^{-3}). A dedicated scoring system was built to properly weight BLAST results: in fact, due to both the reduced size of the sequencing reads and the extreme similarity shared by some species, the first hit method frequently returned a collection of species rather than a single species. In that case, given N outcomes for a sequence, a 1/N score was given to each species. Such a scoring system allows species to be weighted by probability, maintaining a score of 1 when a unique assignment was identified and guaranteeing a balanced attribution.

Determination of SCFAs in Fecal Samples. For determination of SCFAs, we used 1 aliquot of frozen fecal samples (~250 mg). Briefly, fecal samples were homogenized after addition of 1 mL 10% perchloric acid and centrifuged at $15,000 \times g$ for 5 min at 4 °C. Concentrations of SCFAs were determined in a 1:25 dilution of 500 µL supernatant. We used 5 µL of a mixture of deuterated acids containing 50 ng D_3-propionic, 50 ng D_7-butyric acid, and 500 ng D_4-acetic acid as internal standard. A calibration curve was prepared, adding the mixture of internal standards (5 µL) to scalar amounts of the acids. SPME-GC-MS determinations were performed using a Varian Saturn 2000 GC-MS instrument with 8200 CX SPME autosampler. The SPME fiber was a Carboxen/Divinylbenzene 75 µm. The capillary column was an Agilent HP-Innowax 30 m × 0.25 mm, 0.5-µm film thickness. The injector and transfer line temperatures were 290 °C and 260 °C, respectively; the ion trap temperature was 180 °C. Absorption of analyte was performed in the headspace of the sample solution for 3 min at 70 °C; the analytes were desorbed in the GC injector port at 290 °C for 20 min. The GC oven temperature program was as follows: initial temperature 45 °C for 0.15 min, then to 123 °C at 2 °C/min, to 159 °C at 6 °C/min, and to 200 °C at 20 °C/min. The retention times for individual SCFAs were determined by injecting each standard into the column. The Varian MS workstation software (version 6.6) was used for data acquisition and processing. The SCFA concentration in fecal sample was expressed in micromoles per gram (µmol/g) of feces. To determine statistical significance of differences observed between BF and EU populations, we used an unpaired Student *t* test (one tailed).

Taxonomy Assignment to 16S rDNA Reads. Each sequence originating from pyrosequencing that was assigned to a genus by the RDP classifier (v 2.1) with CE > 50% was subject to species assignment using a newly developed experimental algorithm. For each sequence, we were able to identify species with a maximal uncertainty of 3 on average in 87.5% of samples (unique species were found in 60% of sequences). Our system proved to behave well also on scarcely confident genus assignments, as the inclusion of sequences with genus assignment CE < 50% did not alter the percentage of species assignment with an uncertainty of 3 (although the unique species assignments were reduced to 55%). Table S4 reports, for each sample, the percentage of sequences that are classified with a CE < 50% at different phylogenetic ranks, providing a 1%, 4%, and 26% average accuracy at the phylum, family, and genus level, respectively.

Statistical Analyses for BF and EU Comparisons. With the aim of evidencing the statistical significance of differences observed in the two BF and EU groups, the data were further analyzed using both parametric and nonparametric methods, namely the univariate and multivariate ANOVA and Kruskal–Wallis tests respectively (4, 5). As the design is well balanced in terms of quantity of sample, we initially tried to use the MANOVA to test whether the centroids of the two populations were significantly different, considering several variables at the same time, and is based on the following assumptions: (*i*) the response (dependent) variables are continuous; (*ii*) the residuals follow the multivariate-normal probability distribution with means equal to zero; (*iii*) the variance–covariance matrices of each group of residuals are equal; and (*iv*) the individuals are independent. The MANOVAs were balanced so that there was an equal number of observations in each group, guaranteeing the robustness of the analyses. Because normality is an important assumption in the MANOVA, normality was tested with the Shapiro–Wilk and the E-statistic (energy) tests. Both tests indicated that the multivariate data composed of Actinobacteria, Bacteroidetes, Firmicutes, and Proteobacteria were not normal in group BF (Shapiro–Wilk $W = 0.7946$, $P = 4.28 \times 10^{-3}$, energy E-statistic = 1.1967, $P = 8.0 \times 10^{-3}$). Removing potential outliers (the younger children 3BF, 4BF, and 16BF) did not seem to improve the normality of the data. Similar observations could be made in the EU group (Shapiro–Wilk $W = 0.7285$, $P = 5.1 \times 10^{4}$, energy E-statistics = 1.7475, $P = 2.20 \times 10^{-16}$). Similarly to the BF group, the removal of potential outliers (1EU, 2EU, 3EU, and 13EU) did not improve the normality of the data. Despite those results, it has to be underlined that the power of such tests is directly proportional to the sample size: in our analyses, the sample size was quite small and there is some risk to erroneously reject the null hypothesis for data following a multivariate normal distribution.

Despite these limitations, our MANOVA indicated that the BF and EU groups were significantly different ($P = 8.14 \times 10^{-4}$) when all phyla were considered.

Univariate effects (ANOVA) for the above-mentioned phyla were also evaluated, and were in partial agreement with multivariate techniques, as Bacteroidetes and Firmicutes were significantly different between BF and EU groups ($P = 7.89 \times 10^{-5}$ and $P = 1.19 \times 10^{-6}$, respectively) whereas Actinobacteria and Proteobacteria differences were not significant, at least with phylum assignments with CE > 50%. In fact, the latter phyla were found to be significantly different between EU and BF at more tolerant (lower) CE threshold values, indicating that a substantial difference could be seriously considered. Less represented phyla were also evaluated and significant differences were found only in Spirochaetes ($P = 1.09 \times 10^{-3}$).

The nonparametric Kruskal–Wallis rank sum test was used as an alternative for exploring differences in the BF and EU groups. In this case, univariate effects of phyla were considered and results were fully coherent and more explicit than with parametric tests: among the most represented phyla, significant differences were observed in Actinobacteria, Bacteroidetes, and Firmicutes ($P = 8.80 \times 10^{-3}$, $P = 4.80 \times 10^{-4}$ and $P = 3.38 \times 10^{-5}$, respectively), whereas, as above, differences in Proteobacteria were not significant at CE > 50% and significant at lower CE ($P = 5.65 \times 10^{-6}$). Significant differences were also found in less represented phyla such as Spirochaetes ($P = 1.112 \times 10^{-5}$) and Tenericutes ($P = 1.29 \times 10^{-2}$).

1. Nechvatal JM, et al. (2008) Fecal collection, ambient preservation, and DNA extraction for PCR amplification of bacterial and human markers from human feces. *J Microbiol Methods* 72:124–132.
2. Zoetendal EG, et al. (2006) Isolation of RNA from bacterial samples of the human gastrointestinal tract. *Nat Protoc* 1:954–959.
3. Andersson AF, et al. (2008) Comparative analysis of human gut microbiota by barcoded pyrosequencing. *PLoS ONE* 3:e2836.
4. Johnson RA, Wichern DW (2007) In *Applied Multivariate Statistical Analysis*, ed Hall P (Upper Saddle River, NJ), 6th Ed.
5. Conover WJ (1998) *Practical Nonparametric Statistics* (Wiley, New York),3rd Ed.

PNAS PNAS PNAS PNAS

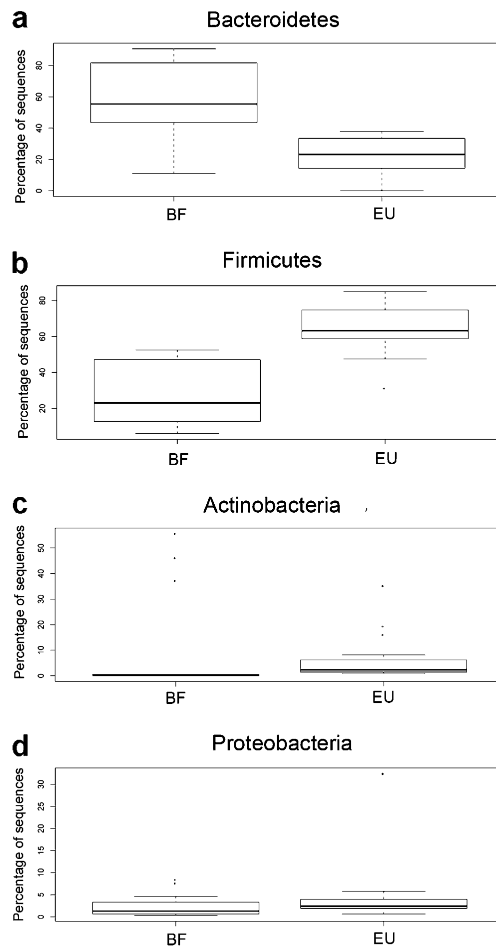

Fig. S1. Boxplots (percentage of sequences) of the four most represented phyla. Bacteroidetes (*A*), Firmicutes (*B*), Actinobacteria (*C*), and Proteobacteria (*D*) in the BF and EU children.

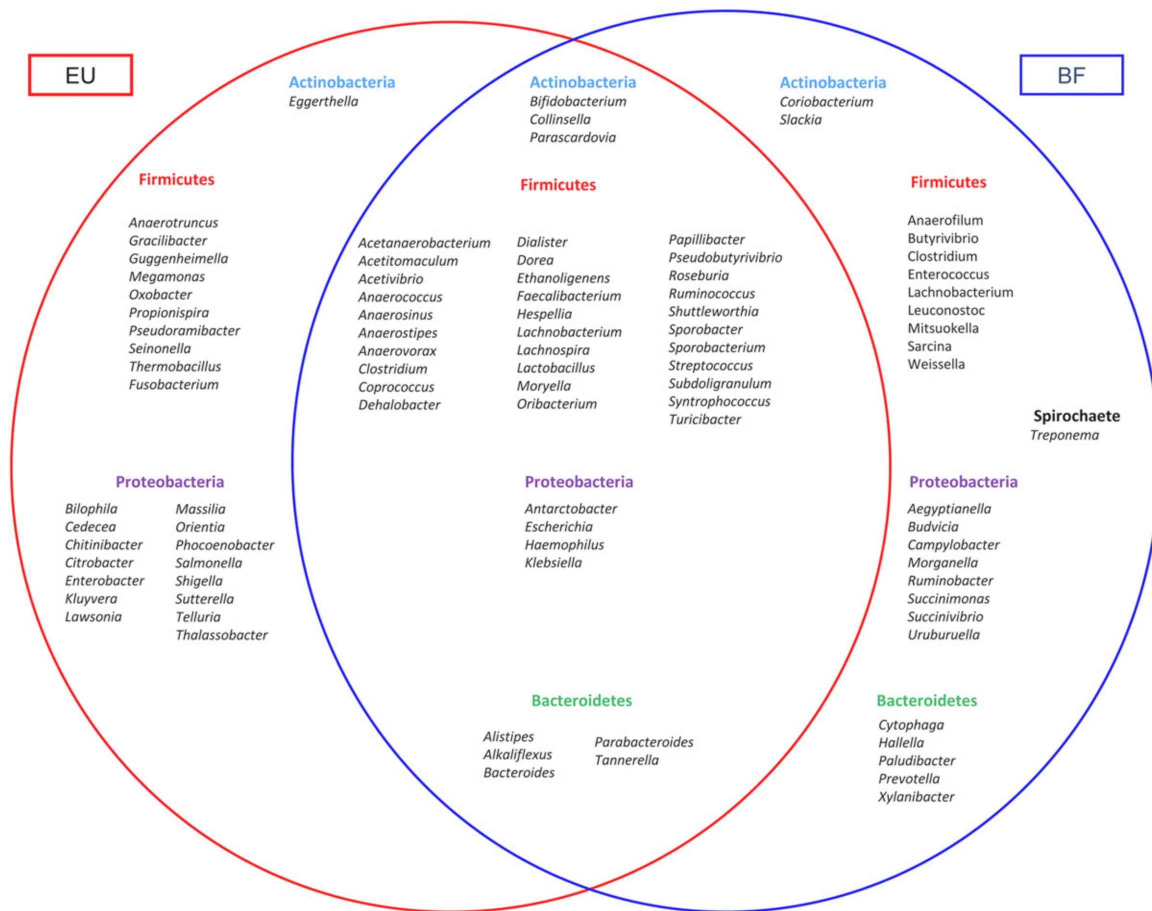

Fig. S2. Venn diagram showing the number of genera belonging to the major phyla differing significantly between BF and EU children. A total of 26 genera are characteristic of EU children, 43 genera are in common between BF and EU children, and 25 genera are characteristic of BF children.

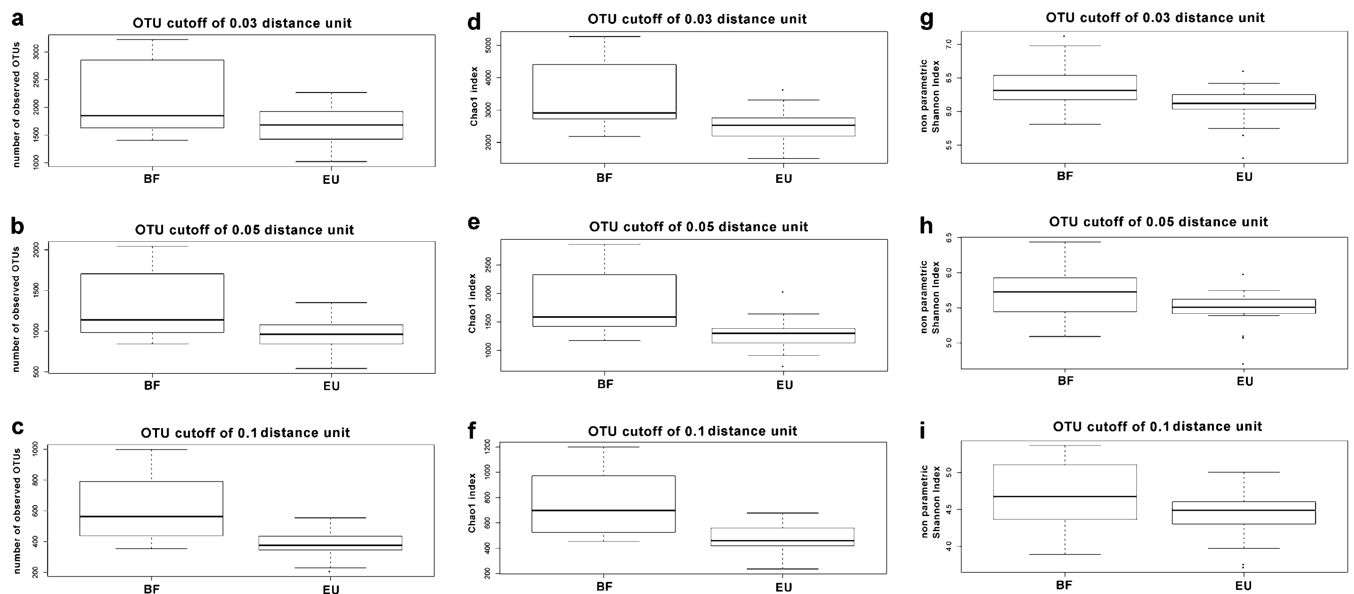

Fig. S3. (*A*–*C*) Boxplots (percentage of sequences) of the observed OTUs at OTU cutoffs of (*A*) 0.03, (*B*) 0.05, and (*C*) 0.10 distance units in the BF and EU children. (*D*–*F*) Boxplots (percentage of sequences) of the Chao1 indexes at OTU cutoffs of (*D*) 0.03, (*E*) 0.05, and (*F*) 0.10 distance units in the BF and EU children. (*G*–*I*) Boxplots (percentage of sequences) of the nonparametric Shannon indexes at OTU cutoffs of (*G*) 0.03, (*H*) 0.05, and (*I*) 0.10 distance units in the BF and EU children.

Table S1. Characteristics of study sample

ID	Age (y)	Sex	Provenance	Clinical condition	Delivery	Months without antibiotics
2BF	5	M	Burkina Faso	No malnutrition	Natural childbirth	Never used
3BF	2	F	Burkina Faso	No malnutrition	Natural childbirth	Never used
4BF	2	M	Burkina Faso	No malnutrition	Natural childbirth	Never used
6BF	6	F	Burkina Faso	No malnutrition	Natural childbirth	Never used
7BF	6	M	Burkina Faso	No malnutrition	Natural childbirth	Never used
8BF	6	M	Burkina Faso	No malnutrition	Natural childbirth	Never used
9BF	6	M	Burkina Faso	No malnutrition	Natural childbirth	Never used
10BF	6	F	Burkina Faso	No malnutrition	Natural childbirth	Never used
11BF	5	M	Burkina Faso	No malnutrition	Natural childbirth	Never used
12BF	6	M	Burkina Faso	No malnutrition	Natural childbirth	Never used
13BF	6	M	Burkina Faso	No malnutrition	Natural childbirth	Never used
15BF	6	M	Burkina Faso	No malnutrition	Natural childbirth	Never used
16BF	1	F	Burkina Faso	No malnutrition	Natural childbirth	Never used
17BF	5	F	Burkina Faso	No malnutrition	Natural childbirth	Never used
1EU	2	M	Tuscany (Italy)	Healthy	Cesarean childbirth	>6 mo
2EU	1	F	Tuscany (Italy)	Healthy	Natural childbirth	>6 mo
3EU	1	F	Tuscany (Italy)	Healthy	Natural childbirth	>6 mo
5EU	5	M	Tuscany (Italy)	Healthy	Natural childbirth	>6 mo
6EU	6	M	Tuscany (Italy)	Healthy	Natural childbirth	>6 mo
8EU	5	M	Tuscany (Italy)	Healthy	Natural childbirth	>6 mo
10EU	5	M	Tuscany (Italy)	Healthy	Natural childbirth	>6 mo
11EU	5	M	Tuscany (Italy)	Healthy	Natural childbirth	>6 mo
12EU	6	M	Tuscany (Italy)	Healthy	Natural childbirth	>6 mo
13EU	5	M	Tuscany (Italy)	Healthy	Natural childbirth	>6 mo
17EU	5	M	Tuscany (Italy)	Healthy	Natural childbirth	>6 mo
18EU	3	F	Tuscany (Italy)	Healthy	Natural childbirth	>6 mo
19EU	4	F	Tuscany (Italy)	Healthy	Natural childbirth	>6 mo
20EU	5	F	Tuscany (Italy)	Healthy	Natural childbirth	>6 mo
21EU	3	F	Tuscany (Italy)	Healthy	Cesarean childbirth	>6 mo

Table S2. Total daily food intake in terms of protein, fat, carbohydrate and fiber in relation to the average of maximum quantity ingested per day relative to BF children

Dish component°	Daily Q* (g)	Food energy (kcal)	Moisture (%)	Protein (g)	Fat (g)	Carbohydrate (including fiber) (g)	Fiber (g)
1–2 y old							
Cereals and starchy component (millet, sorghum)	120.0	344.1	34.3	7.1	2.7	73.2	0.8
Legumes (black-eyed peas, Niebè)	40.0	168.0	2.8	19.0	9.2	6.3	4.3
Vegetables (Nerè)	50.0	46.0	39.3	2.8	3.2	3.6	0.5
Fruit (mango, papaya)	90.0	33.8	80.1	0.6	0.1	8.6	4.4
Milk (breast milk)	120.0	80.4	103.8	1.3	3.7	10.9	0.0
Total daily food intake	420.0	672.2	260.3	30.9	18.9	102.6	10.0
2–6 y old							
Cereals and starchy component (millet, sorghum)	170.0	495.4	46.4	10.1	4.1	105.5	1.1
Legumes (black-eyed peas, Niebè)	70.0	267.4	6.0	25.8	8.5	26.3	6.2
Vegetables (Nerè)	60.0	55.2	47.1	3.4	3.8	4.3	0.6
Fruit (mango, papaya)	130.0	48.8	115.7	0.9	0.2	12.5	6.4
Fat (Karitè butter)	15.0	129.3	0.2	0.0	14.7	0.1	0.0
Total daily food intake	445.0	996.1	215.4	40.2	31.2	148.6	14.2

°Nutritional composition of foods is available at http://www.fao.org/docrep/003/x6877e/X6877E00.htm.
*Average of maximum quantity ingested per child per day.

Table S3. Total daily food intake in terms of protein, fat, carbohydrate and fiber in relation to the average of maximum quantity ingested per day relative to EU children

Dish component°	Daily Q* (g)	Food energy (kcal)	Moisture (%)	Protein (g)	Fat (g)	Carbohydrate (including fiber) (g)	Fiber (g)
1–2 y old							
Cereals and starchy component (bread, biscuits, pasta, rice)	130.0	266.3	35.0	8.6	2.9	107.5	1.6
Vegetables (carrot, potato, tomato, zucchini)	100.0	56.6	82.5	2.2	1.3	20.7	1.8
Fruit (apple, pear, banana)	100.0	50.2	83.0	0.7	0.2	26.5	2.2
Milk (breast milk)	300.0	210.0	262.5	3.0	13.2	20.7	0.0
Milk derivatives (parmesan, mozzarella, yogurt)	100.0	158.4	74.2	8.0	13.8	4.6	0.0
Meat (chicken, beef), fish (cod, sole)	80.0	132.2	53.0	19.3	4.8	6.0	0.0
Oil (extra virgin olive oil)	20.0	179.8	0.0	0.0	20.0	0.0	0.0
Honey	5.0	15.2	0.9	0.0	0.0	4.0	0.0
Total daily food intake	835.0	1068.7	591.1	41.9	56.1	190.0	5.6
2–6 y old							
Cereals and starchy component (bread, biscuits, pasta, rice)	160.0	375.0	68.0	9.7	4.9	152.4	3.5
Legumes (beans, peas)	20.0	10.9	15.4	1.2	0.1	3.7	0.9
Vegetables (carrot, potato, tomato, zucchini)	100.0	50.9	82.3	2.3	1.1	18.4	1.6
Fruit (apple, pear, banana)	140.0	83.5	114.9	1.1	0.3	43.0	2.4
Milk (cow's milk)	140.0	64.4	125.3	4.5	2.2	14.3	0.0
Milk derivatives (parmesan, mozzarella, yogurt)	150.0	250.3	106.8	16.8	19.9	6.5	0.0
Meat (chicken, beef), fish (cod, sole)	120.0	157.7	83.9	26.0	4.6	6.2	0.0
Egg	30.0	91.1	16.9	2.9	8.7	0.4	0.0
Oil and fats (extra-virgin olive oil, butter)	25.0	210.7	1.4	0.1	23.3	0.2	0.0
Sugar	10.0	38.7	0.0	0.0	0.0	10.0	0.0
Snacks	30.0	179.6	18.8	2.2	8.7	34.9	0.0
Total daily food intake	925.0	1512.7	633.6	66.7	73.9	290.0	8.4

°Nutritional composition of foods is available from http://www.inran.it.
*Average of maximum quantity ingested per child per day.

Table S4. Sequences summary: Percentage of sequences classified with a confidence estimation (CE) of <50% at phylum, family, and genus levels relative to each sample

Sample ID	No. of sequences (reads)	Phylum with CE < 50 (%)	Family with CE < 50 (%)	Genus with CE < 50 (%)
2BF	9,171	1.5	5.3	38.1
3BF	14,893	1.3	1.9	11.1
4BF	13,896	1.6	2.8	9.9
6BF	10,415	1.2	6.1	36.9
7BF	12,160	2.1	6.8	35.2
8BF	11,565	1.2	6.1	38.7
9BF	13,267	1.5	6.6	39.2
10BF	19,795	1.2	3.4	29.1
11BF	14,361	1.8	4.1	22.0
12BF	23,742	1.0	2.3	23.5
13BF	16,562	1.2	3.9	29.5
15BF	11,353	2.3	3.7	36.0
16BF	20,003	1.6	3.7	18.1
17BF	17,965	1.1	3.3	21.8
1EU	9,055	0.6	2.3	10.2
2EU	12,789	0.3	0.7	20.2
3EU	13,566	0.2	2.6	29.5
5EU	21,419	0.3	2.0	17.8
6EU	17,012	0.6	2.5	26.9
8EU	17,739	1.0	4.4	37.9
10EU	12,320	1.2	12.3	39.7
11EU	14,955	0.3	2.9	14.8
12EU	12,487	0.6	2.3	30.0
13EU	19,577	0.4	2.4	19.1
17EU	13,067	0.5	2.7	26.1
18EU	16,143	0.3	1.7	20.7
19EU	11,932	0.5	6.4	27.1
20EU	17,466	1.2	11.2	33.3
21EU	19,544	0.3	1.7	27.3

Sequences classified with a confidence estimate <50% are denoted by "uncertain."

Table S5. Report of the species assignment for BF and EU populations relative to the most abundant bacterial genera found by the RDP classifier

Genus	Species in BF	Species in EU
Alistipes		*A. finegoldii*
		A. indistinctus
		A. onderdonkii
		A. putredinis
Bacteroides	*Bacteroides sp. TP-5*	*B. caccae*
		B. coprocola
		B.eggerthii
		B. fragilis
		B. intestinalis
		B. massiliensis
		B. ovatus
		B. plebeius
		B.sp. AR20
		B.sp. XO77B42
		B. thetaiotaomicron
		B. uniformis
		B. vulgatus
		Odoribacter splanchnicus
		Swine fecal bacterium RF3E-Xyl1
		Swine fecal bacterium RF3G-Cel1
Bifidobacterium	*B. adolescentis*	*B. adolescentis*
	B. animalis subsp. Animalis	*B. bifidum*
	B. animalis subsp. Lactis	*B. breve*
	B. bifidum	*B. longum*
	B. breve	*B. longum bv. Infantis*
	B. longum	*B. longum subsp. Infantis*
	B. longum subsp. Infantis	*B. longum subsp. Longum*
	B. longum subsp. Suis	*B. longum subsp. Suis*
	B. pseudocatenulatum	*B. pseudocatenulatum*
	B. pullorum	*B. ruminantium*
	B. ruminantium	*B. sp. H12*
	B. simiae	*B. sp. PL1*
	B. thermophilum	
Faecalibacterium	*F. prausnitzii*	*F. prausnitzii*
	Butyrate-producing bacterium PH07BY04	*Butyrate-producing bacterium PH07BY04*
	Butyrate-producing bacterium PH07AY5	*Butyrate-producing bacterium PH07AY5*
	Butyrate-producing bacterium M21/2	*Butyrate-producing bacterium M21/2*
Prevotella	*P. aurantiaca*	
	P. brevis	
	P. copri	
	P. denticola	
	P. heparinolytica	
	P. paludivivens	
	P. ruminicola	
	P.sp. BI-42	
	P. sp. DJF_B116	
	P.sp. DJF_LS16	
	P.sp. DJF_RP53	
	P.sp. HY-36–2	
	P. sp. oral clone AH005	
	P. sp. oral clone ID019	
	P. stercorea	
	P. bacterium DJF_CR21k6	
	P.bacterium DJF_CR25	
	P.bacterium DJF_CR62	
	P.bacterium DJF_LS10	
	P. bacterium DJF_RP17	
	P. bacterium DJF_VR15	
	P.bacterium WR041	
	P. aff. ruminicola Tc2-24	
Subdoligranulum		*Subdoligranulum. sp. DJF_VR33k2*
		bacterium ic1395

Table S5. Cont.

Genus	Species in BF	Species in EU
Xylanibacter	*X. oryzae*	
	Bacteroidales str. KB13	
	Bacteroidales str. KB11	
Treponema	*Treponema sp.*	

Table S6. Richness and diversity indexes relative to each fecal sample: Number of observed OTUs (Obs. OTU), the Chao1 index (Chao), and the nonparametric Shannon index (Np Shannon) at OTU cutoffs of 0.03, 0.05, and 0.10 distance units

Sample IDs	Obs OTUs (0.03)	Obs OTUs (0.05)	Obs OTUs (0.10)	Chao (0.03)	Chao (0.05)	Chao (0.10)	Np Shannon (0.03)	Np Shannon (0.05)	Np Shannon (0.10)
2BF	1407.0	878.0	417.0	2182.7	1199.6	506.3	6.3	5.6	4.7
3BF	1654.0	934.0	372.0	2602.1	1305.5	488.9	6.0	5.2	4.1
4BF	1471.0	846.0	355.0	2214.3	1174.3	452.1	5.8	5.1	3.9
6BF	1629.0	1039.0	536.0	2727.1	1483.7	704.3	6.3	5.7	4.9
7BF	1743.0	1097.0	540.0	2889.3	1586.6	660.3	3.3	5.7	4.8
8BF	1822.0	1153.0	585.0	2939.6	1483.7	691.1	6.5	5.9	5.1
9BF	2982.0	1970.0	998.0	4885.9	2820.5	119.9	7.1	6.4	5.4
10BF	3224.0	2013.0	924.0	5273.3	2845.5	1143.2	6.8	6.2	5.1
11BF	1877.0	1126.0	463.0	2877.1	1588.9	558.4	6.2	5.4	4.4
12BF	2856.0	1705.0	755.0	4406.7	2236.9	877.5	6.5	5.8	4.7
13BF	3139.0	2042.0	954.0	5138.4	2863.0	1179.1	7.0	6.3	5.2
15BF	1628.0	985.0	438.0	2846.0	1421.5	526.1	6.1	5.4	4.3
16BF	2442.0	1453.0	623.0	3677.2	1952.2	764.1	6.3	5.6	4.5
17BF	2649.0	1658.0	791.0	4230.3	2332.1	971.4	6.4	5.7	4.6
1EU	1357.0	854.0	376.0	2205.4	1165.1	452.8	6.1	5.5	4.2
2EU	1019.0	542.0	206.0	1501.2	720.2	236.0	5.3	4.7	3.7
3EU	1199.0	672.0	230.0	1802.8	914.3	272.8	5.6	5.1	4.0
5EU	1945.0	1074.0	410.0	2814.1	1362.5	548.2	6.2	5.5	4.5
6EU	2014.0	1152.0	460.0	3144.0	1638.1	571.1	6.4	5.8	4.8
8EU	2116.0	1168.0	463.0	3308.1	1547.0	592.5	6.4	5.7	4.8
10EU	1568.0	918.0	400.0	2602.2	1345.5	481.4	6.2	5.6	4.6
11EU	1431.0	804.0	315.0	2035.3	958.6	360.9	5.8	5.1	3.8
12EU	1759.0	1082.0	476.0	2683.4	1408.6	569.7	6.3	5.7	4.6
13EU	1897.0	1008.0	365.0	2703.8	1321.2	459.5	6.2	5.5	4.5
17EU	1484.0	850.0	344.0	2189.4	1117.0	417.2	6.1	5.5	4.4
18EU	1682.0	964.0	387.0	2347.2	1198.8	442.5	6.0	5.4	4.4
19EU	1422.0	839.0	347.0	2304.7	1146.1	471.6	6.1	5.5	4.6
20EU	2267.0	1352.0	555.0	3621.1	2023.8	678.1	6.6	6.0	5.0
21EU	1762.0	986.0	369.0	2528.9	1301.1	423.7	6.1	5.5	4.6
P value	0.1524	0.0401	0.0055	0.0294	0.0136	0.0045	0.093	0.1647	0.2211

Table S7. Amounts of SCFAs in fecal samples from BF and EU children

Subject ID	Age (y)	SCFAs (μmol/g feces)				
		Total SCFAs	Acetic	Propionic	Butyric	Valeric
2BF	5	103.80	58.25	35.00	8.95	1.60
3BF	2	88.35	51.80	16.30	15.95	4.30
4BF	2	39.00	25.85	10.65	1.80	0.70
6BF	6	47.40	29.80	10.90	5.50	1.20
7BF	6	83.40	37.50	24.90	19.80	1.20
8BF	6	205.70	129.90	49.60	24.50	1.70
9BF	6	137.40	74.05	40.45	19.85	3.05
10BF	6	42.15	26.30	11.20	3.95	0.70
11BF	5	95.50	40.95	33.15	20.45	0.95
12BF	6	57.15	29.00	19.55	7.25	1.35
13BF	6	31.75	23.30	4.60	3.45	0.40
15BF	6	33.40	15.55	12.35	5.00	0.50
16BF	1	49.75	42.60	0.10	6.25	0.80
17BF	5	164.95	65.20	79.95	18.40	1.40
Mean ± SEM		67.80 ± 12.8	34.7 ± 4.4	22.98 ± 7.3	9.25 ± 1.9	0.87 ± 0.1
1EU	2	33.80	21.90	8.20	3.50	0.20
2EU	1	27.70	18.20	6.00	3.10	0.40
3EU	1	15.25	10.41	3.83	0.42	0.59
5EU	5	29.30	17.70	5.20	6.00	0.40
6EU	6	36.10	25.60	7.50	2.80	0.20
8EU	5	15.90	11.30	2.70	1.50	0.40
10EU	5	23.10	18.00	3.60	1.10	0.40
11EU	5	72.64	44.65	18.42	8.27	1.30
12EU	6	64.00	44.90	15.60	2.15	1.35
13EU	5	15.50	11.00	2.70	1.40	0.40
17EU	5	21.20	15.50	3.90	1.50	0.30
18EU	3	39.10	26.20	8.50	3.60	0.80
19EU	4	19.90	17.20	1.60	0.70	0.40
20EU	5	19.20	15.00	3.00	0.80	0.40
21EU	3	19.40	15.50	2.80	0.70	0.40
Mean ± SEM		30.14 ± 4.4	20.87 ± 2.7	6.24 ± 1.2	2.50 ± 0.5	0.53 ± 0.09