Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa

Carlotta De Filippo[a], Duccio Cavalieri[a], Monica Di Paola[b], Matteo Ramazzotti[c], Jean Baptiste Poullet[d], Sebastien Massart[d], Silvia Collini[b], Giuseppe Pieraccini[e], and Paolo Lionetti[b,1]

[a]Department of Preclinical and Clinical Pharmacology, University of Florence, 50139 Firenze, Italy; [b]Department of Pediatrics, Meyer Children Hospital, University of Florence, 50139 Firenze, Italy; [c]Department of Biochemical Sciences, University of Florence, 50134 Firenze, Italy; [d]DNA Vision Agrifood S.A., B-4000 Liège, Belgium; and [e]Centro Interdipartimentale di Spettrometria di Massa, University of Florence, 50139 Firenze, Italy

Edited* by Daniel L. Hartl, Harvard University, Cambridge, MA, and approved June 30, 2010 (received for review April 29, 2010)

Gut microbial composition depends on different dietary habits just as health depends on microbial metabolism, but the association of microbiota with different diets in human populations has not yet been shown. In this work, we compared the fecal microbiota of European children (EU) and that of children from a rural African village of Burkina Faso (BF), where the diet, high in fiber content, is similar to that of early human settlements at the time of the birth of agriculture. By using high-throughput 16S rDNA sequencing and biochemical analyses, we found significant differences in gut microbiota between the two groups. BF children showed a significant enrichment in Bacteroidetes and depletion in Firmicutes ($P < 0.001$), with a unique abundance of bacteria from the genus *Prevotella* and *Xylanibacter*, known to contain a set of bacterial genes for cellulose and xylan hydrolysis, completely lacking in the EU children. In addition, we found significantly more short-chain fatty acids ($P < 0.001$) in BF than in EU children. Also, *Enterobacteriaceae* (*Shigella* and *Escherichia)* were significantly underrepresented in BF than in EU children ($P < 0.05$). We hypothesize that gut microbiota coevolved with the polysaccharide-rich diet of BF individuals, allowing them to maximize energy intake from fibers while also protecting them from inflammations and noninfectious colonic diseases. This study investigates and compares human intestinal microbiota from children characterized by a modern western diet and a rural diet, indicating the importance of preserving this treasure of microbial diversity from ancient rural communities worldwide.

metagenomics | nutrigenomics | biodiversity | 454-pyrosequencing | short-chain fatty acids

The human gut "metagenome" is a complex consortium of trillions of microbes, whose collective genomes contain at least 100 times as many genes as our own eukaryote genome (1). This essential "organ," the microbiome, provides the host with enhanced metabolic capabilities, protection against pathogens, education of the immune system, and modulation of gastrointestinal (GI) development (2).

We do not yet completely understand how the different environments and wide range of diets that modern humans around the world experience has affected the microbial ecology of the human gut.

Contemporary human beings are genetically adapted to the environment in which their ancestors survived and which conditioned their genetic makeup. In mammals, both diet and phylogeny influence the increase in bacterial diversity from carnivore to omnivore to herbivore (3). Dietary habits are considered one of the main factors contributing to the diversity of human gut microbiota (2). Profound changes in diet and lifestyle conditions began with the so-called "Neolithic revolution" with the introduction of agriculture and animal husbandry ≈10,000 y ago (4). After that time, food resources became more abundant and constant, the concentration of large populations in limited areas created selective pressure that favored pathogens specialized in colonizing human hosts and probably produced the first wave of emerging human diseases (5). It has been hypothesized that bacteria specialized in human-associated niches, including our gut commensal flora, underwent intense transformation during the social and demographic changes that took place with the first Neolithic settlements (6).

Western developed countries successfully controlled infectious diseases during the second half of the last century, by improving sanitation and using antibiotics and vaccines. At the same time, a rise in new diseases such as allergic, autoimmune disorders, and inflammatory bowel disease (IBD) both in adults and in children has been observed (5), and it is hypothesized that improvements in hygiene together with decreased microbial exposure in childhood are considered responsible for this increase (7). The GI microflora plays a crucial role in the pathogenesis of IBD (8), and recent studies demonstrate that obesity is associated with imbalance in the normal gut microbiota (9, 10).

The aim of this study was to compare the gut microbiota of children aged 1–6 y living in a village of rural Africa in an environment that still resembles that of Neolithic subsistence farmers with the gut microbiota of western European children of the same age, eating the diet and living in an environment typical of the developed world. These two childhood populations provided an attractive model for assessing the impact of many environmental variables on the gut microbiota.

In our study, we address three general questions regarding the geography and evolution of the human microbiota: (*i*) how is bacterial diversity partitioned within and between the two populations studied; (*ii*) is there a possible correlation between bacterial diversity and diet; and (*iii*) what is the distribution of well-known bacterial pathogens in the two populations, given the different hygienic and geographic conditions?

Results and Discussion

Characterization of Dietary Habits of Children from the Boulpon Rural Village and from Florence, Italy. In this study, we characterized the fecal microbiota of 14 healthy children from the Mossi ethnic

Author contributions: C.D.F., D.C., and P.L. designed research; C.D.F., M.D.P., S.M., and S.C. performed research; G.P. contributed new reagents/analytic tools; M.R. and J.B.P. analyzed data; and C.D.F., D.C., M.D.P., and P.L. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

Data deposition: 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.

[1]To whom correspondence should be addressed. E-mail: paolo.lionetti@unifi.it.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1005963107/-/DCSupplemental.

Fig. 1. Life in a rural village of Burkina Faso. (*A*) Village of Boulpon. (*B*) Traditional Mossi dwelling. (*C*) Map of Burkina Faso (modified from the United States CIA's World Factbook, 34). (*D*) Millet and sorghum (basic components of Mossi diet) grain and flour in typical bowls. (*E*) Millet and sorghum is ground into flour on a grinding stone to produce a thick porridge called Tô.

group (BF) living in the small village of Boulpon in Burkina Faso (Fig. 1) and compared it with that of 15 healthy European children (EU) living in the urban area of Florence, Italy (Table S1). The BF children from Boulpon village were selected as representative consumers of a traditional rural African diet. The diet of BF children is low in fat and animal protein and rich in starch, fiber, and plant polysaccharides, and predominantly vegetarian (Table S2). All food resources are completely produced locally, cultivated and harvested nearby the village by women. The BF diet consists mainly of cereals (millet grain, sorghum), legumes (black-eyed peas, called Niébé), and vegetables, so the content of carbohydrate, fiber and nonanimal protein is very high. Millet and sorghum are ground into flour on a flat stone and made into thick porridge called millet-based Tô, dipped into a sauce made of local vegetables (Néré) and herbs. Although the intake of animal protein is very low, sometimes they eat a small amount of meat (chicken) and termites that we verified to be occasionally part of the BF children's diet in the rainy season.

Children are breast-fed up to the age of 2 y as a complement to a mixed diet. The average amount of fiber in BF diet is 10.0 g/d (2.26%) in 1- to 2-y-old children and 14.2 g/d (3.19%) in 2- to 6-y-old children (Table S2). To represent a Western population (EU), we selected children of the same age who are generally concordant for growth, socially homogeneous and eating the diet and living in an environment typical of the developed world. EU children were breast-fed for up to 1 y of age. They were eating a typical western diet high in animal protein, sugar, starch, and fat and low in fiber. The fiber average content in EU diet is 5.6 g/d (0.67%) in 1- to 2-y-old children and 8.4 g/d (0.9%) in 2- to 6-y-old children (Table S3). The amount of calories (average) consumed varies considerably in the two populations (BF children: 1–2 y old, 672.2 kcal/d; 2–6 y old, 996 kcal/d; EU children: 1–2 y old, 1,068.7 kcal/d; 2–6 y old, 1,512.7 kcal/d; Tables S2 and S3). The isolation of the BF village where the children whom we investigated live, in comparison with the urbanized world, suggests that their diet very likely resembles that of the Neolithic African rural populations following the agriculture revolution.

Dominance of the Bacteroidetes in Gut Microbiota of Burkina Faso Compared with European Children. To characterize the bacterial lineages present in the fecal microbiotas of these 29 children, we performed multiplex pyrosequencing of the V5 and V6 hypervariable regions of 16S rRNA gene with a 454 FLX instrument (Roche). We generated a dataset consisting of 438,219 filtered high-quality, classifiable 16S rRNA gene sequences with a mean average (± SD) of 15,111 ± 3,774 sequences per sample (Table S4). More than 94.2% of the sequences in all of the BF and EU samples were found to belong to the four most populated bacterial phyla, namely Actinobacteria, Bacteroidetes, Firmicutes, and Proteobacteria, in agreement with previous studies describing such phyla as those contributing to the majority of human gut microbiota (2, 11). Relevant differences were found in the proportions of four phyla: Actinobacteria and Bacteroidetes were more represented in BF than in EU children's microbiota (10.1% versus 6.7% and 57.7% versus 22.4%, respectively), whereas Firmicutes and Proteobacteria were more abundant in EU than in BF children (63.7% versus 27.3% and 6.7% versus

0.8%, respectively). The differential distribution of Firmicutes and Bacteroidetes delineates profound differences between the two groups (Fig. S1).

Statistical analysis using a parametric test (ANOVA) indicates that Firmicutes ($P = 7.89 \times 10^{-5}$) and Bacteroidetes ($P = 1.19 \times 10^{-6}$) significantly differentiate the BF from the EU children. This result is strengthened by the nonparametric Kruskal–Wallis test, which again indicated significant discriminating factors in Firmicutes ($P = 3.38 \times 10^{-5}$), Bacteroidetes ($P = 4.80 \times 10^{-4}$), Actinobacteria ($P = 8.82 \times 10^{-3}$), and Spirochaetes ($P = 1.11 \times 10^{-5}$) phyla. Firmicutes are twice as abundant in the EU children as evidenced by the different ratio between Firmicutes and Bacteroidetes (F/B ratio ± SD, 2.8 ± 0.06 in EU and 0.47 ± 0.05 in BF), suggesting a dramatically different bacterial colonization of the human gut in the two populations. Interestingly, *Prevotella*, *Xylanibacter* (Bacteroidetes) and *Treponema* (Spirochaetes) are present exclusively in BF children microbiota (Figs. 2 *A* and *B*, Fig. S2, and Table S5). We can hypothesize that among the environmental factors separating the two populations (diet, sanitation, hygiene, geography, and climate) the presence of these three genera could be a consequence of high fiber intake, maximizing metabolic energy extraction from ingested plant polysaccharides.

Diet plays a central role in shaping the microbiota, as demonstrated by the fact that bacterial species associated with a high-fat, high-sugar diet promote obesity in gnotobiotic mice (12). In such a model, indigenous bacteria maintain energy homeostasis by influencing metabolic processes. The ratio of Firmicutes to Bacteroidetes differs in obese and lean humans, and this proportion decreases with weight loss on low-calorie diet (9). It is therefore reasonable to surmise that the increase in the F/B ratio in EU children, probably driven by their high-calorie diet, might predispose them to future obesity. This F/B ratio may also be considered a useful obesity biomarker.

16S rRNA Gene Surveys Reveal Hierarchical Separation of the Two Pediatric Populations. We further assessed differences in the total bacterial community at the single sample level by clustering the EU and BF samples according to their bacterial genera as found by the RDP classifier (Ribosomal Database Project v. 2.1).

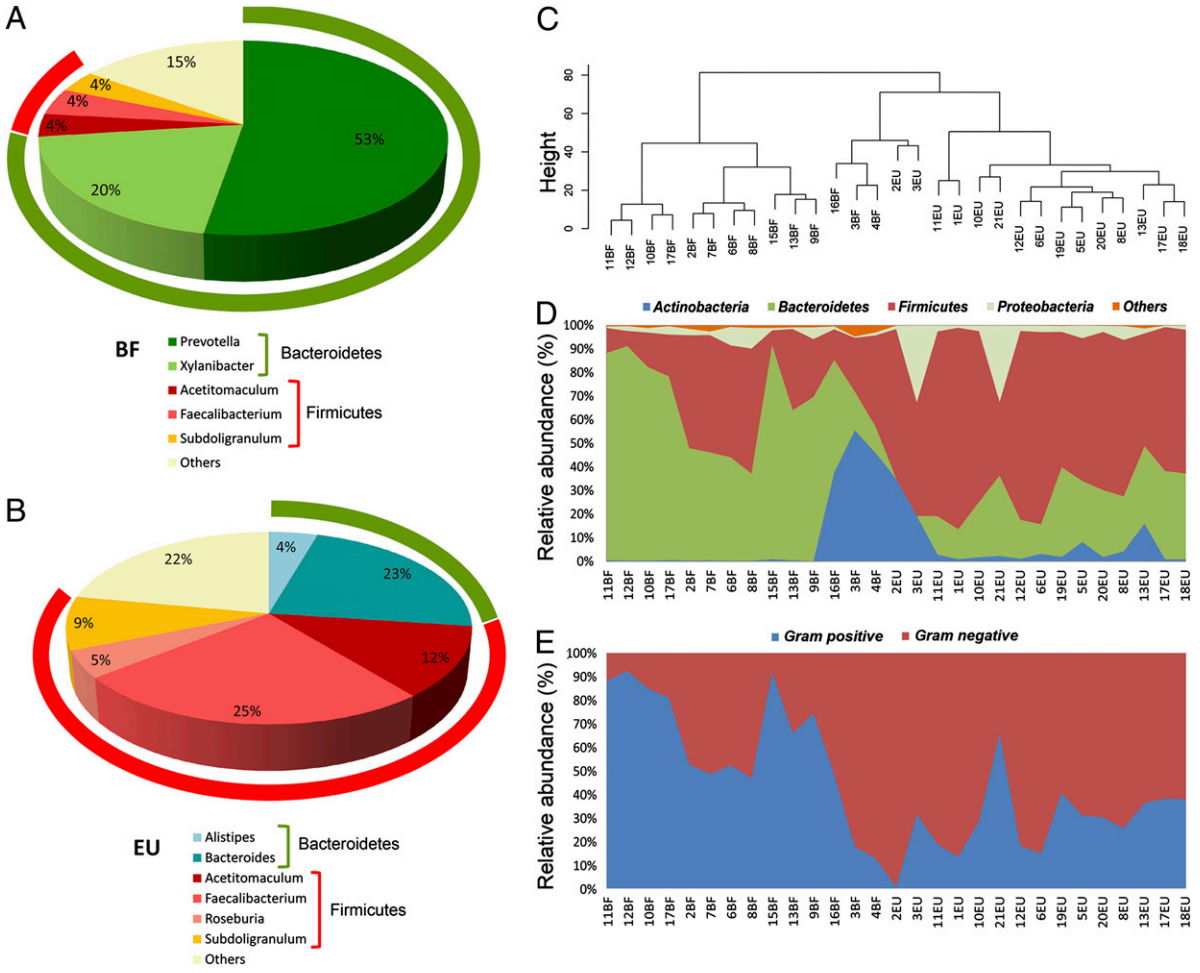

Fig. 2. 16S rRNA gene surveys reveal a clear separation of two children populations investigated. (*A* and *B*) Pie charts of median values of bacterial genera present in fecal samples of BF and EU children (>3%) found by RDP classifier v. 2.1. Rings represent corresponding phylum (Bacteroidetes in green and Firmicutes in red) for each of the most frequently represented genera. (*C*) Dendrogram obtained with complete linkage hierarchical clustering of the samples from BF and EU populations based on their genera. The subcluster located in the middle of the tree contains samples taken from the three youngest (1–2 y old) children of the BF group (16BF, 3BF, and 4BF) and two 1-y-old children of the EU group (2EU and 3EU). (*D*) Relative abundances (percentage of sequences) of the four most abundant bacterial phyla in each individual among the BF and EU children. Blue area in middle shows abundance of Actinobacteria, mainly represented by *Bifidobacterium* genus, in the five youngest EU and BF children. (*E*) Relative abundance (percentage of sequences) of Gram-negative and Gram-positive bacteria in each individual. Different distributions of Gram-negative and Gram-positive in the BF and EU populations reflect differences in the two most represented phyla, Bacteroidetes and Firmicutes.

Complete linkage hierarchical clustering produced a net separation of BF and EU populations (Fig. 2*C*). It is noteworthy that the subcluster that joins the two major clusters (located in the middle of the tree) contains samples taken from the five youngest EU and BF children (1–2 y old). This can be explained by the fact that BF children are breast-fed up to the age of 2 y and resemble the younger EU children who were breast fed-up to 1 y of age. It is also noteworthy that only in these three BF subjects we observed abundant Actinobacteria, mainly represented by *Bifidobacterium* genus (Fig. 2*D*), that were found in all EU subjects (Table S5) and that is known to be strictly related to breast-feeding in infants (13). This result provides a clear indication of the dominant role of diet over the variables mentioned above in shaping the microbial composition of the gut. Studies on the genetic variability between Mossi and Europeans showed polymorphisms in the major histocompatibility complex (MHC) genes (14), but no significant differences in the expression of key genes regulating immune function such as *TGFβ, TGFβR, CTLA4,* and *FOXP3,* suggesting a functional similarity (15). Also, the two populations are different for many other variables such as sanitation, hygiene, geography, and climate. Yet, if any of these variables had prevailed over diet, these five children would have fallen into two main clusters instead of creating a third, significantly separated, cluster.

In western populations, the human intestinal microbiota undergoes maturation from birth to adulthood and aging, with particular emphasis on the F/B ratio, that evolves during different life stages (16). Our results can be explained by the fact that, as soon as breast-feeding is substituted by solid foods, the differences in microbiota between the two populations increase, reflecting the dietary and environmental separation that results in a differentiation between the F/B ratio. As a consequence of the different F/B ratio in the two populations, we found Gram-negative bacteria (mainly Bacteroidetes) more abundant (58.5%) than Gram-positive bacteria (37.4%) in the BF population, whereas Gram-positive (mainly Firmicutes) were more abundant than Gram-negative bacteria (70.4% versus 29.1% respectively) in the EU population (Fig. 2*E*).

Microbial Richness and Biodiversity. We then compared the microbial richness, estimated by the Chao1 index, and the biodiversity, assessed by a nonparametric Shannon index for the two BF and EU groups. In our calculations we took into account different OTU distance unit cutoffs, namely 0.03, 0.05, and 0.10 (Fig. S3). Using the nonparametric Kruskal–Wallis test for comparisons, we found significant differences ($P < 0.01$) in both richness and biodiversity between BF and EU samples at the Operational Taxonomic Unit (OTU) cutoff 0.10, with a higher microbial richness and biodiversity in BF samples than in EU samples (Table S6).

Exposure to the large variety of environmental microbes associated with a high-fiber diet could increase the potentially beneficial bacterial genomes, enriching the microbiome. Reduction in microbial richness is possibly one of the undesirable effects of globalization and of eating generic, nutrient-rich, uncontaminated foods. Both in the Western world and in developing countries diets rich in fat, protein, and sugar, together with reduced intake of unabsorbable fibers, are associated with a rapid increase in the incidence of noninfectious intestinal diseases. The potential protective effects of the diet on bowel disorders was first described by Burkitt (17) who, working in Africa in the 1960s, noticed the remarkable absence of noninfectious colonic diseases in Africans consuming a traditional diet rich in fiber.

***Xylanibacter, Prevotella, Butyrivibrio, and Treponema* Genera May Enhance the Ability to Extract Calories from Indigestible Polysaccharides in BF Children.** Whole grains are concentrated sources of dietary fiber, resistant starch, and oligosaccharides, as well as carbohydrates that escape digestion in the small intestine and are fermented in the gut, producing short-chain fatty acids (SCFAs). *Xylanibacter*, *Prevotella, Butyrivibrio,* and *Treponema* are exclusive to the BF children (Fig. S2) and indicate the presence of a bacterial community using xylane, xylose, and carboxymethylcellulose to produce high levels of SCFAs (18) whose protective role against gut inflammation has been well proven (19). These bacteria can ferment both xylan and cellulose through carbohydrate-active enzymes such as xylanase, carboxymethylcellulase, and endoglucanase (http://www.cazy.org).

Other SCFA-producing bacteria, such as *Bacteroides* and *Faecalibacterium* species, particularly *F. prausnitzii* (Table S5), found in both populations, could generally indicate the importance of maintaining a microflora with potentially anti-inflammatory capability (20).

To associate the presence of SCFA-producing bacterial communities with the effective increase in the concentration of SCFAs in fecal samples, we determined the levels of acetic, propionic, butyric, and valeric acids using solid phase microextraction–gas chromatography–mass spectrometry (SPME-GC-MS) analysis. It is

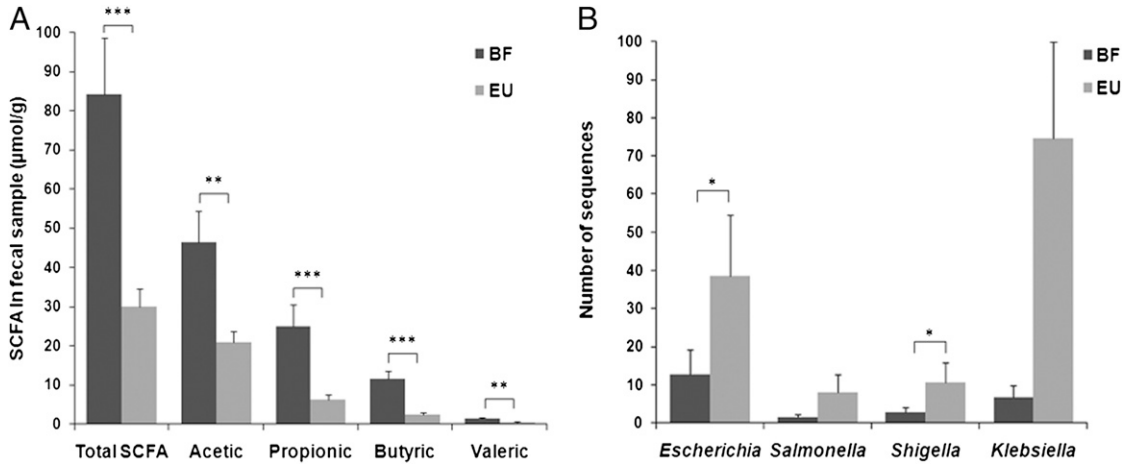

Fig. 3. SCFA-producing bacteria could help to prevent establishment of some potentially pathogenic intestinal bacteria. (*A*) Quantification of SCFAs in fecal samples from BF and EU populations by SPME-GC-MS. (*B*) Number of sequences relative to principal *Enterobacteriaceae* genera, in BF and EU children microbiota. Mean values (±SEM) are plotted. Asterisks indicate significant differences (one-tailed Student *t* test of all data points: $*P < 0.05$; $**P \leq 0.01$; $***P \leq 0.001$).

noteworthy that in BF children we found a significantly higher amount of total SCFAs compared with EU children (one-tailed Student t test, $P = 4.5 \times 10^{-4}$; Fig. 3*A*). In particular, propionic and butyric acids are nearly four times more abundant in BF than in EU fecal samples (one-tailed Student t test, $P = 1.3 \times 10^{-3}$ and $P = 1.6 \times 10^{-4}$, respectively), whereas acetic and valeric acids were comparable in both groups (one-tailed Student's t test, respectively $P = 2.0 \times 10^{-3}$ and $P = 2.4 \times 10^{-3}$ (Fig. 3*A* and Table S7). Normal colonic epithelia derive 60–70% of their energy supply from SCFAs, particularly butyrate (21). Propionate is largely taken up by the liver and is a good precursor for gluconeogenesis, liponeogenesis, and protein synthesis (22). Acetate enters the peripheral circulation to be metabolized by peripheral tissues and is a substrate for cholesterol synthesis (23). Previous analyses on the physiological significance of SCFA (24) showed how SCFA are rapidly absorbed from the colon. Therefore, an abundance of SCFA in the feces indicates production of SCFA from microflora at levels far above the absorption rate. Our results allow us to hypothesize that a diet rich in plant polysaccharides and low in sugar and fat could select SCFA-producing bacteria.

Altogether, our results indicate a correlation between polysaccharide-degrading microbiota and the calories that the host can extract from his/her diet, potentially influencing the survival and fitness of the host. We can hypothesize that microbiota coevolved with the diet of BF individuals, allowing them to maximize the energy intake from indigestible components, such as plant polysaccharides, by producing high levels of SCFAs that supply the host with an additional amount of energy. Given that enhanced ability to obtain energy-rich food is considered to be one factor that has driven human evolution. Substantial microbiota adaptation has probably accompanied the dietary changes that have occurred throughout human history. In fact it is well known that changes in food production agricultural and preparation have profoundly influenced the intestinal microflora.

Our results suggest that diet has a dominant role over other possible variables such as ethnicity, sanitation, hygiene, geography, and climate, in shaping the gut microbiota. We can hypothesize that the reduction in richness we observe in EU compared with BF children, could indicate how the consumption of sugar, animal fat, and calorie-dense foods in industrialized countries is rapidly limiting the adaptive potential of the microbiota. This microbial simplification harbors the risk of depriving our microbial gene pool of potentially useful environmental gene reservoirs that allow adaptation to peculiar diets, as we observed in BF population and as recently shown by diet-induced horizontal gene transfer in Japanese individuals consuming algae in their diet (25).

Gut microbial richness could have several health-related effects. The SCFA-producing bacteria that are abundant in the BF children's gut possibly help to prevent the establishment of some potentially pathogenic intestinal microbes (26) causing diarrhea, as seen by the fact that *Enterobacteriaceae*, such as *Shigella* and *Escherichia*, were significantly underrepresented in BF than in EU children ($P < 0.05$, one-tailed t test; Fig. 3*B*). Increased gut microbial diversity and reduced quantities of potentially pathogenic strains in BF would agree with the "old friend" hypothesis, indicating a role of microbiota in protecting children from pathogens as well as from gastrointestinal diseases (27).

The lessons learned from the BF children's microbiota prove the importance of sampling and preserving microbial biodiversity from regions where the effects of globalization on diet are less profound. The worldwide diversity of the microbiome from ancient communities, where gastrointestinal infections can make the difference between life and death, represents a goldmine for studies aimed at elucidating the role of gut microbiota on the subtle balance between health and disease and for the development of novel probiotics.

Materials and Methods

Population Enrollment, Fecal Sample Collection, and DNA Extraction. We enrolled 15 healthy children (nine male and six female) living in the rural village of Boulpon district of Nanoro, Boulkiemde province, Burkina Faso, and 15 healthy children (nine male and six female) living in the urban area of Florence, Italy. All children were 1–6 y of age, had not taken antibiotics or probiotics in the 6 mo before the sampling dates, and had not been hospitalized in the previous 6 mo (Table S1). A detailed medical and lifestyle report was obtained from EU children's parents, and a 3-d dietary questionnaire and an in-depth interview on BF children's diet was obtained directly from their mothers.

Despite the high incidence of infectious disease, including malaria and malnutrition in the area, all children were apparently healthy at the time of sample collection. Upper midarm measurement excluded both severe and moderate malnutrition. As representative of a healthy Western population (EU), we selected children of the same age who were generally concordant for growth, socially homogeneous, and eating the diet and living in an environment typical of the developed world. Fecal samples were collected by physicians and preserved in RNAlater (Qiagen) at −80 °C until extraction of genomic DNA (28) (details in *SI Materials and Methods*).

Sequencing of 16S rRNA Gene Amplicons. For each sample, we amplified 16S rRNA genes using a primer set specific for V5 and V6 hypervariable 16S RNA region. The forward primer contained the sequence of the Titanium A adaptor and a barcode sequence. Pyrosequencing was carried out using primer A on a 454 Life Sciences Genome Sequencer FLX instrument (Roche) following Titanium chemistry (details in *SI Materials and Methods*). 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.

Taxonomic Assignment to 16S Reads. RDP classifier (v 2.1) software was used (29) to classify the sequences according to the taxonomy proposed by Garrity et al. (30), maintained at the Ribosomal Database Project (RDP 10 database, Update 18). RDP classifier also emits, for each taxonomic rank, a confidence estimate (CE) based on a bootstrapping procedure, allowing to append the notation of "_uncertain" to assignments with CE lower than a defined cutoff , usually 50% (Table S4). Bacterial species were assigned using a speed-optimized procedure based on BLAST and on the creation of genus specific subsamples of the RDP 10 database (details in *SI Materials and Methods*).

Quantifying and Comparing Diversity Between BF and EU Populations. Differences between populations have been analyzed using parametric (ANOVA) and nonparametric (Kruskal-Wallis test) statistical methods. Even if, in principle, multivariate ANOVA would be more appropriate to catch the whole information available from such dataset, some of the assumptions (e.g., normality of residuals) were not met, as tested with Shapiro–Wilk W test and the energy E test. We then preferred to use univariate methods such as ANOVA and the nonparametric, rank-driven Kruskall-Wallis test, which performs well in the absence of distributional assumptions (details in *SI Materials and Methods*).

Complete Linkage Hierarchical Clustering. The clustering of EU and BF samples was performed on genera obtained from RDP Classifier by means of a complete linkage hierarchical clustering technique using the R package hclust (details in *SI Materials and Methods*).

Richness and Diversity Index. To obtain the matrix containing pairwise sequence distances, all reads were first aligned with muscle v3.7 (31) and converted to Phylip format for downstream calculations. Richness and biodiversity indices were obtained with the Mothur software package (32). For richness estimation, related to the number of observed operational taxonomic units (OTUs), we used the Chao1 index. Biodiversity that depends how uniformly the sequences are spread into the different observed OTUs, was instead estimated with the nonparametric Shannon formula (33). Both indexes were evaluated at a different distance unit cutoff, to test different selectivity in the definition of OTUs.

Determination of SCFAs in Fecal Samples. For determination of SCFAs we used 250 mg frozen fecal samples. Concentrations of SCFAs were determined in a 1:25 dilution of 500 μL supernatant. SPME-GC-MS determinations were performed using a Varian Saturn 2000 GC-MS instrument with 8200 CX SPME autosampler (details in *SI Materials and Methods*).

ACKNOWLEDGMENTS. We thank the volunteers for their participation, G. Capponi from St. Camille Hospital and Solange Zemba from Nanoro (Burkina Faso), L. Turbanti (Centro Interdipartimentale di Spettrometria di Massa) and A.P. Femia (Department of Pharmacology) at the University of Florence, and P. Rocca Serra (EMBL-EBI) for assistance. This work was supported by Ministero dell'Istruzione, dell'Università e della Ricerca, Italy Grant PRIN 2007-N352CP_001, Ente Cassa di Risparmio di Firenze Grant 0875, and Meyer's Children Hospital.

1. Gill SR, et al. (2006) Metagenomic analysis of the human distal gut microbiome. *Science* 312:1355–1359.
2. Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI (2005) Host-bacterial mutualism in the human intestine. *Science* 307:1915–1920.
3. Ley RE, et al. (2008) Evolution of mammals and their gut microbes. *Science* 320: 1647–1651.
4. Cordain L, et al. (2005) Origins and evolution of the Western diet: Health implications for the 21st century. *Am J Clin Nutr* 81:341–354.
5. Blaser MJ (2006) Who are we? Indigenous microbes and the ecology of human diseases. *EMBO Rep* 7:956–960.
6. Mira A, Pushker R, Rodríguez-Valera F (2006) The Neolithic revolution of bacterial genomes. *Trends Microbiol* 14:200–206.
7. Strachan DP (1989) Hay fever, hygiene, and household size. *BMJ* 299:1259–1260.
8. Peterson DA, Frank DN, Pace NR, Gordon JI (2008) Metagenomic approaches for defining the pathogenesis of inflammatory bowel diseases. *Cell Host Microbe* 3: 417–427.
9. Ley RE, Turnbaugh PJ, Klein S, Gordon JI (2006) Microbial ecology: Human gut microbes associated with obesity. *Nature* 444:1022–1023.
10. Turnbaugh PJ, et al. (2009) A core gut microbiome in obese and lean twins. *Nature* 457:480–484.
11. Qin J, et al.; MetaHIT Consortium (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464:59–65.
12. Turnbaugh PJ, et al. (2009) The effect of diet on the human gut microbiome: A metagenomic analysis in humanized gnotobiotic mice. *Sci Transl Med* 1:6ra14.
13. Gueimonde M, Laitinen K, Salminen S, Isolauri E (2007) Breast milk: A source of bifidobacteria for infant gut development and maturation? *Neonatology* 92:64–66.
14. Lulli P, et al. (2009) HLA-DRB1 and -DQB1 loci in three west African ethnic groups: Genetic relationship with sub-Saharan African and European populations. *Hum Immunol* 70:903–909.
15. Torcia MG, et al. (2008) Functional deficit of T regulatory cells in Fulani, an ethnic group with low susceptibility to Plasmodium falciparum malaria. *Proc Natl Acad Sci USA* 105:646–651.
16. Mariat D, et al. (2009) The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age. *BMC Microbiol* 9:123.
17. Burkitt DP (1973) Epidemiology of large bowel disease: The role of fibre. *Proc Nutr Soc* 32:145–149.
18. Flint HJ, Bayer EA, Rincon MT, Lamed R, White BA (2008) Polysaccharide utilization by gut bacteria: Potential for new insights from genomic analysis. *Nat Rev Microbiol* 6: 121–131.
19. Scheppach W, Weiler F (2004) The butyrate story: Old wine in new bottles? *Curr Opin Clin Nutr Metab Care* 7:563–567.
20. Sokol H, et al. (2008) *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci USA* 105:16731–16736.
21. Scheppach W (1994) Effects of short chain fatty acids on gut morphology and function. *Gut* 35(1, Suppl):S35–S38.
22. Wolever TM, Spadafora P, Eshuis H (1991) Interaction between colonic acetate and propionate in humans. *Am J Clin Nutr* 53:681–687.
23. Wolever TM, Brighenti F, Royall D, Jenkins AL, Jenkins DJ (1989) Effect of rectal infusion of short chain fatty acids in human subjects. *Am J Gastroenterol* 84: 1027–1033.
24. Cummings JH, Pomare EW, Branch WJ, Naylor CP, Macfarlane GT (1987) Short chain fatty acids in human large intestine, portal, hepatic and venous blood. *Gut* 28: 1221–1227.
25. Hehemann JH, et al. (2010) Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* 464:908–912.
26. Hermes RG, et al. (2009) Effect of dietary level of protein and fiber on the productive performance and health status of piglets. *J Anim Sci* 87:3569–3577.
27. Rook GAW, Brunet LR (2005) Microbes, immunoregulation, and the gut. *Gut* 54: 317–320.
28. Zoetendal EG, et al. (2006) Isolation of RNA from bacterial samples of the human gastrointestinal tract. *Nat Protoc* 1:954–959.
29. Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73:5261–5267.
30. Garrity GM, et al. (2007) *The Taxonomic Outline of Bacteria and Archaea. TOBA Release 7.7,* Michigan State University Board of Trustees. Available at http://www.taxonomicoutline.org/index.php/toba/index.
31. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.
32. Schloss PD, et al. (2009) Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537–7541.
33. Chao A, Shen TJ (2003) Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. *Environ Ecol Stat* 10:429–443.
34. Central Intelligence Agency, The world Factbook. Available at https://www.cia.gov/library/publications/the-world-factbook/index.html.

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 × *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.

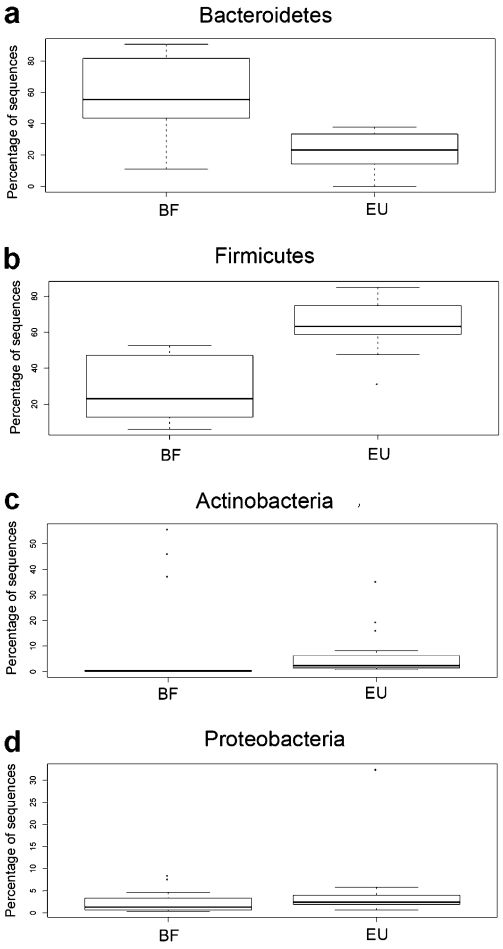

Fig. S1. Boxplots (percentage of sequences) of the four most represented phyla. Bacteroidetes (*A*), Firmicutes (*B*), Actinobacteria (*C*), and Proteobacteria (*D*) in the BF and EU children.

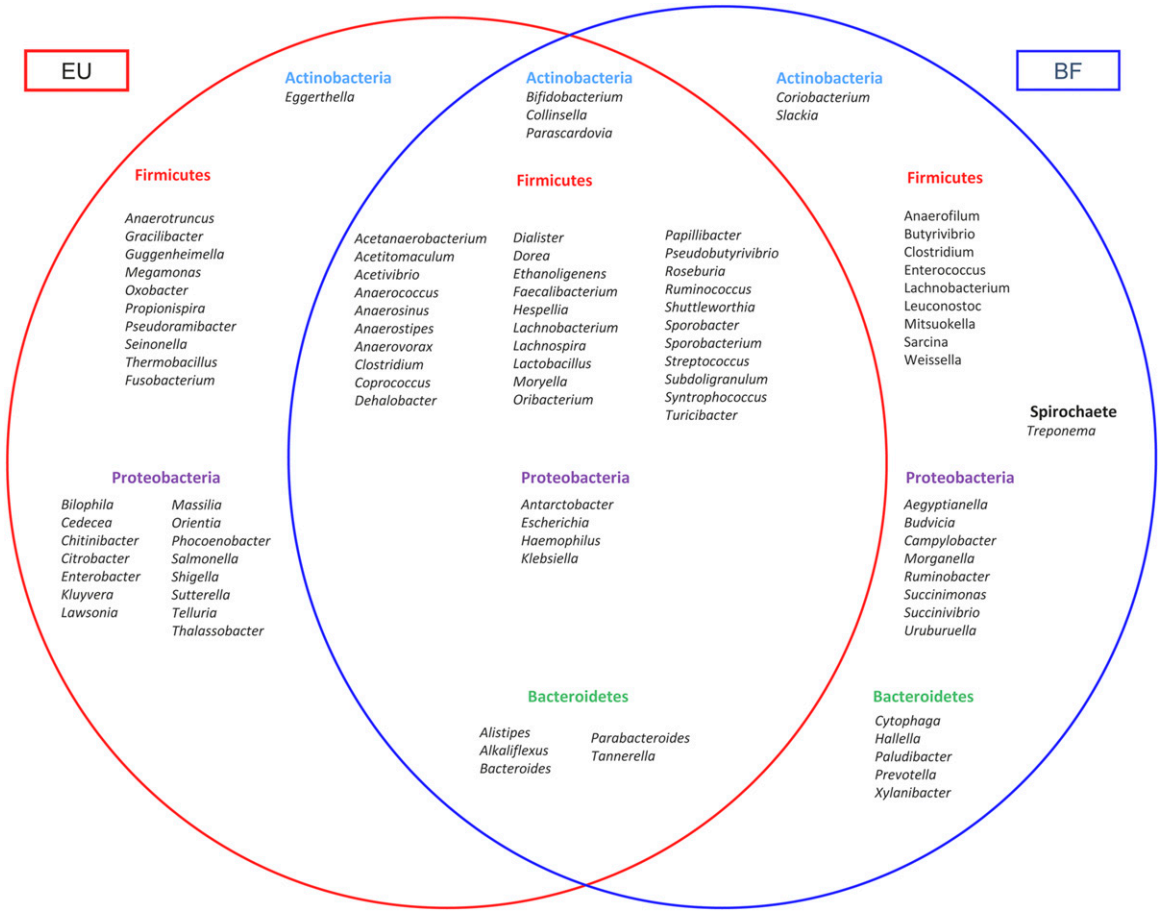

Fig. S2. Venn diagram showing the number of genera belonging to the major phyla differing significantly between BF and EU children. A total of 26 genera are characteristic of EU children, 43 genera are in common between BF and EU children, and 25 genera are characteristic of BF children.

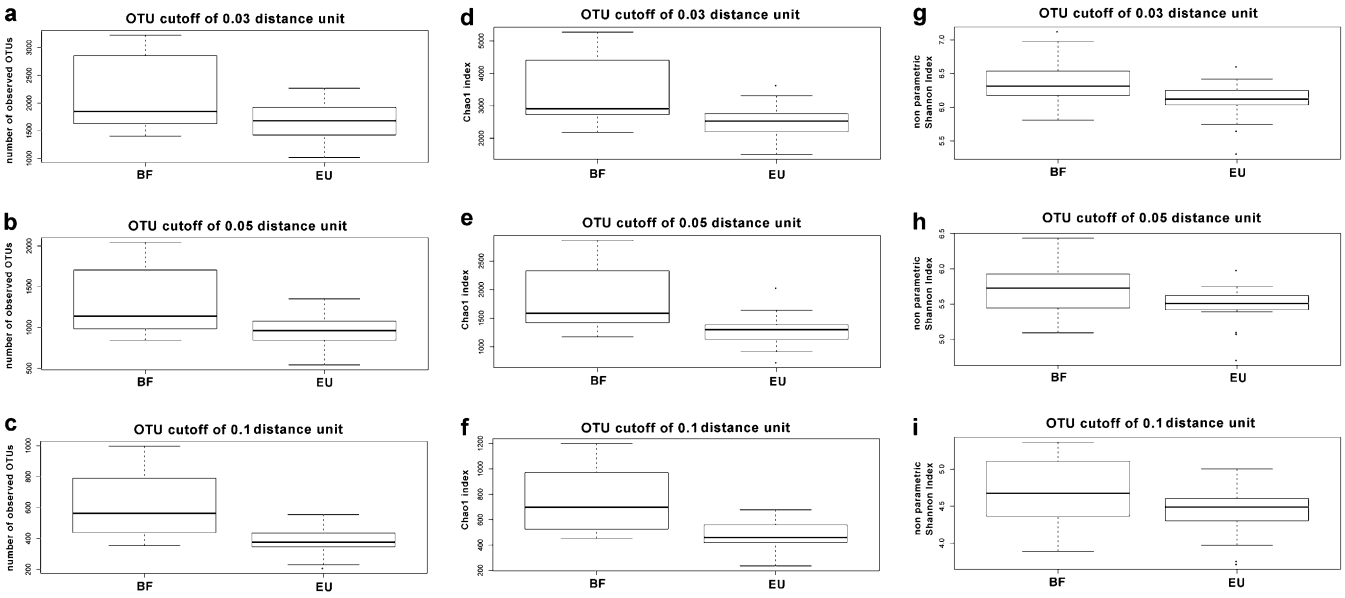

Fig. S3. (*A*–*C*) Boxplots (percentage of sequences) of the observed OTUs at OTU cutoffs of (*A*) 0.03, (*B*) 0.05, and (*C*) 0.10 distance units in the BF and EU children. (*D*–*F*) Boxplots (percentage of sequences) of the Chao1 indexes at OTU cutoffs of (*D*) 0.03, (*E*) 0.05, and (*F*) 0.10 distance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

		SCFAs (μmol/g feces)				
Subject ID	Age (y)	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