Supporting information

Method of dietary information collected during the first year of life

Information on the child’s diet was collected weekly throughout the first year of life on whether the child consumed any of the following 23 categories of food during the last week: milk, butter, yoghurt, other dairy products, grains with and without gluten, 4 categories of vegetables (white roots and tubers, Vit A rich vegetables and tubers, dark green leafy vegetables and other vegetables), 2 categories of fruits (Vit A rich fruits and other fruits), meat, eggs, fish or seafood, peanut, almond, other nuts, soy products, legumes, oils, sweets and spices, which was based on diet diversity classification developed by the Food and Agriculture Organization of the United Nations with some adjustment (<http://www.fao.org/3/i1983e/i1983e.pdf>).

A-1

A-2

A-3



B-3

B-2

B-1



Supplementary figure 1. Principal coordinates analysis (PCoA) plot based on A) weighted UniFrac distances (which takes relative abundance into account) and B) unweighted UniFrac distances among samples from each 90 days (1), 180 days (2) and 360 days (3) analysed using 16S rRNA gene sequencing among n=51, 56 and 61 samples, respectively. Each point represents the sample, and the clustering of the points describe how similar the communities are between the samples. The clustering of samples changed over time similarly with no clear separation between AD group and non-AD group at each time point.

Supplementary table 1. Quantitative PCR assay. Target group and gene, primers used for qPCR analysis, correction for multiple 16S rRNA gene copies and upper and lower limit of quantification of each primer set used.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Targeted group  | Target gene  | Primer  | Sequence (5'-3') | Correction for multiple 16S rRNA gene copies | Limit of quantification log cells/g feces | Reference |
| Upper | Lower |
| All bacteria | 16S rRNA gene | Eub338 | ACWCCTACGGGWGGCAGCAG  | 5 | 11 | 3.2 | (1) |
| Eub518 | ATTACCGCGGCTGG  |
| *Faecalibacterium prausnitzii* | *butCoAT* | G\_Fprsn\_F | GACAAGGGCCGTCAGGTC | 1 | 8 | 2.4 | (2) |
|  | G\_Fprsn\_R | GGACGAGATRAAGCT |
| *Eubacterium rectale/ Roseburia spp* | *butCoAT* | G\_RosEub\_F | TCAAATCMGG5GACTGGGTWGA | 1 | 9.1 | 2.4 | (2) |
|  | G\_Ros\_R | TCGATACCGGACATATGCCAKGAG  |
|  | G\_Eub\_R | TCATAACCGCCCATATGCCATGAG |
| *Anaerobutyricum hallii*  | *pduC* | pduF | CGTTATGCTCCATTTAATGCT | 1 | 9.5 | 2.4 | (3) |
|  | pduR | CCAAGGAGTATCATCACCATC |
| *Clostridium* sensu stricto | *buk* | Gbuk-F | TGCTGTWGTTGGW AGAGGYGGA | 1 | 9.8 | 2.2 | (2) |
|   | Gbuk-R | GCAACIGCYTTTTGATTTAATGCATGG |

Supplementary table 2. Comparison of beta diversity using different distance metrics between AD and non-AD group by each time point

|  |  |  |  |
| --- | --- | --- | --- |
| **Time point** | **Diversity Metric** | **R2** | **P value** |
| 90 days | Bray Curtis | 0.02476 | 0.23 |
| Jaccard | 0.023655 | 0.142 |
| Weighted Unifrac | 0.022382 | 0.327 |
| Unweighted Unifrac | 0.029558 | 0.089 |
| 180 days | Bray Curtis | 0.015985 | 0.55 |
| Jaccard | 0.017899 | 0.516 |
| Weighted Unifrac | 0.005138 | 0.942 |
| Unweighted Unifrac | 0.019266 | 0.373 |
| 360 days | Bray Curtis | 0.014387 | 0.64 |
| Jaccard | 0.020492 | 0.057 |
| Weighted Unifrac | 0.006642 | 0.906 |
| Unweighted Unifrac | 0.018343 | 0.278 |

Different distance metrics were used to analyze the difference in beta diversity between AD and non-AD group based on 16S rRNA gene sequencing data among n=51, 56 and 61 samples at 90, 180 and 360 days, respectively.

****

Supplementary figure 2. Relative abundance of major bacterial families in infants without AD (A) and with AD up to one year (B). Faecal microbiota composition at 90, 180 and 360 days was determined using 16S rRNA gene sequencing among n=51, 56 and 61 samples, respectively. Shown are the mean relative abundance of bacterial families in feces.

Supplementary table 3. Numbers of observed species in children with and without AD in the first year of life at each time point

|  |  |  |  |
| --- | --- | --- | --- |
|   | 90 days  | 180 days  | 360 days  |
|   | non-AD (n=28) | AD (n=22) |   | non-AD (n=34) | AD (n=19) |   | non-AD (n=31) | AD (n=25) |   |
|   | median | IQR | median | IQR | p-value  | median | IQR | median | IQR | p-value  | median | IQR | median | IQR | p-value  |
| **Total microbiota**  | 45 | 41.5-49.5 | 39.5 | 27-49 | 0.02 | 59.5 | 51-68 | 47 | 42-63 | 0.14 | 84 | 68-102 | 72 | 64-87 | 0.13 |
| **Phylum level** |   |  |  |  |   |   |  |  |  |   |   |  |  |  |   |
|  **Firmicutes** | 26 | 21-32 | 21 | 12-28 | 0.07 | 37 | 30-43 | 27 | 23-44 | 0.17 | 55 | 46-72 | 49 | 43-65 | 0.13 |
|  **Bacteroidetes** | 5.5 | 3-8.5 | 3.5 | 1-6 | 0.09 | 5 | 3-8 | 5 | 2-8 | 0.40 | 8 | 5-10 | 6 | 4-10 | 0.36 |
|  **Actinobacteria** | 10 | 6-14 | 8.5 | 6-12 | 0.65 | 10 | 8 -14 | 11 | 7-14 | 0.90 | 11 | 9-14 | 13 | 9-14 | 0.85 |
| **Firmicutes family level** |   |  |  |  |   |   |  |  |  |   |   |  |  |  |   |
|  ***Ruminococcaceae*** | 1 | 0-2 | 0.5 | 0-2 | 0.70 | 2 | 1-4 | 1 | 0-3 | 0.08 | 6 | 4-10 | 5 | 3-8 | 0.13 |
|  ***Peptostreptococcaceae***  | 1.5 | 1-3 | 0.5 | 0-2 | 0.10 | 3 | 2-4 | 2 | 1-3 | 0.14 | 4 | 2-5 | 4 | 3-4 | 0.65 |
|  ***Lachnospiraceae*** | 6 | 3-9.5 | 5.5 | 3-9 | 0.49 | 8.5 | 5-15 | 9 | 3-13 | 0.28 | 22 | 17-32 | 19 | 11-29 | 0.10 |
|  ***Clostridiaceae***  | 4 | 1-6 | 3 | 1-5 | 0.40 | 5 | 3-7 | 4 | 3-5 | 0.18 | 6 | 4-7 | 5 | 3-7 | 0.76 |

Faecal microbiota composition was determined using 16S rRNA gene sequencing and samples with more than 15,000 reads were included in this analysis (n=50, 53 and 56 at 90, 180 and 360 days, respectively). P-values are derived by Mann-Whitney U test

Supplementary table 4. Relative abundance of the major bacterial families in children without AD and with AD during the first year of life at each time point

|  |  |  |  |
| --- | --- | --- | --- |
|  | **90 days** | **180 days** | **360 days** |
|  | **non-AD (n=29)** | **AD (n=22)** |  | **non-AD (n=36)** | **AD (n=20)** |  | **non-AD (n=35)** | **AD (n=26)** |  |
| **Major bacterial families** | med | IQR | med | IQR | p-value | med | IQR | med | IQR | p-value | med | IQR | med | IQR | p-value |
| *Bifidobacteriaceae* | 72.76 | 63.67-89.01 | 71.65 | 52.61-88.2 | 0.49 | 61.53 | 33.09-78.3 | 48.68 | 30.35-80.75 | 0.70 | 27.94 | 11.02-47.66 | 30.11 | 10.81-40.81 | 0.71 |
| *Enterobacteriaceae* | 2.51 | 0.36-11.77 | 3.27 | 1.81-16.45 | 0.43 | 5.24 | 2.12-11.05 | 6.35 | 2.08-12.23 | 0.77 | 0.91 | 0.34-2.12 | 1.20 | 0.51-3.45 | 0.55 |
| *Enterococcaceae* | 0.74 | 0.15-2.81 | 0.25 | 0.13-1.95 | 0.41 | 0.79 | 0.18-2.32 | 0.81 | 0.18-1.75 | 0.90 | 0.19 | 0.08-0.34 | 0.35 | 0.1-0.71 | 0.10 |
| *Streptococcaceae* | 0.64 | 0.22-1.33 | 0.48 | 0.22-0.85 | 0.47 | 0.82 | 0.2-3.76 | 0.84 | 0.44-1.32 | 0.85 | 0.89 | 0.36-2.11 | 0.62 | 0.27-2.08 | 0.69 |
| *Coriobacteriaceae/Actinomycetaceae* | 0.45 | 0.06-2.65 | 0.09 | 0.04-0.36 | 0.21 | 0.23 | 0.07-2.69 | 0.31 | 0.08-1.2 | 0.74 | 0.11 | 0.05-0.97 | 0.10 | 0.02-0.22 | 0.19 |
| *Veillonellaceae* | 0.35 | 0.07-0.8 | 0.22 | 0.03-1.27 | 0.37 | 1.48 | 0.34-6.51 | 2.65 | 0.64-6.08 | 0.43 | 1.29 | 0.4-3.9 | 1.41 | 0.53-4.76 | 0.89 |
| *Bacteroidaceae* | 0.29 | 0.08-4.69 | 0.08 | 0.02-6.01 | 0.25 | 0.51 | 0.05-6.04 | 1.44 | 0.07-4.72 | 0.56 | 2.75 | 0.61-17.83 | 3.36 | 0.15-16.35 | 0.65 |
| *Clostridiaceae*  | 0.23 | 0.07-0.94 | 0.20 | 0.02-1.13 | 0.90 | 0.65 | 0.18-2.79 | 0.35 | 0.16-2.71 | 0.49 | 0.76 | 0.29-1.51 | 0.46 | 0.23-1.07 | 0.37 |
| *Lachnospiraceae* | 0.20 | 0.13-0.71 | 0.17 | 0.07-0.91 | 0.50 | 2.67 | 0.17-7.49 | 0.50 | 0.14-5.5 | 0.33 | 22.10 | 15.88-32.3 | 36.34 | 10.49-56.47 | 0.20 |
| *Others*  | 0.16 | 0.09-0.4 | 0.15 | 0.07-0.22 | 0.40 | 0.11 | 0.04-0.32 | 0.12 | 0.06-0.31 | 0.66 | 0.12 | 0.02-0.41 | 0.09 | 0.02-0.17 | 0.73 |
| *Erysipelotrichaceae*  | 0.09 | 0.03-0.15 | 0.00 | 0-0.06 | 0.01 | 0.21 | 0.03-0.54 | 0.09 | 0.02-0.52 | 0.80 | 1.04 | 0.4-1.97 | 1.07 | 0.69-1.93 | 0.89 |
| *Peptostreptococcaceae*  | 0.07 | 0.02-0.25 | 0.00 | 0-0.08 | 0.10 | 0.44 | 0.06-0.85 | 0.08 | 0.02-0.92 | 0.14 | 0.73 | 0.2-2.55 | 0.60 | 0.29-2.52 | 0.81 |
| *Eggerthellaceae* | 0.03 | 0-0.2 | 0.01 | 0-0.23 | 0.96 | 0.10 | 0-0.22 | 0.05 | 0-0.34 | 0.96 | 0.26 | 0.09-0.61 | 0.13 | 0.05-0.28 | 0.06 |
| *Lactobacillaceae* | 0.02 | 0-0.39 | 0.19 | 0-1.9 | 0.29 | 0.05 | 0-1.31 | 0.20 | 0-0.63 | 0.65 | 0.01 | 0-0.34 | 0.07 | 0-0.48 | 0.45 |
| *Bacillaceae/Paenibacillaceae* | 0.02 | 0-0.05 | 0.00 | 0-0.02 | 0.11 | 0.01 | 0-0.04 | 0.02 | 0.01-0.06 | 0.15 | 0.03 | 0.01-0.06 | 0.03 | 0.01-0.06 | 0.73 |
| *Ruminococcaceae* | 0.01 | 0-0.05 | 0.00 | 0-0.04 | 0.77 | 0.08 | 0.02-0.33 | 0.01 | 0-0.04 | 0.03 | 2.36 | 0.67-10.67 | 1.49 | 0.35-4.38 | 0.21 |
| *Pasteurellaceae* | 0.00 | 0-0.02 | 0.00 | 0-0.08 | 0.71 | 0.01 | 0-0.21 | 0.08 | 0-0.51 | 0.36 | 0.01 | 0-0.28 | 0.06 | 0.02-0.32 | 0.15 |
| *Akkermansiaceae* | 0.00 | 0-0 | 0.00 | 0-0 | 0.21 | 0.00 | 0-0 | 0.00 | 0-0 | 0.91 | 0.00 | 0-0.37 | 0.00 | 0-0 | 0.02 |
| *Barnesiellaceae/Tarnerellaceae/Prevotellaceae* | 0.00 | 0-0.11 | 0.00 | 0-0.01 | 0.30 | 0.01 | 0-0.15 | 0.01 | 0-0.05 | 0.39 | 0.12 | 0-0.89 | 0.02 | 0-0.3 | 0.32 |

Relative abundance of each bacterial family at 90, 180 and 360 days was determined using 16S rRNA gene sequencing among n=51, 56 and 61 samples at 90, 180 and 360 days, respectively. P-values were derived by Mann-Whitney U test.

Supplementary table 5. Absolute abundance (log10 cells/g feces) of the major bacterial families in children without AD and with AD during the first year of life at each time point

|  |  |  |  |
| --- | --- | --- | --- |
|   | **90 days**  | **180 days**  | **360 days**  |
|   | **non-AD (n=26)** | **AD (n=21)** |  | **non-AD (n=33)** | **AD (n=18)** |  | **non-AD (n=35)** | **AD (n=26)** |  |
| **Major bacterial families**  | med | IQR | med | IQR | P-value  | med | IQR | med | IQR | p-value | med | IQR | med | IQR | p-value |
| *Bifidobacteriaceae* | 8.87 | 8.47-9.94 | 8.70 | 8.2-9.55 | 0.43 | 9.11 | 8.38-9.69 | 8.05 | 7.81-9.14 | 0.03 | 9.26 | 8.68-9.77 | 8.68 | 8.09-9.62 | 0.07 |
| *Enterobacteriaceae* | 7.81 | 6.25-8.86 | 7.61 | 6.91-9.3 | 0.99 | 8.15 | 7.32-8.8 | 7.24 | 6.35-8.58 | 0.10 | 7.81 | 7.23-8.32 | 7.58 | 7.07-8.44 | 0.54 |
| *Bacteroidaceae* | 7.22 | 5.58-8.17 | 6.65 | 5.34-7.98 | 0.44 | 6.81 | 6.14-8.21 | 6.75 | 4.88-7.89 | 0.66 | 8.14 | 7.19-9.33 | 7.86 | 6.23-8.95 | 0.24 |
| *Veillonellaceae* | 7.18 | 5.75-7.54 | 5.82 | 5.07-7.84 | 0.30 | 7.53 | 6.84-8.19 | 7.14 | 6.35-8.17 | 0.33 | 8.05 | 7.33-8.52 | 7.73 | 6.13-8.47 | 0.27 |
| *Streptococcaceae* | 7.03 | 6.56-7.81 | 6.93 | 6.3-7.41 | 0.30 | 7.22 | 6.68-7.73 | 6.56 | 5.96-7.31 | 0.07 | 7.83 | 7.23-8.18 | 7.14 | 6.58-8.36 | 0.28 |
| *Lachnospiraceae* | 6.76 | 5.6-7.75 | 6.85 | 5.68-7.48 | 0.73 | 7.28 | 6.37-8.44 | 6.65 | 5.77-7.79 | 0.12 | 9.17 | 8.87-9.54 | 9.21 | 8.18-9.58 | 0.54 |
| *Enterococcaceae* | 6.75 | 5.86-7.85 | 6.71 | 5.72-7.83 | 0.79 | 7.28 | 6.55-7.63 | 6.52 | 6.08-7.58 | 0.12 | 7.09 | 6.57-7.56 | 6.99 | 6.39-7.64 | 0.72 |
| *Coriobacteriaceae/Actinomycetaceae* | 6.73 | 5.71-7.75 | 6.40 | 5.54-7.09 | 0.36 | 6.39 | 5.96-7.43 | 6.17 | 5.22-7.02 | 0.21 | 7.01 | 6.35-7.71 | 6.60 | 5.32-6.93 | 0.02 |
| *Clostridiaceae*  | 6.62 | 5.06-7.69 | 6.28 | 5.48-7.75 | 0.89 | 7.03 | 6.25-7.99 | 6.10 | 5.49-7.22 | 0.16 | 7.52 | 7.14-8.28 | 7.37 | 6.29-7.87 | 0.08 |
| *Others*  | 6.55 | 5.75-7.14 | 6.28 | 5.75-6.76 | 0.60 | 6.03 | 5.55-7.07 | 5.47 | 5.18-6.38 | 0.09 | 7.05 | 6.18-7.4 | 6.24 | 5.28-7.03 | 0.13 |
| *Peptostreptococcaceae*  | 6.21 | 3.98-6.77 | 0.00 | 0-6.99 | 0.24 | 6.75 | 5.95-7.74 | 5.90 | 4.65-6.78 | 0.07 | 7.71 | 7.07-8.48 | 7.46 | 6.86-8.02 | 0.20 |
| *Erysipelotrichaceae*  | 5.78 | 4.46-7.04 | 0.00 | 0-5.99 | 0.04 | 6.22 | 4.96-7.51 | 5.76 | 4.58-6.62 | 0.35 | 7.99 | 7.38-8.41 | 7.81 | 6.68-8.09 | 0.10 |
| *Eggerthellaceae* | 5.37 | 0-6.15 | 5.72 | 0-6.35 | 0.99 | 5.85 | 0-6.81 | 5.56 | 0-6.51 | 0.77 | 7.27 | 6.62-7.71 | 6.77 | 4.63-7.3 | 0.03 |
| *Lactobacillaceae* | 5.06 | 0-6.69 | 6.72 | 0-7.85 | 0.23 | 5.79 | 0-7.03 | 5.50 | 0-7.6 | 0.96 | 6.23 | 0-7.24 | 6.10 | 0-7.49 | 0.95 |
| *Ruminococcaceae* | 4.60 | 0-5.56 | 0.00 | 0-6.57 | 0.94 | 6.25 | 4.86-7.13 | 4.83 | 0-6.03 | 0.03 | 8.42 | 7.71-8.78 | 7.68 | 6.57-8.57 | 0.05 |
| *Bacillaceae/Paenibacillaceae* | 4.44 | 0-5.95 | 0.00 | 0-4.65 | 0.08 | 4.41 | 0-5.46 | 4.95 | 4.13-5.63 | 0.30 | 6.27 | 5.61-7.04 | 5.84 | 5.16-6.68 | 0.13 |
| *Pasteurellaceae* | 3.16 | 0-5.89 | 3.83 | 0-6.63 | 0.57 | 4.59 | 0-6.62 | 5.30 | 3.28-6.84 | 0.59 | 5.58 | 0-7.09 | 6.36 | 5.35-7 | 0.26 |
| *Akkermansiaceae* | 0.00 | 0-0 | 0.00 | 0-0 | 0.20 | 0.00 | 0-0 | 0.00 | 0-0 | 0.86 | 0.00 | 0-7.24 | 0.00 | 0-0 | 0.02 |
| *Barnesiellaceae/Tarnerellaceae/**Prevotellaceae* | 0.00 | 0-6.12 | 0.00 | 0-5.44 | 0.58 | 5.36 | 0-6.65 | 1.86 | 0-5.58 | 0.21 | 6.94 | 0-7.65 | 5.44 | 0-6.99 | 0.15 |

Absolute abundance (log10 cells/g feces) of each bacterial family was calculated from relative abundance data based on 16S rRNA gene sequencing and from total bacteria cells counts determined by qPCR for n=47, 51 and 61 samples at 90, 180 and 360 days, respectively. P-values were derived by Mann-Whitney U test.



Supplementary figure 3. Microbial fermentation metabolites compared between children with and without AD during their first year of life at each time point. Metabolite measurements were completed among n=60, 65 and 64 samples at 90, 180 and 360 days, respectively. Each box plot shows the interquartile range with a marker for the median value. The dots represent the individual samples.

p-values are derived by Mann-Whitney U test.

Supplementary table 6. Occurrence of samples with detectable levels of fermentation metabolites among children with and without AD during their first year of life

|  |  |  |  |
| --- | --- | --- | --- |
|   | 90 days  | 180 days  | 360 days |
|   | no AD n=36 (%) | AD n=24 (%) | p-value | no AD n=38 (%) | AD n=27 (%) | p-value | no ADn=38 (%) | ADn=26 (%) | p-value |
| **fucose** | 36 (100%) | 24 (100%) | - | 35 (92.1%) | 26 (96.3%) | 0.64 | 33 (86.8%) | 20 (76.9%) | 0.3 |
| **succinate** | 33 (91.7%) | 20 (83.3%) | 0.42 | 33 (86.8%) | 22 (81.5%) | 0.73 | 34 (89.5%) | 25 (96.2%) | 0.64 |
| **lactate** | 31 (86.1%) | 18 (75.0%) | 0.32 | 25 (71.1%) | 21 (77.8%) | 0.58 | 16 (42.1%) | 13 (50%) | 0.53 |
| **formate** | 17 (47.2%) | 12 (50.0%) | 0.55 | 14 (36.8%) | 8 (29.6%) | 0.55 | 4 (10.5%) | 1 (3.8%) | 0.64 |
| **acetate** | 36 (100%) | 24 (100%) | - | 38 (100%) | 27 (100%) | - | 38 (100%) | 26 (100%) | - |
| **1,2-PD** | 19 (52.8%) | 22 (91.7%) | <0.01 | 15 (39.5%) | 11 (40.7%) | 0.92 | 9 (23.7%) | 11 (42.3%) | 0.11 |
| **propionate** | 22 (61.1%) | 12 (50.0%) | 0.4 | 36 (94.7%) | 21 (77.8%) | 0.06 | 37 (97.4%) | 26 (100%) | 1 |
| **butyrate**  | 13 (36.1%) | 8 (33.35%) | 0.83 | 24 (63.2%) | 16 (59.3%) | 0.75 | 35 (92.1%) | 24 (92.3%) | 1 |

Metabolite measurements were completed among n=60, 65 and 64 samples at 90, 180 and 360 days, respectively. Proportion of samples with detectable amounts of metabolites between AD and non-AD groups are compared using chi-squared test or Fisher's exact test if expected value is less than 5.

Supplementary table 7. Association between demographic, environmental and dietary factors and butyrate, butyrate producers and *R. bromii* at 360 days

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|   | Butyrate level at 360 days  | *F. Prausnitzii* cell count at 360 days  | *E. Rectale* cell count at 360 days  | *A. Hallii* cell count at 360 days  | *R. bromii* cell count at 360 days  |
|   | <75percentile | >75percentile | p.value | <75percentile | >75percentile | p.value | <75percentile | >75percentile | p.value | <75percentile | >75percentile | p.value | <75percentile (N=49) | >75percentile | p.value |
| (N=48) | (N=16) | (N=49) | (N=17) | (N=49) | (N=17) | (N=49) | (N=17) | (N=17) |
| Maternal allergic disease\* |   |  |   |  |  |   |  |  |   |  |  |  |   |  |  |
|  no | 17 (73.9%) | 6 (26.1%) | 0.88 | 17 (70.8%) | 7 (29.2%) | 0.63 | 19 (79.2%) | 5 (20.8%) | 0.49 | 18 (75.0%) | 6 (25.0%) | 0.92 | 16 (66.7%) | 8 (33.3%) | 0.29 |
|  yes | 31 (75.6%) | 10 (24.4%) |   | 32 (76.2%) | 10 (23.8%) |   | 30 (71.4%) | 12 (28.6%) |   | 31 (73.8%) | 11 (26.2%) |  | 33 (78.6%) | 9 (21.4%) |  |
| Paternal allergic disease\* |   |  |   |  |  |   |  |  |   |  |  |  |   |  |  |
|  no | 28 (80.0%) | 7 (20.0%) | 0.31 | 25 (69.4%) | 11 (30.6%) | 0.33 | 27 (75.0%) | 9 (25.0%) | 0.88 | 25 (69.4%) | 11 (30.6%) | 0.33 | 28 (77.8%) | 8 (22.2%) | 0.47 |
|  yes | 20 (69.0%) | 9 (31.0%) |   | 24 (80.0%) | 6 (20.0%) |   | 22 (73.3%) | 8 (26.7%) |   | 24 (80.0%) | 6 (20.0%) |  | 21 (70.0%) | 9 (30.0%) |  |
| Mode of delivery  |   |  |   |  |  |   |  |  |   |  |  |  |   |  |  |
|  vaginal | 37 (74.0%) | 13 (26.0%) | 0.73 | 39 (75.0%) | 13 (25.0%) | 0.79 | 39 (75.0%) | 13 (25.0%) | 0.79 | 39 (75.0%) | 13 (25.0%) | 0.79 | 38 (73.1%) | 14 (26.9%) | 0.68 |
|  caesarean section | 11 (78.6%) | 3 (21.4%) |   | 10 (71.4%) | 4 (28.6%) |   | 10 (71.4%) | 4 (28.6%) |   | 10 (71.4%) | 4 (28.6%) |  | 11 (78.6%) | 3 (21.4%) |  |
| Antibiotics use up to 4 months |   |  |   |  |  |   |  |  |   |  |  |  |   |  |  |
|  none | 42 (76.4%) | 13 (23.6%) | 0.53 | 40 (71.4%) | 16 (28.6%) | 0.22 | 40 (71.4%) | 16 (28.6%) | 0.22 | 40 (71.4%) | 16 (28.6%) | 0.22 | 41 (73.2%) | 15 (26.8%) | 0.65 |
|  once or more | 6 (66.7%) | 3 (33.3%) |   | 9 (90.0%) | 1 (10.0%) |   | 9 (90.0%) | 1 (10.0%) |   | 9 (90.0%) | 1 (10.0%) |  | 8 (80.0%) | 2 (20.0%) |  |
| Any siblings  |   |  |   |  |  |   |  |  |   |  |  |  |   |  |  |
|  none | 35 (83.3%) | 7 (16.7%) | 0.03 | 34 (77.3%) | 10 (22.7%) | 0.43 | 31 (70.5%) | 13 (29.5%) | 0.32 | 36 (81.8%) | 8 (18.2%) | 0.05 | 35 (79.5%) | 9 (20.5%) | 0.16 |
|  one or more  | 13 (59.1%) | 9 (40.9%) |   | 15 (68.2%) | 7 (31.8%) |   | 18 (81.8%) | 4 (18.2%) |   | 13 (59.1%) | 9 (40.9%) |  | 14 (63.6%) | 8 (36.4%) |  |
| Pet dog or cat during pregnancy |   |  |   |  |  |   |  |  |   |  |  |  |   |  |  |
|  no | 36 (72.0%) | 14 (28.0%) | 0.3 | 40 (76.9%) | 12 (23.1%) | 0.34 | 39 (75.0%) | 13 (25.0%) | 0.79 | 39 (75.0%) | 13 (25.0%) | 0.79 | 39 (75.0%) | 13 (25.0%) | 0.79 |
|  yes | 12 (85.7%) | 2 (14.3%) |   | 9 (64.3%) | 5 (35.7%) |   | 10 (71.4%) | 4 (28.6%) |   | 10 (71.4%) | 4 (28.6%) |   | 10 (71.4%) | 4 (28.6%) |  |
| Duration of breastfeeding\* |   |  |   |  |  |   |  |  |   |  |  |  |   |  |  |
| less than median  | 22 (78.6%) | 6 (21.4%) | 0.54 | 22 (75.9%) | 7 (24.1%) | 0.76 | 21 (72.4%) | 8 (27.6%) | 0.54 | 22 (75.9%) | 7 (24.1%) | 1 | 22 (75.9%) | 7 (24.1%) | 0.75 |
| more than median  | 20 (71.4%) | 8 (28.6%) |   | 21 (72.4%) | 8 (27.6%) |   | 23 (79.3%) | 6 (20.7%) |   | 22 (75.9%) | 7 (24.1%) |  | 23 (79.3%) | 6 (20.7%) |  |
| Number of food items introduced in the first year of life\*  |   |  |   |  |  |   |  |  |   |  |  |  |   |  |  |
| less than median  | 30 (78.9%) | 8 (21.1%) | 0.32 | 31 (79.5%) | 8 (20.5%) | 0.18 | 30 (76.9%) | 9 (23.1%) | 0.79 | 32 (82.1%) | 7 (17.9%) | 0.12 | 29 (74.4%) | 10 (25.6%) | 0.40 |
| more than median  | 12 (66.7%) | 6 (33.3%) |   | 12 (63.2%) | 7 (36.8%) |   | 14 (73.7%) | 5 (26.3%) |   | 12 (63.2%) | 7 (36.8%) |  | 16 (84.2%) | 3 (15.8%) |  |
| Timing of solid introduction\*  |   |  |   |  |  |   |  |  |   |  |  |  |   |  |  |
| earlier than median  | 21 (72.4%) | 8 (27.6%) | 0.64 | 21 (70.0%) | 9 (30.0%) | 0.46 | 21 (70.0%) | 9 (30.0%) | 0.28 | 22 (73.3%) | 8 (26.7%) | 0.64 | 21 (70.0%) | 9 (30.0%) | 0.15 |
| later than median  | 21 (77.8%) | 6 (22.2%) |   | 22 (78.6%) | 6 (21.4%) |   | 23 (82.1%) | 5 (17.9%) |   | 22 (78.6%) | 6 (21.4%) |  | 24 (85.7%) | 4 (14.3%) |  |

\* Analysis with dietary factors were conducted using data from children with sufficient dietary history collected during their first year of life and measurements completed. Total numbers of samples included are n= 56 for butyrate, n=58 for qPCR measurements. Duration of breastfeeding, number of items introduced in the first year of life and timing of solid introduction were each categorized at the median value, 40.5 weeks, 17 items and 20 weeks, respectively.



Supplementary figure 4. Duration of consumption of selected food items during the first year of life and butyrate, abundance of major butyrate producers (*F. Prausnitzii*, *E. rectale* and *A. hallii*) and *R. bromii* at 360 days. Children with sufficient history of diet during the first year of life and butyrate or qPCR measurement completed at 360 days (n=56 and n=58, respectively) were included in this analysis.

Duration of A. yogurt and B. fish consumption were each categorized at the median value, 14 weeks and 5 weeks, respectively. Duration of C. vegetable and fruit consumption was calculated as the sum of the duration of the 6 categories in the questionnaire and categorized at the median value of 145 weeks. For each food item, the proportion of children with high level of butyrate or abundance of selected bacterial groups using the 75-percentile value as a cut-off were compared between children who had consumed the food less and more than the median. P-values were derived by Chi-squared test.

References

1. Fierer N, Jackson JA, Vilgalys R, Jackson RB. Assessment of soil microbial community structure by use of taxon-specific quantitative PCR assays. Appl Environ Microbiol. 2005;71(7):4117-20.

2. Vital M, Penton CR, Wang Q, Young VB, Antonopoulos DA, Sogin ML, et al. A gene-targeted approach to investigate the intestinal butyrate-producing bacterial community. Microbiome. 2013;1(1):8.

3. Appert O, Garcia AR, Frei R, Roduit C, Constancias F, Neuzil-Bunesova V, et al. Initial butyrate producers during infant gut microbiota development are endospore formers. Environ Microbiol. 2020;22(9):3909-21.