Supplementary Materials and Methods

Study cohort

Newborn infants were recruited to the birth cohort as part of the international DIABIMMUNE study (http://www.diabimmune.org/). Recruitment took place in Espoo, Finland between September 2008 and May 2010. Inclusion criteria for this study included receiving either none or at least 9 antibiotic courses in the first three years of life. The participating children were monitored prospectively for infections, use of antibiotics, and other life events. Data on breastfeeding and introduction of complementary foods were registered in a study booklet and interview at each study visit (3, 6, 12, 18, 24, and 36 months). Subject metadata is available in Table S1. The DIABIMMUNE study was conducted according to the guidelines laid down in the Declaration of Helsinki, and all procedures involving human subjects were approved by the local ethical committee. The parents gave their written informed voluntary consent prior to sample collection.

Stool sample collection and DNA extraction

Stool samples were collected by the participants' parents and stored in the household freezer (−20°C) until the next scheduled visit to the local study center; samples were then shipped on dry ice to the DIABIMMUNE Core Laboratory. Samples were stored at −80°C until shipping to the Broad Institute for DNA extraction. DNA extractions from stool were carried out using the QIAamp DNA Stool Mini Kit (QIAGEN).

Sequencing and analysis of the 16S rRNA gene and whole-genome shotgun sequencing

16S rRNA gene sequencing was performed essentially as previously described (*3*). Taxonomy was assigned using version 1.8.0 of Qiime (*54*) and the Greengenes reference database of OTUs (*55*). A mean sequence depth of 48,131 per sample was obtained, and samples with less than 3,000 sequences were excluded from analysis.

We selected 240 samples for whole-genome shotgun sequencing (also referred to as metagenomic sequencing) based on two criteria: (1) samples from all children at ages 2, 12, 24, and 36 months, and (2) at least four samples before and after selected antibiotic treatments (with minimal additional treatments in this time period).

Metagenome library construction

Metagenomic DNA samples were quantified by Quant-iT PicoGreen dsDNA Assay (Life Technologies) and normalized to a concentration of 50 pg μL^{-1}. Illumina sequencing libraries were prepared from 100-250 pg DNA using the Nextera XT DNA Library Preparation kit (Illumina) according to the manufacturer's recommended protocol, with reaction volumes scaled accordingly. Batches of 24, 48, or 96 libraries were pooled by transferring equal volumes of each library using a Labcyte Echo 550 liquid handler. Insert sizes and concentrations for each pooled library were determined using an Agilent Bioanalyzer DNA 1000 kit (Agilent Technologies).

Diversity analysis based on 16S rRNA gene sequencing data

Microbial richness (alpha diversity) was measured using the Chao1 metric, as implemented in Qiime (*54*) version 1.8.0. To account for the decrease in diversity caused by the varying sequencing depth, we subsampled each sample to 10,000 reads and reported the average alpha

diversity across 100 subsampling iterations.

Stability analysis based on 16S rRNA gene sequencing data

The Jaccard index for a given sample pair is defined as |sample A ∩ sample B| / |sample A ∪ sample B|, i.e. the number of items (here, OTUs) in common between samples A and B divided by the total number of items present in either sample A or sample B. Jaccard indices was calculated for all within-subject sample pairs; for samples collected 1 month apart, the median of the samples was calculated for the overall stability measure per child.

To estimate the variation in the stability measure per group, we performed the following analysis for each group (Abx^- and Abx^+). We denoted the stability measure of a group by S, where n = |S|. We sampled n times with replacement from S, and calculated the variance of the sampled set. We performed this step 1,000 times and calculated the standard deviation of these measures.

Analysis of whole-genome shotgun (WGS) sequencing

WGS libraries were sequenced on the Illumina HiSeq 2500 platform, targeting ~2.5 Gb of sequence per sample with 101 bp paired-end reads. Reads were quality controlled by trimming low-quality bases, dropping reads below 60 nucleotides, and filtering out potential human contamination. Quality controlled samples were profiled taxonomically using MetaPhlAn 2.0 (*40*), following Bowtie 2-2.1.0 (*56*) alignment to the MetaPhlAn 2.0 unique marker database (http://huttenhower.sph.harvard.edu/metaphlan2).

Estimating differences between strains

For each species, we used the output of the ConStrains method, as explained above, and extracted the SNPs profiles for all strains in that species across all individuals. We calculated the mutation distance between all strain pairs and constructed the phylogenetic tree based on this distance matrix, where branch lengths correspond to the mutation distance; trees were constructed using the nearest neighbors approach.

Next, using the phylogenetic tree, we identified the most common recent ancestor (MRCA) for each subject, as the root of the minimal sub-tree that contains all strains of that subject as leaves. We then calculated for each subject the median distance from its strains to its MRCA, or the median distance from the MRCA to all other MRCAs. We used these measures to plot the species in fig. S7.

Measuring antibiotic resistance genes

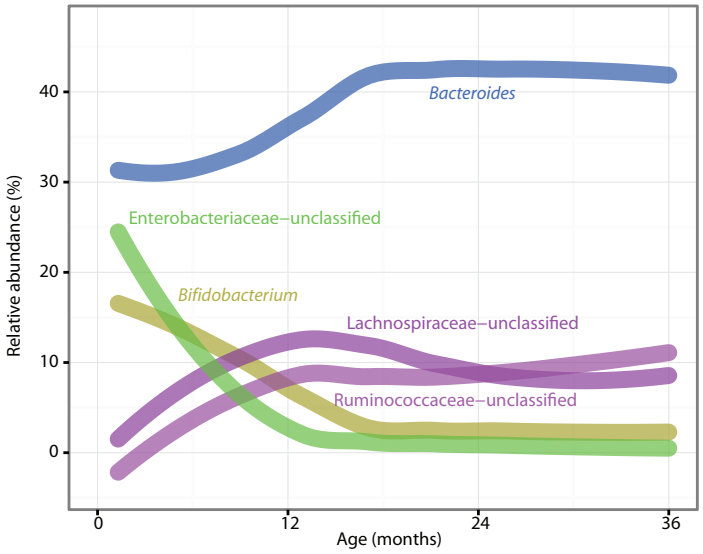
To detect and quantify the abundance of the antibiotic resistance (AR) genes in our WGS data, we used a recently developed tool called shortBRED (http://huttenhower.sph.harvard.edu/shortbred). Briefly, given a set of AR protein sequences, shortBRED clusters them into similar families based on their sequence, extracts a set of distinctive strings ("markers") per family, and then searches for these markers in metagenomic data. We did not take into consideration genes that are normally present in the core genome of the species and in which point mutations can give rise to antibiotic resistance, as we need very high read coverage to clearly identify these mutations. Instead, we focused on genes whose presence is sufficient to confer resistance. Specifically, we used the sequences of 3,060 proteins from The Comprehensive Antibiotic Resistance Database (*45*) (http://arpcard.mcmaster.ca/).

Statistics

Information regarding statistical tests is included in figure legends and/or in the detailed analytical methods above.

Figure S1

A. Average profiles of highly abundant genera



B. Average profiles of lowly abundant genera

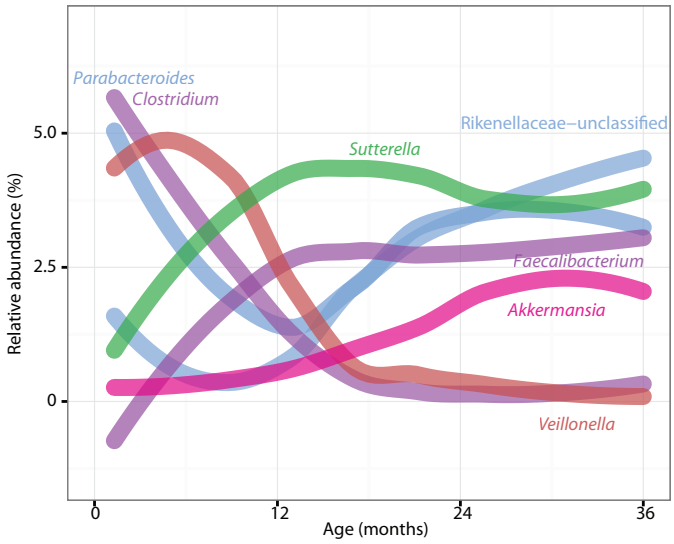


Fig. S1. Average relative abundance of dominant genera in all 39 children. (**A** and **B**) Highly (A) and lowly (B) abundant genera are shown, color-coded as in Figure 1C.

Figure S2

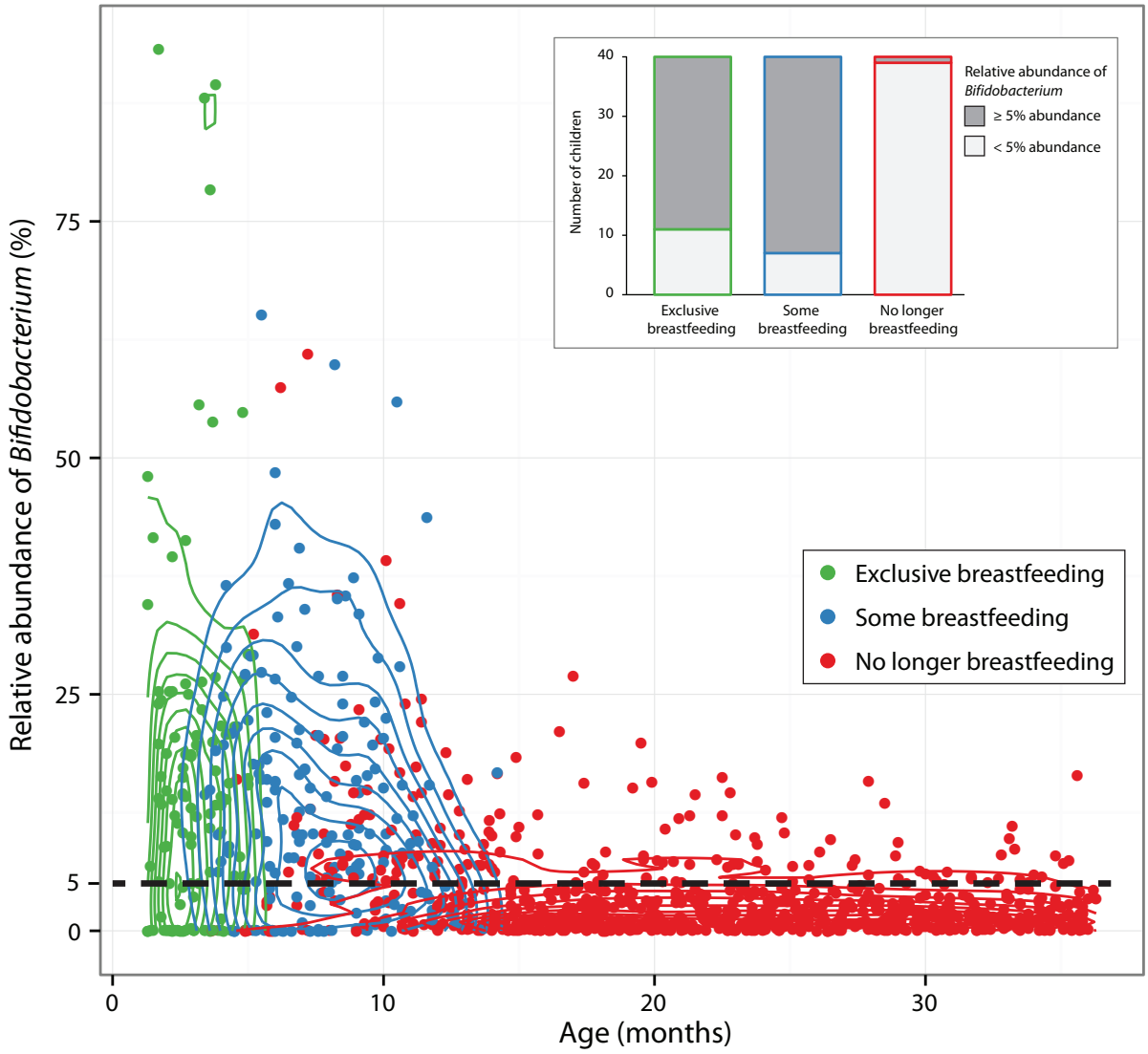

Relative abundance of Bifidobacterium (%)
75
50
25
5
0
0
10
20
30
Age (months)
Exclusive breastfeeding
Some breastfeeding
No longer breastfeeding
Number of children
40
30
20
10
0
Exclusive breastfeeding
Some breastfeeding
No longer breastfeeding
Relative abundance of Bifidobacterium
≥ 5% abundance
< 5% abundance

Fig. S2. *Bifidobacterium* abundance together with early feeding data. Samples are plotted by their age (x-axis) and their relative abundance of *Bifidobacterium* species (y-axis), and are colored by their early feeding state (exclusive breastfeeding, green; some breastfeeding, blue; no breastfeeding, red). Inset shows the number of children with a median *Bifidobacterium* abundance of less than 5%, at each feeding state.

Figure S3

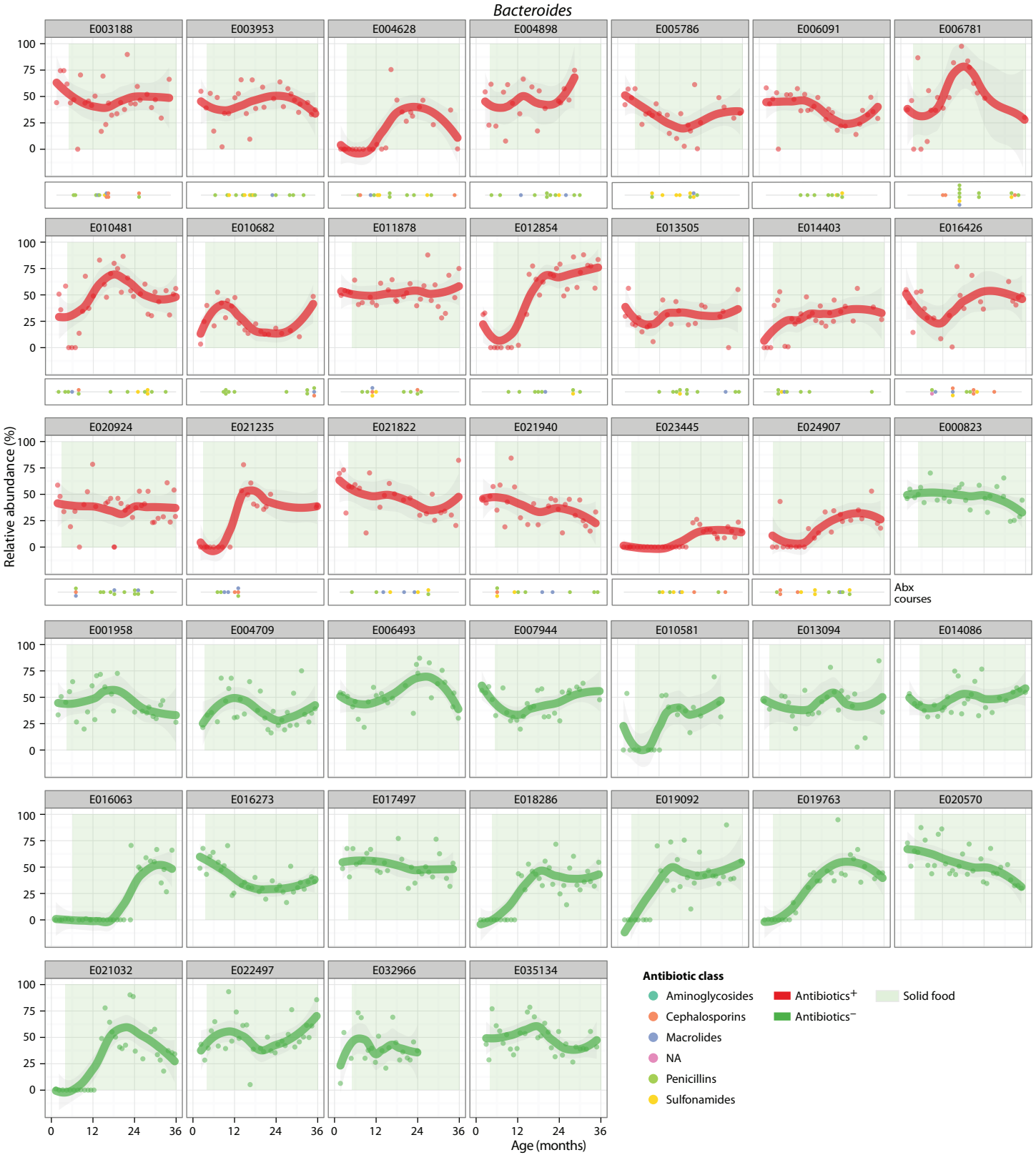

Bacteroides
E003188
E003953
E004628
E004898
E005786
E006091
E006781
E010481
E010682
E011878
E012854
E013505
E014403
E016426
E020924
E021235
E021822
E021940
E023445
E024907
E000823
E001958
E004709
E006493
E007944
E010581
E013094
E014086
E016063
E016273
E017497
E018286
E019092
E019763
E020570
E021032
E022497
E032966
E035134
Relative abundance (%)
100
75
50
25
0
Abx courses
0
12
24
36
Age (months)
Antibiotic class
Aminoglycosides
Cephalosporins
Macrolides
NA
Penicillins
Sulfonamides
Antibiotics+
Antibiotics−
Solid food

Fig. S3. Individual profiles of *Bacteroides* abundance together with solid food consumption. Data are divided by antibiotic treatment (Abx^+, red lines; Abx^-, green lines). Light green shading represents the time period during which the child consumed solid food. Shaded gray regions indicate 95% confidence intervals. The number and order of antibiotic courses are shown with each antibiotic class indicated by color.

Figure S4

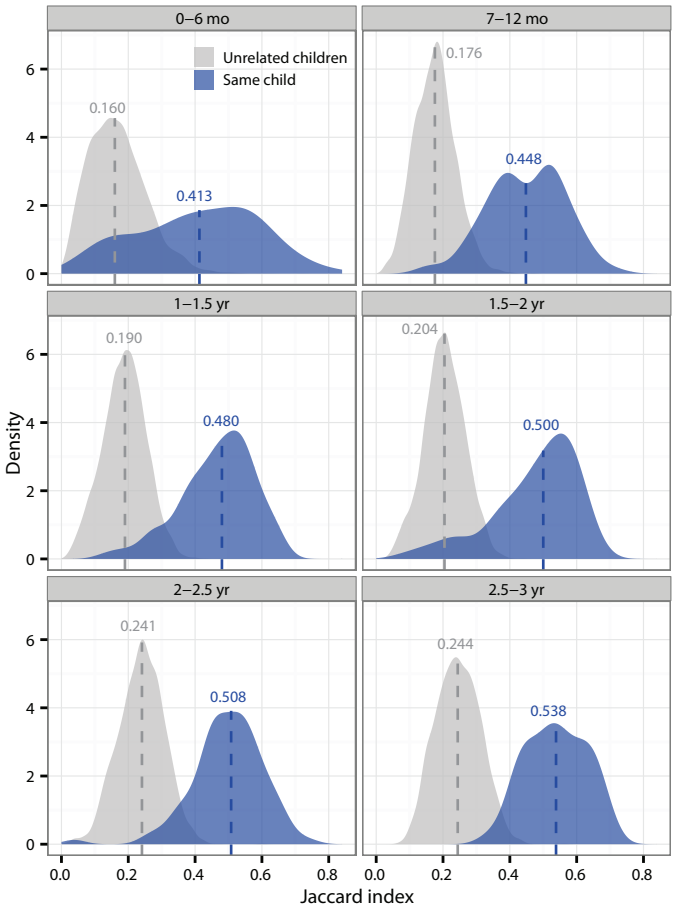

0–6 mo
7–12 mo
1–1.5 yr
1.5–2 yr
2–2.5 yr
2.5–3 yr
Unrelated children
Same child
0.160
0.413
0.176
0.448
0.190
0.480
0.204
0.500
0.241
0.508
0.244
0.538
Density
Jaccard index
0
2
4
6
0.0
0.2
0.4
0.6
0.8

Fig. S4. Consistency of the infant gut microbiome. Shown are distributions of Jaccard indices calculated using samples collected 1 month apart either from the same individual (blue) or from age-matched samples from different individuals (gray), separated into 6-month periods, with the median of each sub-population.

Figure S5

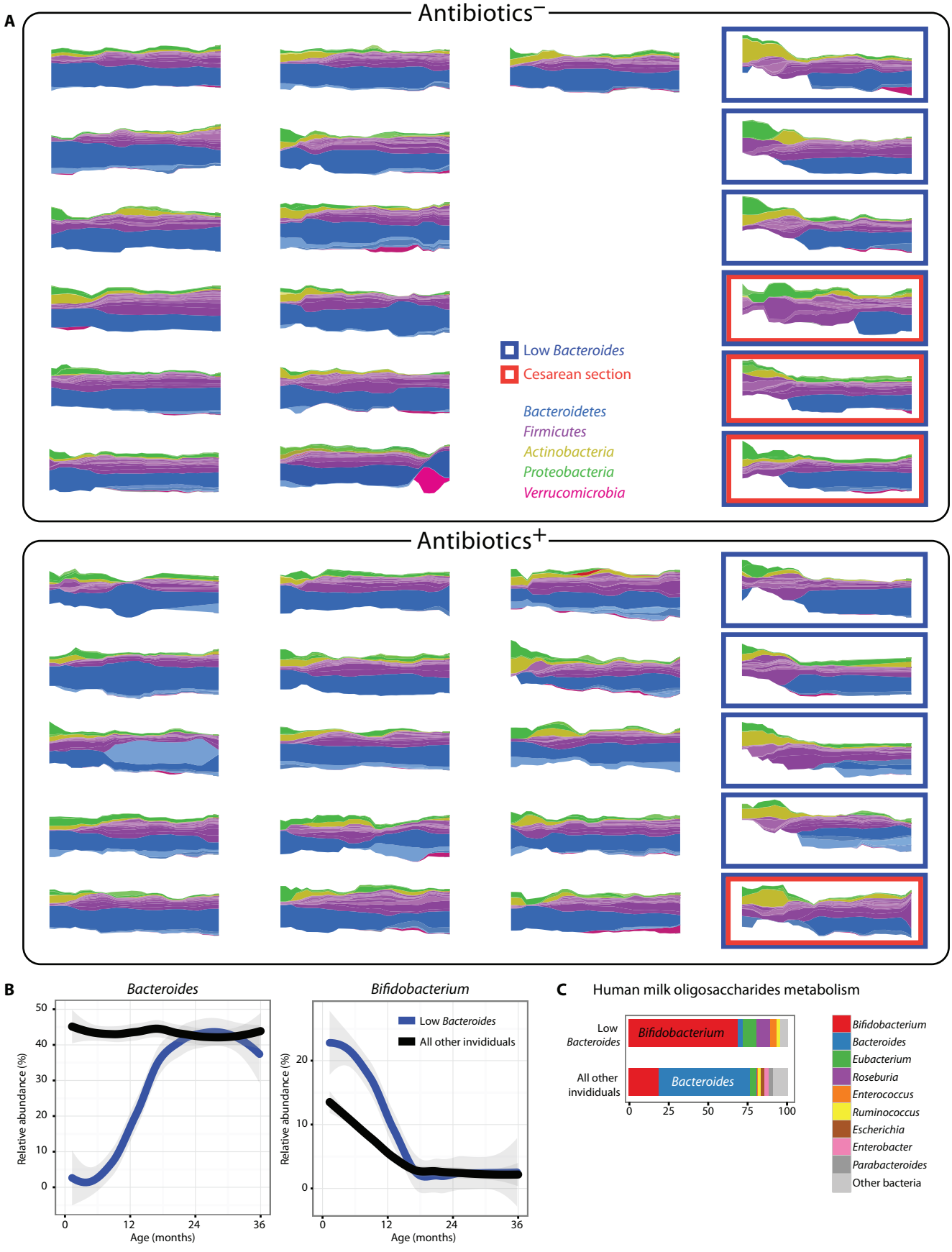

A
Antibiotics⁻
Antibiotics⁺
Low Bacteroides
Cesarean section
Bacteroidetes
Firmicutes
Actinobacteria
Proteobacteria
Verrucomicrobia
B
Bacteroides
Bifidobacterium
Low Bacteroides
All other invididuals
Relative abundance (%)
Age (months)
C
Human milk oligosaccharides metabolism
Low Bacteroides
All other invididuals
Bifidobacterium
Bacteroides
Bifidobacterium
Bacteroides
Eubacterium
Roseburia
Enterococcus
Ruminococcus
Escherichia
Enterobacter
Parabacteroides
Other bacteria

Fig. S5. Microbial trajectories for all children in the study. (**A**) Stream plots, as in Figure 1C, for all individuals. The low *Bacteroides* group is highlighted in blue and children born by Cesarean section are highlighted in red. (**B**) Average abundance profiles of the *Bacteroides* and *Bifidobacterium* genera, as in figure S1, differentiating the low *Bacteroides* group (blue) from all other children (black). Shaded gray regions indicate 95% confidence intervals. (**C**) Median contribution of various species (colored bars) to the metabolism of human milk oligosaccharides, differentiating the low *Bacteroides* group (top) from all other children (bottom).

Figure S6

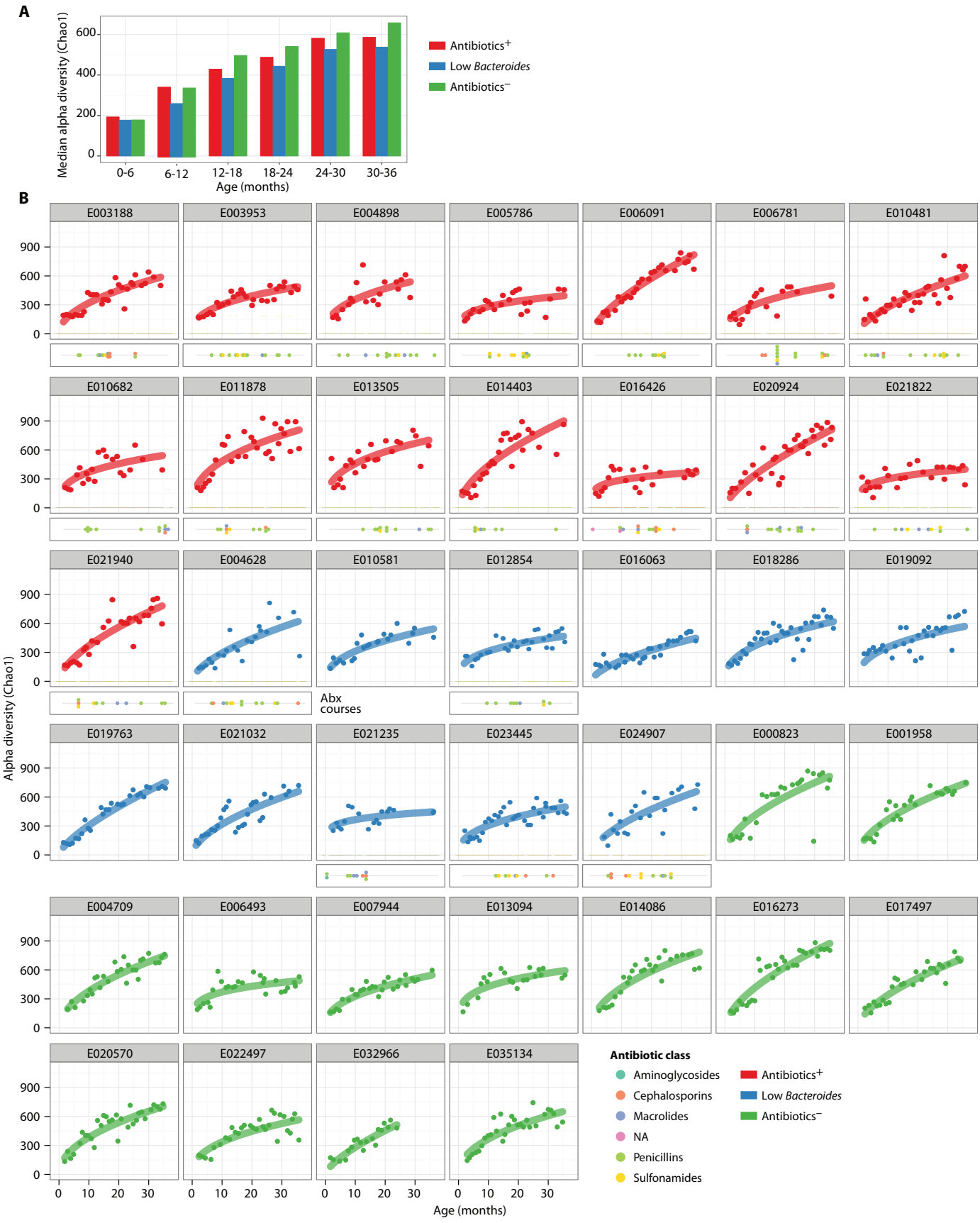

A
Median alpha diversity (Chao1)
600
400
200
0
0-6
6-12
12-18
18-24
24-30
30-36
Age (months)
Antibiotics+
Low Bacteroides
Antibiotics−
B
E003188
E003953
E004898
E005786
E006091
E006781
E010481
E010682
E011878
E013505
E014403
E016426
E020924
E021822
E021940
E004628
E010581
E012854
E016063
E018286
E019092
Abx courses
E019763
E021032
E021235
E023445
E024907
E000823
E001958
E004709
E006493
E007944
E013094
E014086
E016273
E017497
E020570
E022497
E032966
E035134
Alpha diversity (Chao1)
900
600
300
0
0
10
20
30
Age (months)
Antibiotic class
Aminoglycosides
Cephalosporins
Macrolides
NA
Penicillins
Sulfonamides
Antibiotics+
Low Bacteroides
Antibiotics−

Fig. S6. Richness of the infant gut microbiome. Microbial richness (Chao1) of the community as a function of age as measured in all samples, using 16S rRNA gene sequencing data. (**A**) Median richness values are shown at 6-month intervals, colored according to three groups: children who received antibiotics (red), children with low *Bacteroides* (blue), and children who received no antibiotics (green). (**B**) Plots are shown for each child, together with antibiotic treatment profile (when present). Samples are colored as in (A). The number and order of antibiotic courses are shown (colored dots), with each antibiotic class indicated by color.

Figure S7

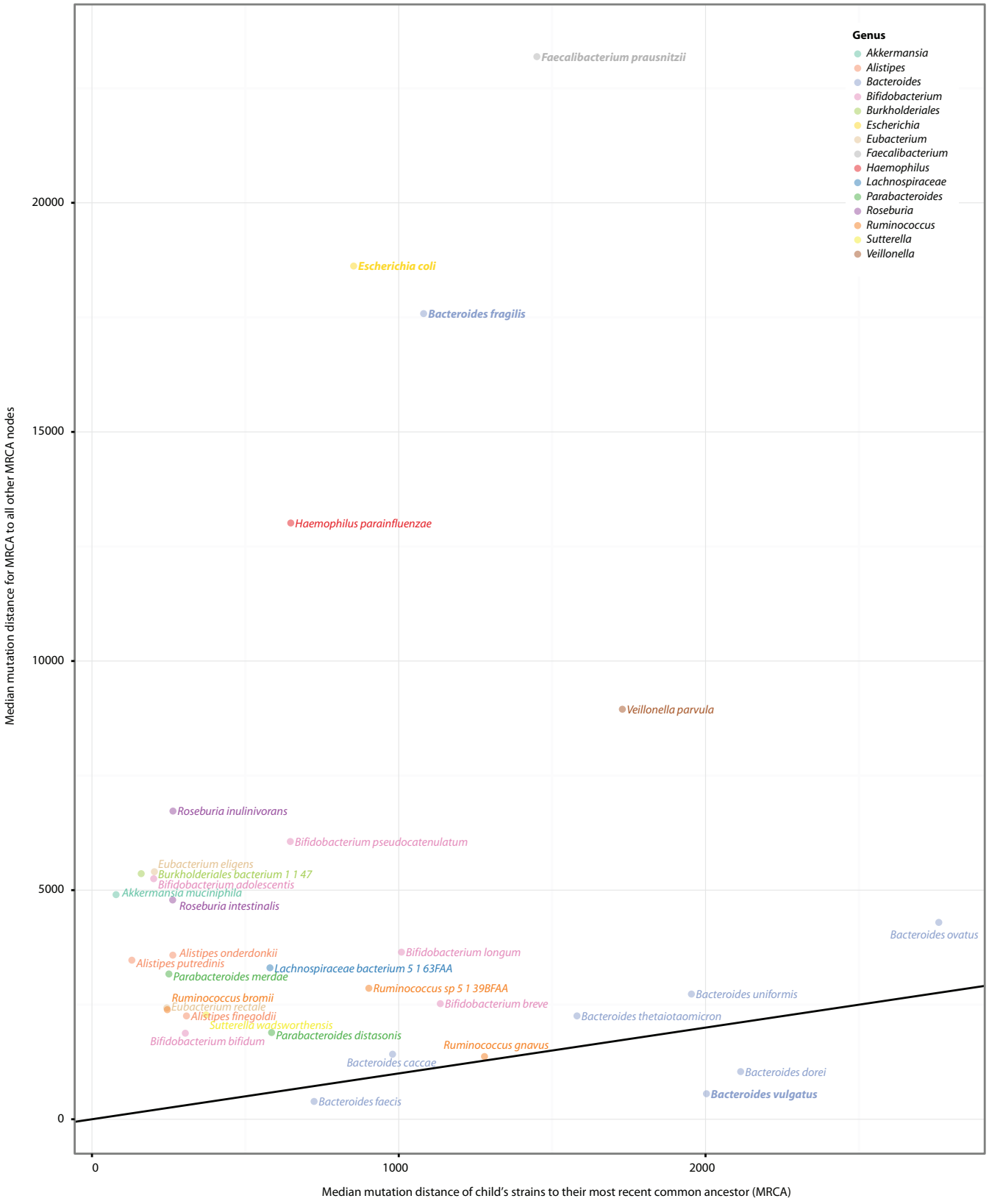

Genus
Akkermansia
Alistipes
Bacteroides
Bifidobacterium
Burkholderiales
Escherichia
Eubacterium
Faecalibacterium
Haemophilus
Lachnospiraceae
Parabacteroides
Roseburia
Ruminococcus
Sutterella
Veillonella
Faecalibacterium prausnitzii
Escherichia coli
Bacteroides fragilis
Haemophilus parainfluenzae
Veillonella parvula
Roseburia inulinivorans
Bifidobacterium pseudocatenulatum
Eubacterium eligens
Burkholderiales bacterium 1 1 47
Bifidobacterium adolescentis
Akkermansia muciniphila
Roseburia intestinalis
Bacteroides ovatus
Alistipes onderdonkii
Alistipes putredinis
Parabacteroides merdae
Bifidobacterium longum
Lachnospiraceae bacterium 5 1 63FAA
Ruminococcus sp 5 1 39BFAA
Bifidobacterium breve
Bacteroides uniformis
Ruminococcus bromii
Eubacterium rectale
Alistipes finegoldii
Sutterella wadsworthensis
Parabacteroides distasonis
Bacteroides thetaiotaomicron
Bifidobacterium bifidum
Ruminococcus gnavus
Bacteroides caccae
Bacteroides dorei
Bacteroides vulgatus
Bacteroides faecis
20000
15000
10000
5000
0
0
1000
2000
Median mutation distance for MRCA to all other MRCA nodes
Median mutation distance of child's strains to their most recent common ancestor (MRCA)

Fig. S7. Strain similarity patterns of abundant species. Species are plotted according to the median distance of strains to the child's most recent common ancestor (MRCA; x-axis), and the median distance between all MRCAs (y-axis). "Single-introduction species" cluster at higher y-axis values.

Figure S8

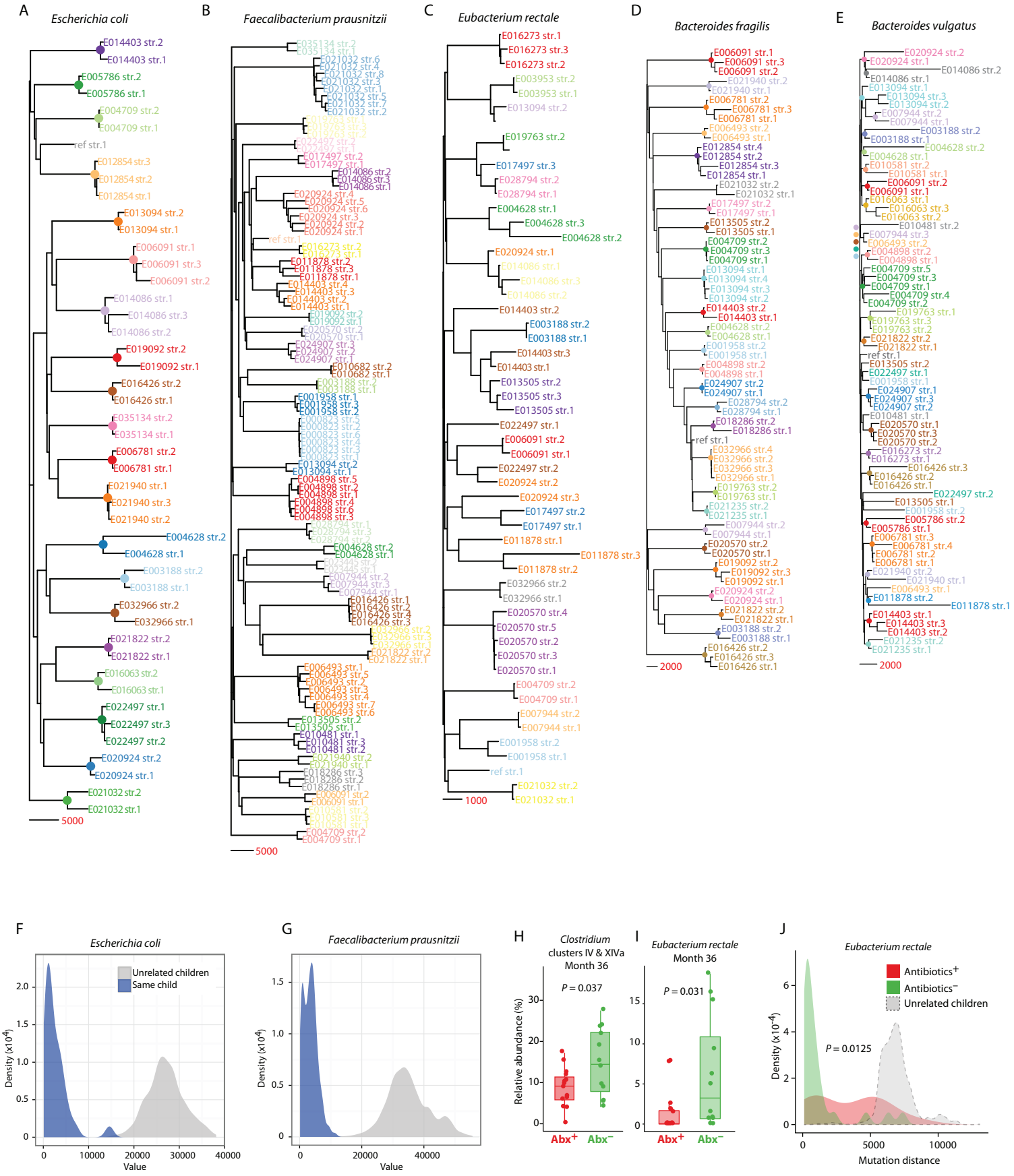

A
Escherichia coli
B
Faecalibacterium prausnitzii
C
Eubacterium rectale
D
Bacteroides fragilis
E
Bacteroides vulgatus
F
Escherichia coli
Unrelated children
Same child
Density (x10^4)
Value
G
Faecalibacterium prausnitzii
Density (x10^4)
Value
H
Clostridium clusters IV & XIVa
Month 36
P = 0.037
Relative abundance (%)
Abx+
Abx−
I
Eubacterium rectale
Month 36
P = 0.031
Abx+
Abx−
J
Eubacterium rectale
Antibiotics+
Antibiotics−
Unrelated children
P = 0.0125
Density (x10^-4)
Mutation distance

Fig. S8. Strain diversity. **(A** to **E)** Phylogenetic trees (as in Fig. 2D) based on the mutation distance between all strains of *Escherichia coli* (A), *Faecalibacterium prausnitzii* (B), *Eubacterium rectale* (C), *Bacteroides fragilis* (D) and *Bacteroides vulgatus* (E). Scale bars of the mutation distances are shown per tree. **(F-G)** Mutation distance distributions (as in Fig. 2E), for strains of *Escherichia coli* (F), and *Faecalibacterium prausnitzii* (G). **(H)** Total relative abundance of all members of Clostridium clusters IV and XIVa, as measured at age 36 months. Box boundaries are the 25th and 75th percentiles, and the median is highlighted. **(I)** Relative abundance of *Eubacterium rectale*, the most abundant member of the Clostridium clusters IV and XIVa, at age 36 months. **(J)** Strain similarity distribution as in Fig. 2F for the E. rectale strains (colored as in Fig. 2F with gray for across-individual comparisons), with a P value for the separation of the Abx– and Abx+ distributions (KS-test).

Figure S9

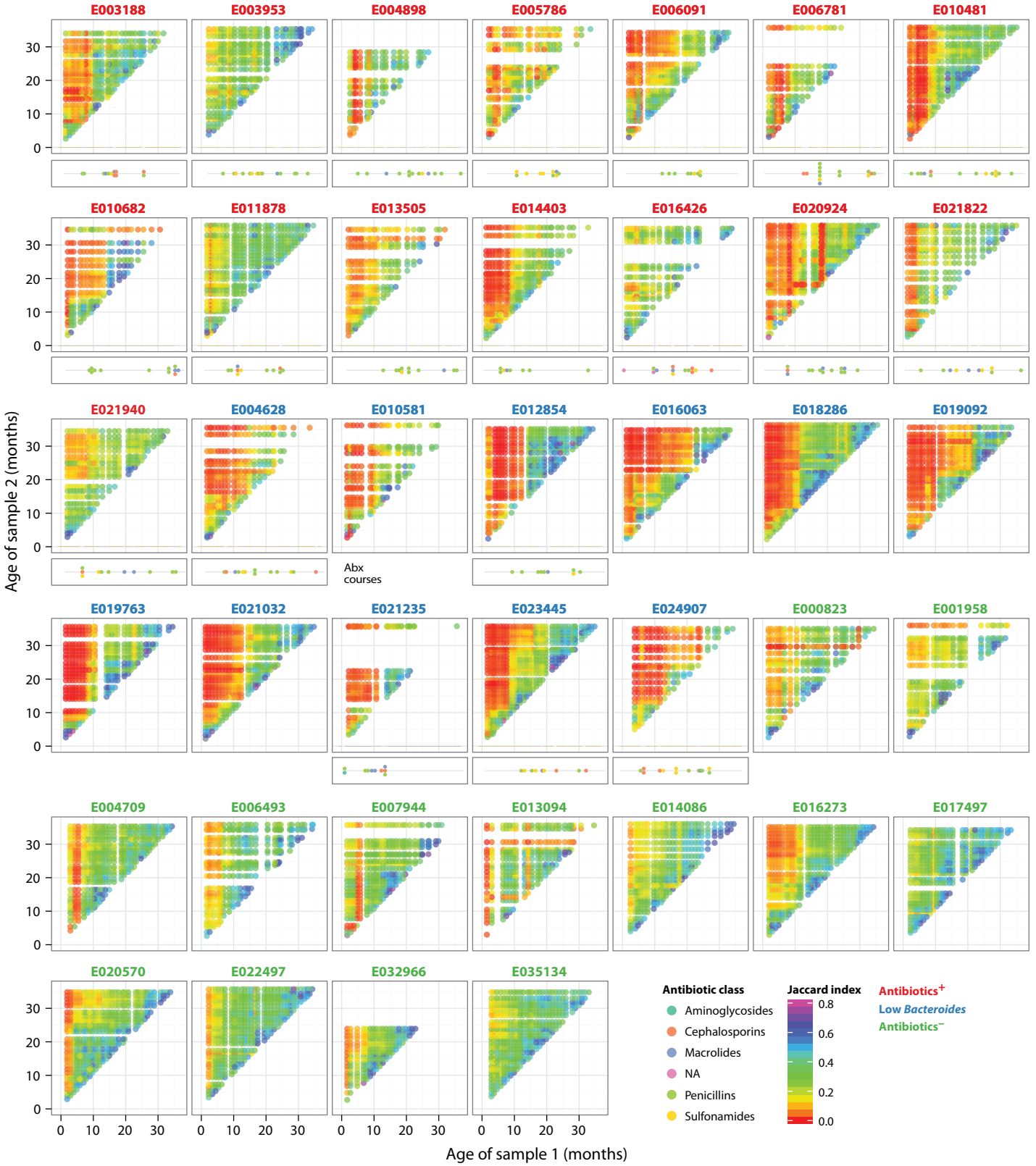

E003188
E003953
E004898
E005786
E006091
E006781
E010481
E010682
E011878
E013505
E014403
E016426
E020924
E021822
E021940
E004628
E010581
E012854
E016063
E018286
E019092
E019763
E021032
E021235
E023445
E024907
E000823
E001958
E004709
E006493
E007944
E013094
E014086
E016273
E017497
E020570
E022497
E032966
E035134
Abx courses
Age of sample 2 (months)
Age of sample 1 (months)
0
10
20
30
Antibiotic class
Aminoglycosides
Cephalosporins
Macrolides
NA
Penicillins
Sulfonamides
Jaccard index
0.8
0.6
0.4
0.2
0.0
Antibiotics+
Low Bacteroides
Antibiotics−

Fig. S9. Stability of the infant gut microbiome. Plots as in Figure 3A-D for all subjects in the study. Child identifiers are colored as Abx^+ (red), low *Bacteroides* (blue), or Abx^- (green). The number and order of antibiotic courses are shown with each antibiotic class indicated by color.

Figure S10

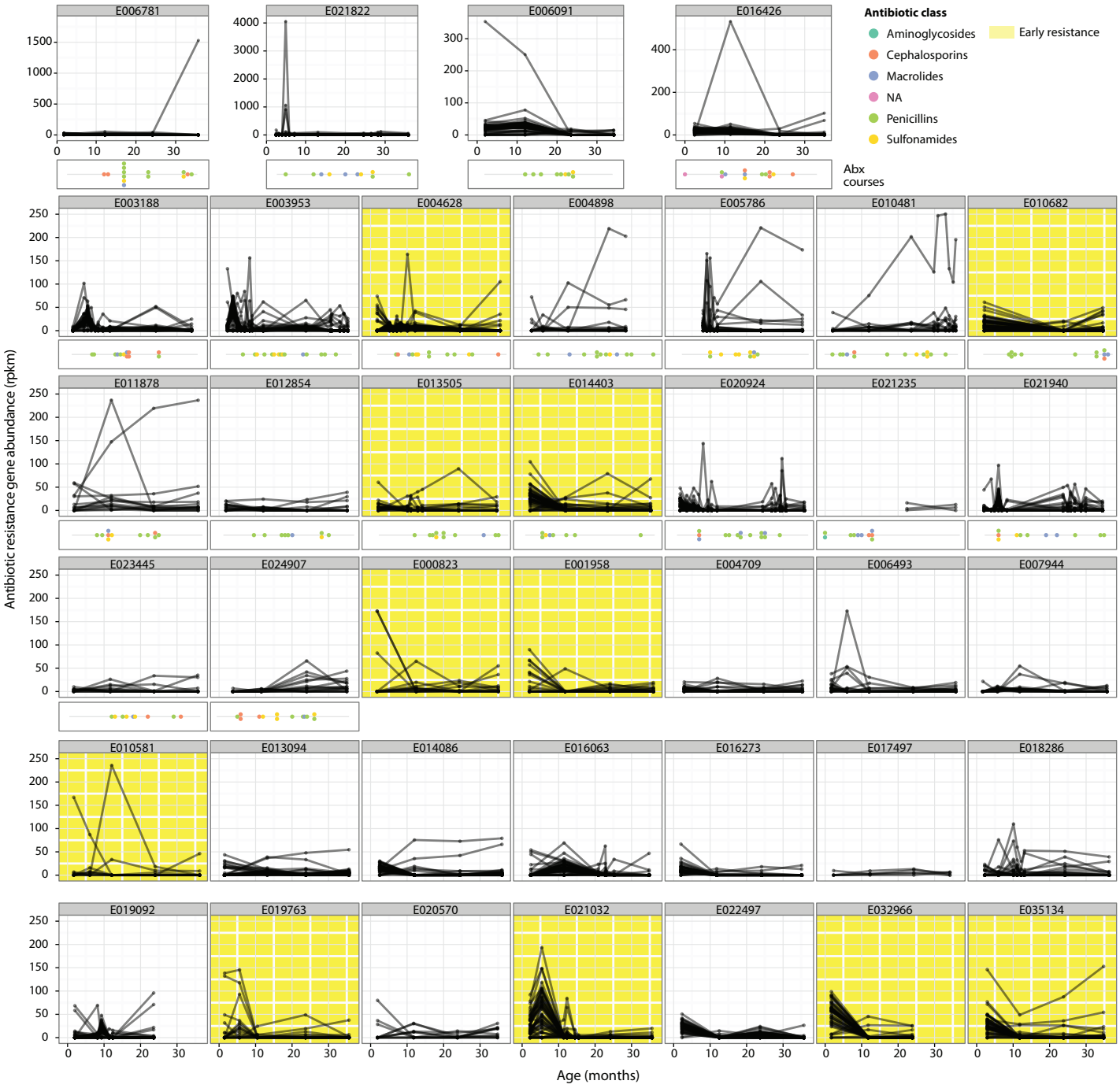

E006781
E021822
E006091
E016426
Antibiotic class
Aminoglycosides
Cephalosporins
Macrolides
NA
Penicillins
Sulfonamides
Early resistance
Abx courses
E003188
E003953
E004628
E004898
E005786
E010481
E010682
E011878
E012854
E013505
E014403
E020924
E021235
E021940
E023445
E024907
E000823
E001958
E004709
E006493
E007944
E010581
E013094
E014086
E016063
E016273
E017497
E018286
E019092
E019763
E020570
E021032
E022497
E032966
E035134
Antibiotic resistance gene abundance (rpkm)
Age (months)

Fig. S10. Abundance profiles for antibiotic resistance (AR) genes. As in Figure 4, abundance of AR genes in all children over time, together with the timing of individual antibiotic courses. Children with an early abundance of AR genes are highlighted in yellow.

Table S1. Clinical variables used in this study including birth mode, early feeding history, and antibiotic treatments.

The members of the DIABIMMUNE Study Group are Mikael Knip, Katriina Koski, Matti Koski, Taina Härkönen, Samppa Ryhänen, Heli Siljander, Anu-Maaria Hämäläinen, Anne Ormisson, Aleksandr Peet, Vallo Tillmann, Valentina Ulich, Elena Kuzmicheva, Sergei Mokurov, Svetlana Markova, Svetlana Pylova, Marina Isakova, Elena Shakurova, Vladimir Petrov, Natalya V. Dorshakova, Tatyana Karapetyan, Tatyana Varlamova, Jorma Ilonen, Minna Kiviniemi, Kristi Alnek, Helis Janson, Raivo Uibo, Tiit Salum, Erika von Mutius, Juliane Weber, Helena Ahlfors, Henna Kallionpää, Essi Laajala, Riitta Lahesmaa, Harri Lähdesmäki, Robert Molder, Viveka Öling, Janne Nieminen, Terhi Ruohtula, Outi Vaarala, Hanna Honkanen, Heikki Hyöty, Anita Kondrashova, Sami Oikarinen, Hermie J.M. Harmsen, Marcus C. De Goffau, Gjal Welling, Kirsi Alahuhta, Tuuli Korhonen, Suvi M. Virtanen, and Taina Öhman.