**Gut microbiota evolution during infancy: impact of introducing allergenic foods**

**Online Materials**

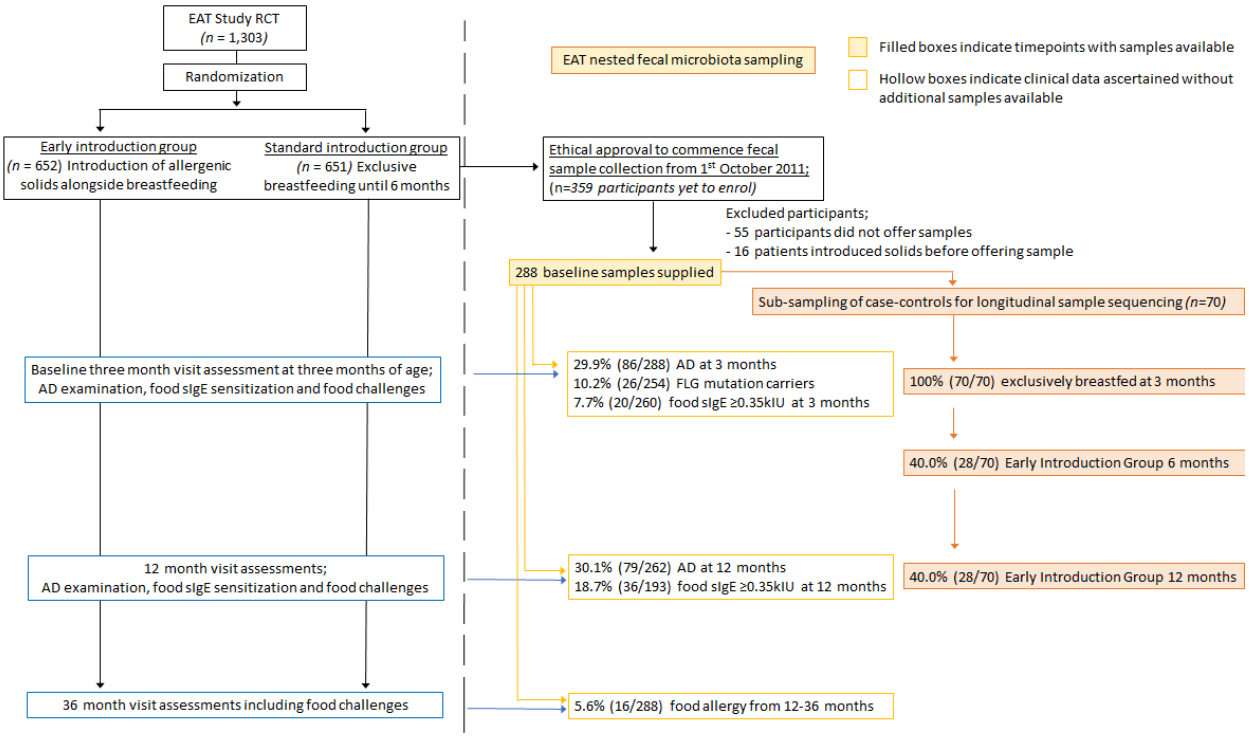


**FIG E1. Participant flow diagram.**

Of the potential participants (n=359) in the nested study, 288 participants provided a baseline (3 months) fecal sample. For the subset of case-control longitudinal sampling (n=70), participants provided additional fecal samples at 6 and 12 months.

FIG E2

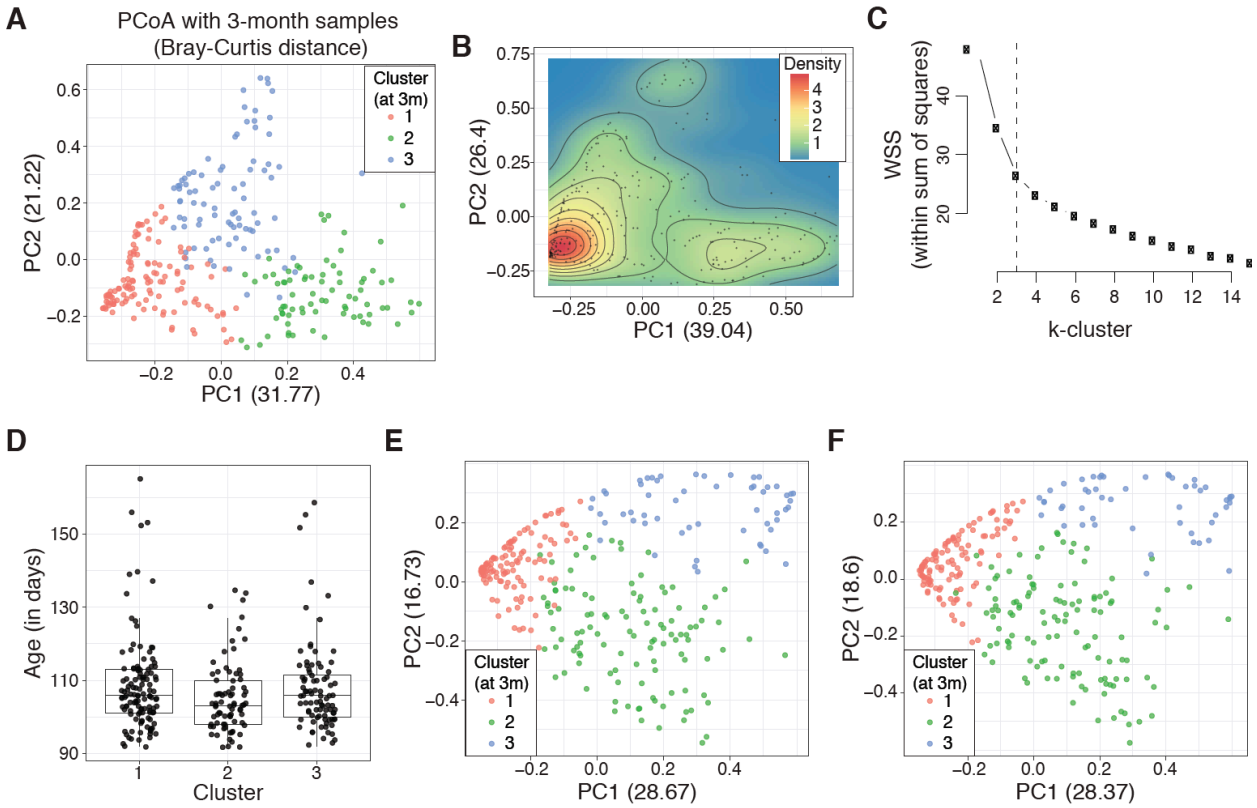


**FIG E2. Patterns of gut microbiota of baseline fecal samples**

A, PCoA of the gut microbiome at 3 months. Pair-wise distances (Bray-Curtis distance) among all samples were calculated and two major axes (PC1 and PC2) from the multi-dimensional distance space were calculated and depicted on a scatter plot. Colors indicate different clusters, according to k-means clustering.

B, Density distribution of three-month-old infants' microbiome based on PCoA (same as Fig 1A). There are three high-density peaks pronounced, suggesting 3 clusters of microbial community populations.

C, Evaluation of optimal clustering using 'within sum of squares' by each k (number of clusters). Based on the graphical visualization of within sum of squares or the 'elbow method', k=3 (3 clusters) is the optimal number of clusters. Increasing the number of clusters beyond 3 (k>3) results in overfitting, whereas k<3 is not as effective in minimizing intra-cluster variation.

D, Distribution of age (in days) of the infants when their 3 months samples were acquired, grouped by each cluster (p=NS; Kruskal-Wallis test).

E, PCoA of microbiome communities at 3 months with species-level classification. The colors indicate clusters.

F, PCoA of microbiome communities at 3 months with 97% *de novo* OTU level classification. The colors indicate clusters.

# FIG E3

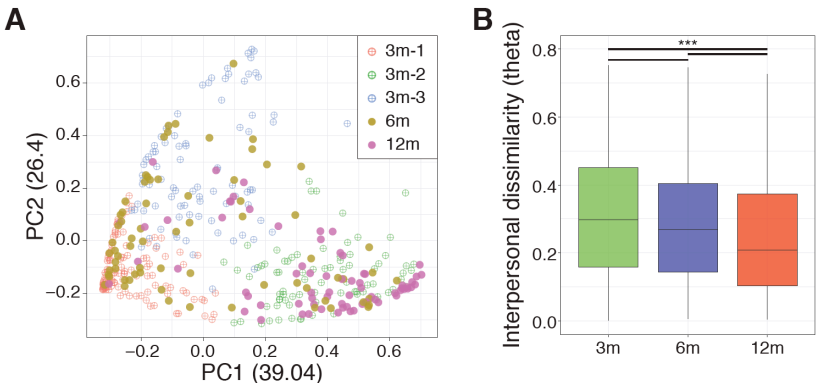


**FIG E3. Longitudinal changes of EAT gut microbiome**

A, Gut microbiota for 6 and 12 months samples were overlaid on the PCoA ordination of the microbiota of three-months samples (same as Fig 1A). The colors indicate clustering and age groups. At twelve months of age the microbiota largely converge to cluster #2 of the three-months gut microbiota.

B, Boxplot showing changes of interpersonal dissimilarity at different time points. All-to-all theta distances within each age group were calculated and plotted. Lower value (close to zero) indicates communities are more similar.

# FIG E4

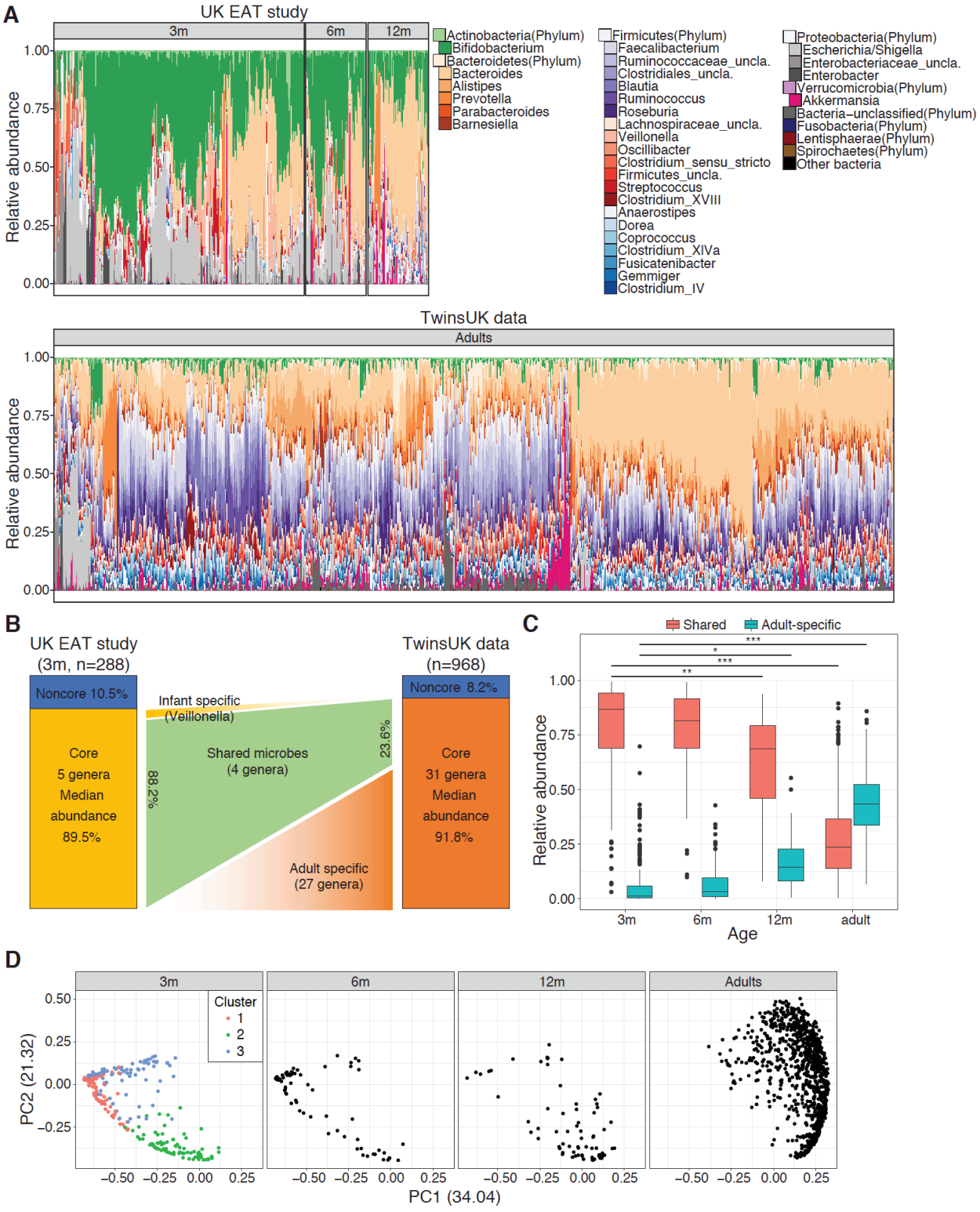
A
UK EAT study
3m
6m
12m
Relative abundance
1.00
0.75
0.50
0.25
0.00
Actinobacteria(Phylum)
Bifidobacterium
Bacteroidetes(Phylum)
Bacteroides
Alistipes
Prevotella
Parabacteroides
Barnesiella
Firmicutes(Phylum)
Faecalibacterium
Ruminococcaceae_uncla.
Clostridiales_uncla.
Blautia
Ruminococcus
Roseburia
Lachnospiraceae_uncla.
Veillonella
Oscillibacter
Clostridium_sensu_stricto
Firmicutes_uncla.
Streptococcus
Clostridium_XVIII
Anaerostipes
Dorea
Coprococcus
Clostridium_XIVa
Fusicatenibacter
Gemmiger
Clostridium_IV
Proteobacteria(Phylum)
Escherichia/Shigella
Enterobacteriaceae_uncla.
Enterobacter
Verrucomicrobia(Phylum)
Akkermansia
Bacteria_unclassified(Phylum)
Fusobacteria(Phylum)
Lentisphaerae(Phylum)
Spirochaetes(Phylum)
Other bacteria
TwinsUK data
Adults
Relative abundance
1.00
0.75
0.50
0.25
0.00
B
UK EAT study
(3m, n=288)
Noncore 10.5%
Core
5 genera
Median
abundance
89.5%
Infant specific
(Veillonella)
Shared microbes
(4 genera)
88.2%
23.6%
Adult specific
(27 genera)
TwinsUK data
(n=968)
Noncore 8.2%
Core
31 genera
Median
abundance
91.8%
C
Shared
Adult-specific
***
*
***
**
Relative abundance
1.00
0.75
0.50
0.25
0.00
3m
6m
12m
adult
Age
D
3m
6m
12m
Adults
Cluster
1
2
3
PC2 (21.32)
0.50
0.25
0.00
−0.25
−0.50
−0.25
0.00
0.25
PC1 (34.04)

**FIG E4. Microbiota differences between infants and adults**

A, Stacked bar charts for relative abundances of major genera in all samples, from both EAT (3, 6 and 12 months) and TwinsUK (adults) cohorts. Each bar indicates an individual.

B, Schematic diagram describing differences of core/non-core genera between three-month-old infants and adults. There are 5 core genera (genera exist more than 95% of individual, *Bifidobacterium*, *Bacteroides*, *Streptococcus*, *Escherichia*/*Shigella* and *Veillonella*) from three-month-old infants. Among them, 4 genera (except *Veillonella*) are also core in adults.

C, Boxplot showing relative changes of shared and adult-specific core microbes (*$p<0.05$, **$p<0.01$, ***$p<0.001$; Wilcoxon rank-sum test after Kruskal-wallis test). At 12 months, relative abundances of adult-specific core genera are increased and shared cores are decreased, compared to 3 months.

D, PCoA plot showing relative changes of gut microbiome during maturation. 12 months samples clustered more closely with adults.

# FIG E5

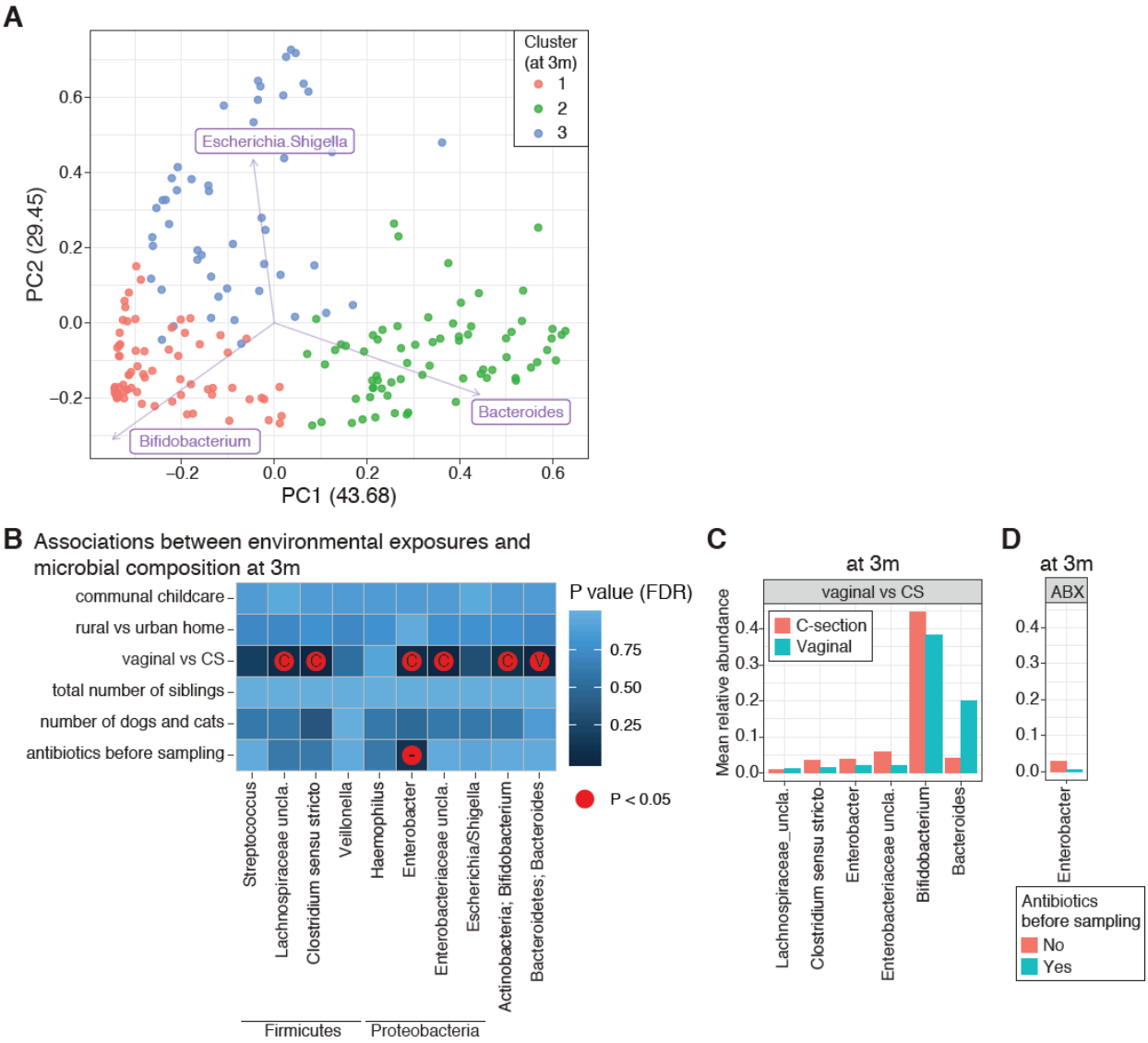


**FIG E5. Role environmental factors on microbial compositions.**

A, PCoA of baseline gut microbiota of vaginally born infants. Different colors indicate clusters and arrows indicate specific genera significantly correlated with PCoA ordination ($p<0.05$, lengths of arrows are proportional to $R^2$ (calculated by EnvFit in R)).

B, Heatmap showing associations between environmental exposures and microbial composition. Darker colors indicate lower P values and red marks indicate associations with FDR-corrected P values lower than 0.05.

C. Mean relative abundance differences of selected taxa from B (vaginal vs CS).

D. Mean relative abundance differences of selected taxa from B (Antibiotics before sampling).

# FIG E6

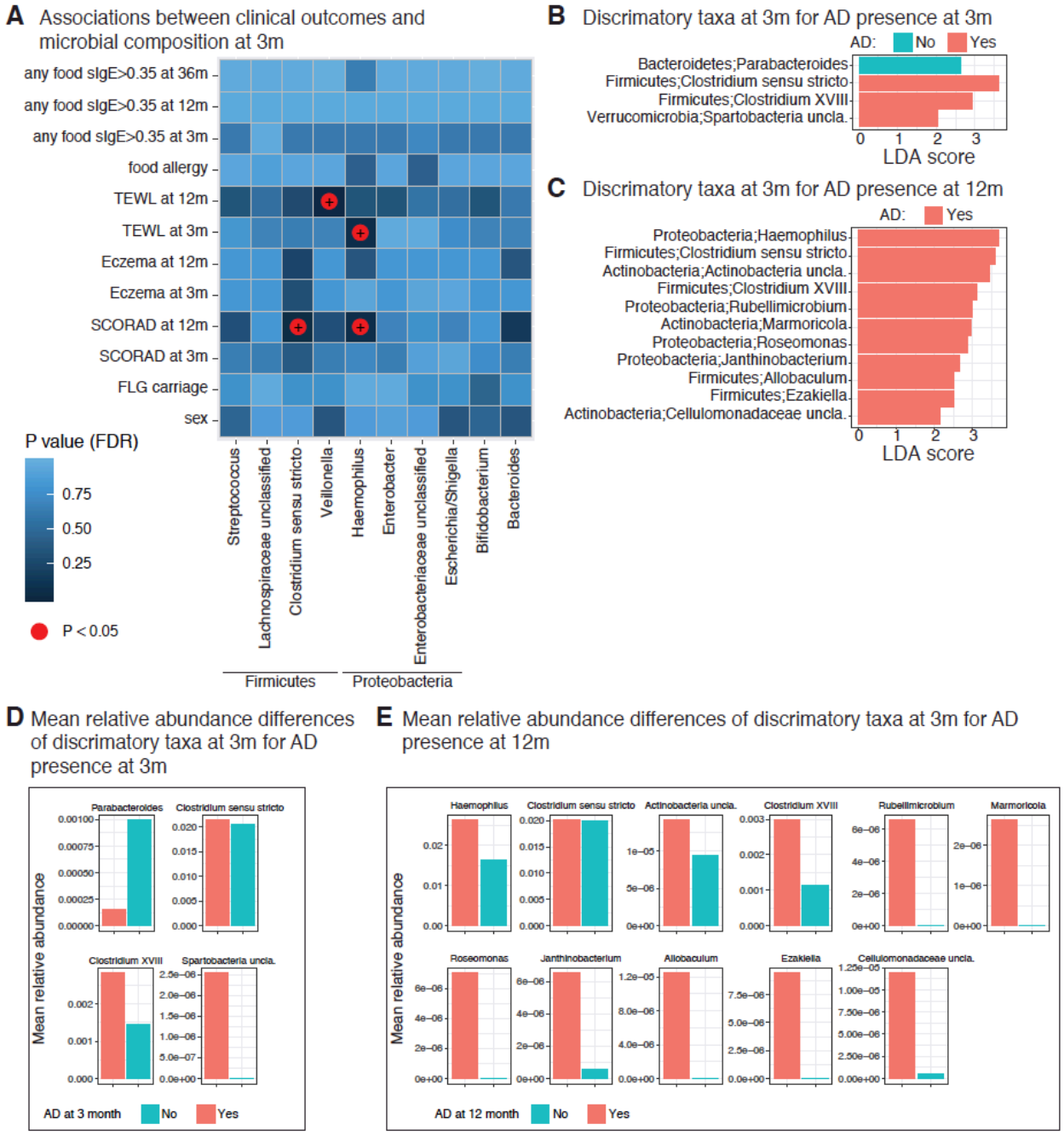


**FIG E6. Association of clinical measurements with microbial compositions.**

A, Heatmap showing associations between various clinical measurements and microbial composition. Darker colors indicate lower P values and red marks indicate associations with FDR-corrected P values lower than 0.05.

B, Discriminatory taxa from three-month gut microbiota in infants with or without eczema at three months of age (LDA score: Linear discriminant analysis score). LDA score is calculated by LEfSe(1).

C, Discriminatory taxa from three-month gut microbiota in infants with or without eczema at twelve months of age. LDA score is calculated by LEfSe.
D. Mean relative abundance differences of selected taxa from B (with or without eczema at three months).
E. Mean relative abundance differences of selected taxa from C (with or without eczema at twelve months).

FIG E7

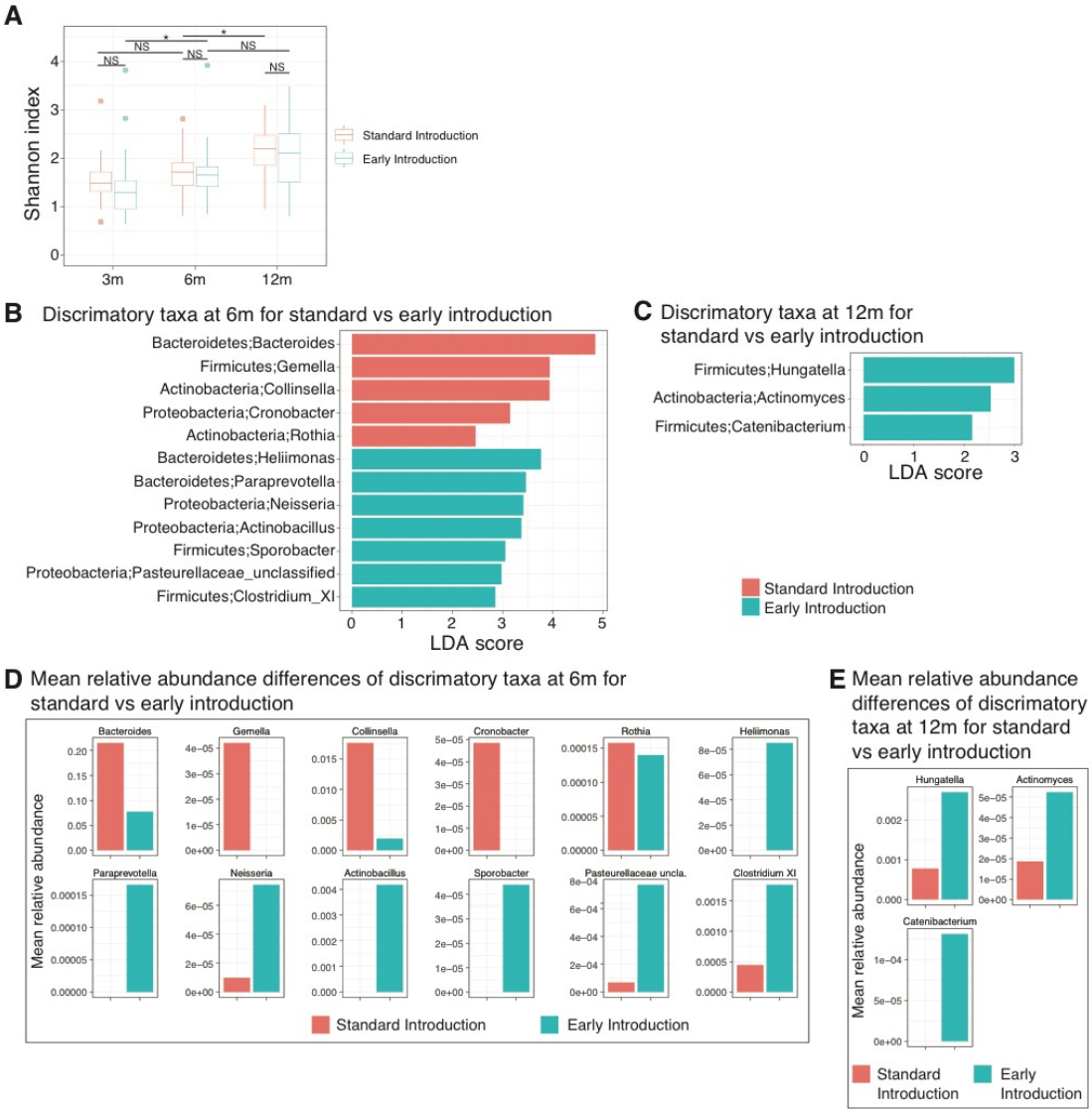


**FIG E7. Differential microbiota dynamics by early food introduction in EAT cohort**

A, Boxplot comparing Shannon diversity changes amongst participants' longitudinal samples according to randomized allocation to continued exclusive breastfeeding (standard introduction group) or the introduction of allergenic solids (early introduction group). (*p<0.05; paired Wilcoxon rank-sum test)

B,Discriminatory taxa from six-month gut microbiota of infants in standard versus early introduction (LDA score: Linear discriminant analysis score). LDA score is calculated by LEfSe.

C, Discriminatory taxa from twelve-month gut microbiota of infants in standard versus early introduction (LDA score: Linear discriminant analysis score). LDA score is calculated by LEfSe.

D, Mean relative abundance differences of selected taxa from B (standrard versus early introduction).

E, Mean relative abundance differences of selected taxa from C (standrard versus early introduction).

# FIG E8

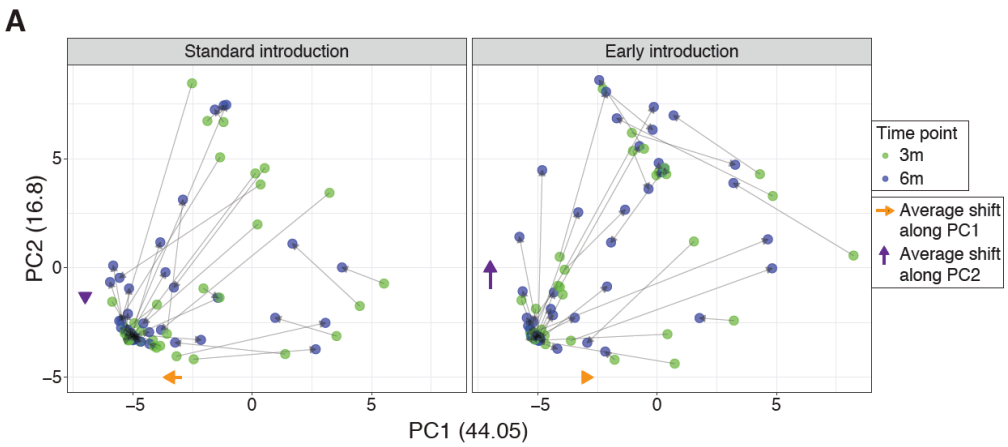


**FIG E8. Differential microbiota dynamics by early food introduction in TEDDY cohort**
A, PCoA scatter plot demonstrating longitudinal transition from 3 to 6 months of age of defined samples from the TEDDY cohort. From the TEDDY dataset samples from infants who were i) exclusively breastfed before solid food introduction and ii) longitudinal data (>2 sampling before 6 months of age) were selected. Grey arrows connect samples from the same individual. Yellow and purple arrows on the sides indicate the average shift of the microbiota in the PCoA axes.

***Supplementary appendix***

*Supplementary methods*

Chemical lysis was undertaken by adding 200µls of Guanidinium thiocyanate–EDTA–sarkosyl and 900µls of phosphate-buffered saline to the samples before they underwent bead-beating. Cell disruption was undertaken using Qiagen stainless steel beads and tungsten carbide beads on a Fastprep-24 Instrument (MP Biomedicals Europe, Illkirch, France) running at 6.5meters/second for 45 seconds. Two cycles of thermolysis followed by alternating incubation at 90ºC and -20ºC for 10 minutes each, before cell debris was pelleted by centrifugation at 13,000 g for 10 minutes. Supernatant was transferred to a fresh microfuge tube, where it was inverted with 140µl of 5 molar sodium chloride and 374µl 40% polyethylene glycol and precipitated for one hour at 4ºC. DNA was pelleted by centrifugation at 13,000 g for 10 min and resuspended in 500µl of sterile distilled water. 300µl Phenol/chloroform (1:1) was added and samples were vortexed before centrifugation at 13,000 g for 5 minutes. The upper phase was then transferred to a fresh microfuge tube. Total DNA was then precipitated by the addition of an equal volume of isopropanol and 0.1 volume of 10 molar ammonium acetate and stored at -20ºC for one hour. DNA was pelleted by centrifugation at 13,000 g for 10 minutes. Pelleted DNA was washed in 70% ethanol, dried and resuspended in 30µl of sterile distilled water (2, 3).

DNA extracts were amplified using universal bacterial primers targeting the 16S ribosomal RNA gene at hypervariable region V4 (515F-806R), which are tailed with sequences to incorporate Illumina (San Diego, CA) adapters and indexing barcodes (4). Sequencing was performed on the MiSeq instrument using version 2 chemistry and 250 cycles, stratifying the amplicon samples according to three, six and twelve month time points between each plate.

**References**


1. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. Genome Biol. 2011;12(6):R60.
2. Green GL, Brostoff J, Hudspith B, Michael M, Mylonaki M, Rayment N, et al. Molecular characterization of the bacteria adherent to human colorectal mucosa. J Appl Microbiol. 2006;100(3):460-9.
3. Rogers GB, Cuthbertson L, Hoffman LR, Wing PA, Pope C, Hooftman DA, et al. Reducing bias in bacterial community analysis of lower respiratory infections. ISME J. 2013;7(4):697-706.
4. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci U S A. 2011;108 Suppl 1:4516-22.