Compromised gut microbiota networks in children with anti-islet cell autoimmunity

Running title: Gut microbiome and anti-islet cell autoimmunity

David Endesfelder[1,2*], Wolfgang zu Castell[1*], Alexandria Ardissone[3], Austin G. Davis-Richardson[3], Peter Achenbach[2], Michael Hagen[1], Maren Pflueger[2], Kelsey A. Gano[3], Jennie R. Fagen[3], Jennifer C. Drew[3], Christopher T. Brown[3], Bryan Kolaczkowski[3], Mark Atkinson[4], Desmond Schatz[4], Ezio Bonifacio[5,6], Eric W. Triplett[3**], Anette-G. Ziegler[2**]

[1] Scientific Computing Research Unit, Helmholtz Zentrum München, Germany

[2] Institute of Diabetes Research, Helmholtz Zentrum München, and Forschergruppe Diabetes, Klinikum rechts der Isar, Technische Universität München, Germany

[3] Department of Microbiology and Cell Science, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, U.S.A.

[4]Department of Pediatrics, University of Florida, Gainesville, FL, U.S.A.

[5]Center for Regenerative Therapies, Dresden, and Paul Langerhans Institute Dresden, Technische Universität Dresden, Germany

[6]Institute for Diabetes and Obesity, Helmholtz Zentrum München, Germany

[*] these authors contributed equally

[**] these authors co-directed the project

Corresponding author contact information:

Prof. Dr. med. Anette-G. Ziegler

Diabetes Publish Ahead of Print, published online March 7, 2014

Institute of Diabetes Research, Helmholtz Zentrum München,

Ingolstaedter Landstr. 1, 85764 Neuherberg, Germany

Phone: 0049-89-3187 2896, Fax: 0049-89-3187 3144

E-mail: anette-g.ziegler@helmholtz-muenchen.de

Number of Figures: 4

Number of Tables: 0

Word count: 4000

Abstract

The gut microbiome is suggested to play a role in the pathogenesis of autoimmune disorders such as type 1 diabetes. Evidence of anti-islet cell autoimmunity in type 1 diabetes appears in the first years of life, however little is known regarding establishment of the gut microbiome in early infancy. Here, we sought to determine whether differences were present in early composition of the gut microbiome in children who developed anti-islet cell autoimmunity. We investigated the microbiome of 298 stool samples prospectively taken up to age 3 years from 22 case children who developed anti-islet cell autoantibodies, and 22 matched control children who remained islet autoantibody negative in follow-up. The microbiome changed markedly during the first year of life, and was further affected by breast-feeding, food introduction, and birth delivery mode. No differences between anti-islet cell autoantibody positive and negative children were found in bacterial diversity, microbial composition, or single genus abundances. However, substantial alterations in microbial interaction networks were observed at age 0.5 and 2 years in the children who developed anti-islet cell autoantibodies. The findings underscore a role of the microbiome in the pathogenesis of anti-islet cell autoimmunity and type 1 diabetes.

Introduction

Type 1 diabetes is the result of a complex interplay of genetic susceptibility (1) and environmental determinants leading to anti-islet cell autoimmunity against pancreatic islet beta cells and autoimmune beta cell destruction (2). Anti-islet cell autoimmunity precedes the clinical onset of type 1 diabetes and often develops within the first years of life (3). This suggests that early shaping of the immune system in children is critical for the initiation of autoimmunity (3). There is increasing evidence that the immune response is shaped by factors that include how the host establishes a stable ecosystem with a large cohort of accompanying bacteria (4-7). With this, the role of microbiota in type 1 diabetes pathogenesis has become an important subject of investigation (8-12). The largest community of bacteria is established in the gastrointestinal tract (13, 14) where beneficial host-bacteria interactions have been demonstrated for food degradation or pathogen defense (14-16).

Relatively few studies of the human gut microbiome have been performed in children less than 5 years old. These studies suggest that the phylogenetic composition of the bacterial communities evolves towards an adult-like configuration within the three-year period after birth (14, 17-19). Hence, it is conceivable that the evolution of the microbiome in infancy could influence the risk of anti-islet cell autoimmunity in susceptible children. Indeed, studies from Finland have provided evidence for this hypothesis (10, 20). The aim of our study was to investigate gut bacterial community structures during the early period from birth to the age of 3 years from the perspective of complex interaction networks. We estimated interaction on the basis of co-variation of bacterial abundances to compare children who developed anti-islet cell autoantibodies with children who did not develop such autoantibodies. We took advantage of the prospective BABYDIET study (21) where infants at increased risk of type 1 diabetes were monitored from birth for the development of anti-islet cell autoantibodies and type 1 diabetes. The gut microbiome composition was estimated based on measurements of

16S rRNA gene sequences from fecal samples that were obtained at 3 months intervals up to the age of 3 years. Analyses were focused on bacterial diversity, community composition, individual bacterial species and microbial interaction networks. Results show that complex bacterial interaction networks, rather than single genera, appear to be relevant to early pre-clinical type 1 diabetes.

Research Design and Methods

BABYDIET study material

Analysis of microbiota was performed on 298 stool samples from 44 children participating in the BABYDIET study (21). The BABYDIET study randomized 150 infants with a first degree relative with type 1 diabetes and with the type 1 diabetes risk HLA genotypes DR3/4-DQ8 or DR4/4-DQ8 or DR3/3 to gluten exposure at 6 months or at 12 months of age. The intervention had no effect on anti-islet cell autoimmunity outcome. Blood and stool samples were collected at 3 month intervals from age 3 to 36 months and subsequently at 6 month intervals. Anti-islet cell autoantibodies (i.e., autoantibodies to insulin, GAD, insulinoma-associated antigen-2, and zinc transporter 8) were measured at each study visit. Written informed consent was obtained from the parents. The study was approved by the ethics committee of the Ludwig-Maximilian-University, Munich, Germany (Ethikkommission der Medizinischen Fakultät der Ludwig-Maximilians Universität No. 329/00).

Stool samples chosen for the study included 147 samples from the 22 BABYDIET cohort children who developed persistent anti-islet cell autoantibodies at a median age of 1.54 years (IQR: 0.90 years and maximum 2.45 years), and 151 samples from 22 children who remained anti-islet cell autoantibody negative, and were matched for date of birth. Of the 22 children with persistent islet autoantibodies, 15 had developed persistent multiple islet autoantibodies, and 10 developed diabetes after a median follow-up of 5.3 years. For the 44 children, stool samples were taken from age 0.24 to 3.2 years with an average of 6.8 probes per child (Supplementary Table 1).

Sample processing and deep sequencing

Stool samples were collected at home and shipped by express courier overnight to the clinical study center where they were processed and immediately frozen at -80°C. DNA was extracted

from the stool samples as described previously (10). Bacterial 16S rRNA genes present within fecal samples were amplified using the primers 515F and 806R (22) modified with a sample specific barcode sequence and Illumina adapter sequences.

PCR was performed at an initial denaturation temperature of 94^{o}C for 3 min, followed by 20 cycles of 94^{o}C for 45 sec, 50^{o}C for 30 sec, and 65^{o}C for 90 sec. A final elongation step at 65^{o}C was run for 10 min. PCR products were purified using the Qiagen™ PCR purification kit following the manufacturer's protocol. Illumina high throughput sequencing of 16S rRNA genes was conducted as described (23). Illumina sequencing was performed with 101 cycles each. Sequences were trimmed based on quality scores using a modified version of Trim2 (24) and the first 11 bases of each paired read were removed to eliminate degenerate bases derived from primer sequences. The prokaryotic database (25) used for 16S rRNA gene analysis was formatted using TaxCollector (26). Sequences were compared to the TaxCollector-modified RDP database using CLC Assembly Cell version 3.11 utilizing the paired reads and global alignment parameters. Two parameters were used in this step, a 98% length fraction and similarity values dependent on the desired taxonomic level, i.e., 80% at Domain/Phylum, 90% to Class/Order/Family, 95% to Genus levels (27). Pairs that matched different references at the species level were classified at the lowest common taxonomic level. Unresolved pairs were discarded. Henceforth, successfully paired reads are referred to as reads.

Confounding variables

Data on breastfeeding (yes, no), the duration of breastfeeding (weeks) and the introduction of solid food (gluten free and gluten-containing cereals, vegetables, fruits), were analysed from daily food records as previously described (21). Data on Caesarean section (yes, no) were obtained from obstetric records.

Statistical analysis

Shannon evenness and Chao richness indices were estimated at genus level as described (28, 29). To correct bacterial diversity for the influence of confounding factors, stepwise multiple regression was performed with diversity as dependent variable. Age, breast-feeding at sampling time, introduction of solid food, first gluten exposure and delivery by Caesarean section were used as confounding variables. Akaike's Information Criterion (AIC) (30) was used in the stepwise regression procedure to select confounding factors associated to diversity. To avoid bias due to violations of normality, rank regression (31) was used to estimate p-values of the regression coefficients corresponding to confounding factors associated with diversity. The R package fields (32) was used for cubic spline regression of age versus Shannon evenness. Chao richness was corrected for the influence of Caesarean section by using the residuals of a regression model with richness as dependent and Caesarean section as independent variable. Diversity analyses were performed on the entire age range and after grouping reads into three age classes of 0.5±0.25, 1.0±0.25 and 2.0±0.5 years. At most one single probe closest to 0.5, 1 and 2 years was used, respectively, for each child.

For further analyses, phyla and genera with less than 0.01% abundance within the total number of reads were neglected. This reduced the number of genera from 452 to 75 and the number of phyla from 21 to 8. For the analysis of bacterial community compositions, Bray-Curtis distances (33) were estimated on Hellinger transformed data (34). Differences in community compositions were tested with the non-parametric Multivariate Analysis of Variance (npMANOVA) (35) method. To visualize the results, Principal Coordinate Analysis (PCoA) was performed. Relative abundances of individual phyla and genera were compared by Wilcoxon-Mann-Whitney tests. To account for heterogeneity in variances, Brunner-Munzel tests (36) were used for bacteria where Bartlett's test (37) showed evidence for unequal variances ($P<0.05$). Second, we adjusted for confounding factors by using the

residuals of stepwise AIC models with bacterial abundance as dependent variable and the confounding variables as independent variables. For all independent variables with a p-value <0.1, the model was again estimated and the resulting residuals were used as adjusted abundance values. P-values were corrected for multiple testing with the Benjamini&Hochberg method (38).

Correlation based networks reflecting co-variations of bacterial abundances were used as a surrogate for bacterial interaction networks. To construct interaction networks, we computed Spearman's rank correlation ρ for all possible pairs of genera. We used 1000 random permutations and set an edge, if $P<0.05$ and $\rho>0.3$, considering positively correlated genera. Networks were plotted with the Fruchterman-Reingold (39) algorithm. Eigenvector centrality was estimated as described (40), and Kolmogorov-Smirnov tests were used to test for differences in distributions (41, 42). Differences in the number of isolated nodes were analysed by comparing the number of nodes of degree ≤ 1. To test whether the observed differences were due to lower sample size in children who became anti-islet cell autoantibody positive, networks were estimated with all combinations of N autoantibody negative samples, where N denotes the number of autoantibody positive samples. All statistical analyses were performed with the statistical software R version 2.15.2.

Results

Anti-Islet Cell Autoantibodies and Diversity of the Gut Microbiome

Bacterial diversity was analyzed for 452 bacterial genera. Diversity can be described by its evenness and richness. Evenness measures the similarity of proportions of taxa within a community, while richness represents the number of taxa in the community. We first analyzed covariates for Shannon evenness and Chao richness via stepwise regression models for all 298 stool probes. We observed an association of evenness with age (P=0.025), Caesarean section (P=0.0026), gluten exposure (P=0.0095), and an association of richness with Caesarean section (P=0.0025). Cubic spline regression of evenness with age showed that evenness increased until age 2 years and saturated for older children (Figure 1A). In contrast, richness remained constant over time (Figure 1B). We found no association of anti-islet cell autoantibody positive/negative status with evenness (P=0.27; Figure 1C) and richness (P=0.56; Figure 1D). Richness and evenness were also not different after adjustment for associated covariates (Shannon evenness, P=0.62; Chao richness, P=0.40).

Supporting the association of bacterial evenness with increasing age, the analysis of the distribution of all 21 phyla revealed a considerable shift between age 6 months and 1 year (Figure 1E). This shift was primarily due to an increase in the relative abundance of *Firmicutes* in both the autoantibody positive (P= 8.7×10^{-6}) and negative (P= 0.016) groups. The distribution of phyla remained relatively constant between age 1 year and 2 years.

To account for the effect of age and for sample dependent colinearities, the data were grouped into three age intervals: 0.5+/-0.25 years (anti-islet cell autoantibody positive: N=19; autoantibody negative: N=21), 1+/-0.25 years (anti-islet cell autoantibody positive: N=16; autoantibody negative: N=19) and 2+/-0.5 (anti-islet cell autoantibody positive: N=18; autoantibody negative: N=20) years. At each age interval, no differences between anti-islet

cell autoantibody positive and negative children were observed for evenness ($P_{0.5}$=0.22, P_1= 0.83, P_2=0.29, Supplementary Figure 1A-C) and richness ($P_{0.5}$=0.12, P_1=0.1, P_2=0.63, Supplementary Figure 1D-F).

Anti-Islet Cell Autoantibodies and Bacterial Community Composition

Differences in bacterial community composition were tested by comparing the intra-group distances of bacterial abundances between case and control children based on Bray-Curtis distances (33) estimated on the 75 genera that remained after filtering bacteria with low abundances (<0.01%). Single variable and multivariable npMANOVA (35) models, including the covariates age at sampling time, Caesarean section, duration since solid food introduction, duration since introduction of gluten and breast feeding at sampling time were applied for each of the three age intervals.

The covariates with the strongest effects on bacterial community composition at age 0.5 years were breast-feeding (P=0.002, Supplementary Table 2) and the duration since first solid food introduction (P=0.001, Supplementary Table 2). At age 1 year, only duration since first solid food introduction (P=0.049, Supplementary Table 2) was associated with bacterial community composition and at age 2 years the effects of nutrition vanished. Children who became anti-islet cell autoantibody positive did not show significant differences in community composition in univariable ($P_{0.5}$=0.52, P_1=0.36 and P_2=0.35) and multivariable npMANOVA ($P_{0.5}$=0.20, P_1=0.38 and P_2=0.18, Supplementary Table 2) analysis for all of the three analyzed age intervals. PCoA plots did not show a noticeable clustering of anti-islet cell autoantibody positive and negative children (Figure 2).

Anti-Islet Cell Autoantibodies and Bacterial Abundances

Abundances at the phylum and genus level were assessed at age 0.5, 1 and 2 years. None of the 8 analyzed phyla differed in bacterial abundances between anti-islet cell autoantibody

positive and negative children. Of the 75 analyzed genera, *Dorea* and *Barnesiella* abundances at age 0.5 years (P=0.003 and P=0.035), *Candidatus Nardonella* abundances at age 1 year (P=0.031), *Erwinia* and *Enterobacter* abundances at age 2 years (P=0.024 and P=0.045) differed between anti-islet cell autoantibody positive and negative children (Supplementary Tables 3-5). None of these were significant after correction for multiple testing. None of the genera showed a persistent difference between anti-islet cell autoantibody positive and negative children across all three age groups.

Since nutrition affected bacterial composition in our cohort, we also compared bacterial abundances between children who developed anti-islet cell autoantibodies and children who remained autoantibody negative after adjustment for confounding factors. None of the phyla abundances were significantly different after the adjustment. Some additional genera showed differences after adjusting for confounding factors (Supplementary Tables 3-5). These include *Veillonella* abundances which were lower in children who developed anti-islet cell autoantibodies (average 3%) than in children who remained autoantibody negative (average 10%, P=0.0098) at age 0.5 years and *Enterococcus* abundances which were higher in children who developed anti-islet cell autoantibodies (average 3%) than in children who remained autoantibody negative (average 0.8%, P=0.00011) at age 0.5 years. Again, these differences were not significant after correction for multiple comparisons.

Anti-Islet Cell Autoantibodies and Bacterial Interaction Networks

Since the gut microbiome constitutes an ecosystem, where bacteria depend on each other and in particular compete for nutrition, we hypothesized that a functional interplay of bacteria is crucial for the development of the gut microbiome, and that differences in the interaction between bacteria might be associated with the development of anti-islet cell autoimmunity. We therefore used microbial correlation networks at the genus level (N=75) as an approximation for bacterial interactions using two different scores: eigenvector centrality and

the number of isolated nodes. Correlation based bacterial interaction networks were estimated at age 0.5, 1 and 2 years for the anti-islet cell autoantibody positive and negative groups.

Eigenvector centrality is a measure for the relative importance and the connectivity of each node in a network (40). Eigenvector centrality of a node accounts for centrality of its neighbors assuming that a node is more central if the surrounding neighbors also have high centrality (43). Differences in eigenvector centrality indicate that the information flow varies throughout the network. The networks of children who became anti-islet cell autoantibody positive showed significantly different centrality distributions at age 0.5 years (P=0.0024; Figure 3A, D and G) and again at age 2 years (P=0.013; Figure 3C, F and I). Most of the genera that had high centrality at age 0.5 years had also high centrality at age 2 years for children who became autoantibody positive (88%) and children who remained autoantibody negative (77%). No differences were observed between the two groups at age 1 year (Figure 3B, E and H).

In both groups, a cluster of nodes had high eigenvector centralities (EC>0.5). In contrast to the autoantibody negative group, more nodes with intermediate levels of eigenvector centrality (0.05<EC<0.5) were found in the anti-islet cell autoantibody positive group (Figure 3). At age 0.5 years, the bacterial genera *Enterococcus*, *Sarcina*, *Prevotella* and *Corynebacterium* showed high centrality in networks of children who became anti-islet cell autoantibody positive and low (EC<0.05) centrality in networks of children who remained autoantibody negative (Supplementary Figure 2A). A detailed overview of eigenvector centrality values at age 1 year can be found in Supplementary Figure 2B. At age 2 years, *Barnesiella* and *Candidatus Nardonella* showed high centrality in the autoantibody positive and low centrality in the autoantibody negative group (Supplementary Figure 2C). In contrast, *Staphylococcus* and *Nocardioides* had high centrality in the autoantibody negative group and low centrality in the autoantibody positive group (Supplementary Figure 2C).

While eigenvector centrality measures the capacity of overall information flow, node degree measures the connectivity in a topological sense. In the following we refer to genera with node degree ≤ 1 as isolated nodes. More isolated bacterial genera (Figure 3A-F) were found in children who developed anti-islet cell autoantibodies at age 0.5 years (P=0.00012) and 2 years (P=0.0044, Figure 4). A detailed overview of node degrees of all genera in the bacterial networks for anti-islet cell autoantibody positive and negative children is shown in Supplementary Figure 3. Sample sizes differed slightly between anti-islet cell autoantibody positive ($N_{0.5}$=19, N_1=16, N_2=18) and negative ($N_{0.5}$=21, N_1=19, N_2=20) children in the three age groups. We therefore performed interaction network estimates for all possible equal number subsets of children who remained autoantibody negative. At age 0.5 and at age 2 years, there was no single combination of children who remained autoantibody negative that resulted in a similar high number of isolated bacterial genera as observed for children who became anti-islet cell autoantibody positive. No differences in the number of isolated bacterial genera between the two groups were observed at age 1 year.

Discussion

In murine models, associations between gut microbiome composition and type 1 diabetes or anti-islet cell autoimmunity have been found (11, 44-46). Little is known regarding the early establishment of the gut microbiome in children who develop anti-islet cell autoimmunity. In this study, stool samples from children participating in the prospective BABYDIET cohort were analysed within the first 3 years of life to compare bacterial diversity, composition, individual phyla and genera abundances and interaction networks for children who became anti-islet cell autoantibody positive to those of children who remained autoantibody negative. No differences in bacterial diversity or community composition were observed between autoantibody positive and negative children. After correction for multiple testing, there were no individual bacterial genera that showed significantly different abundances between anti-islet cell autoantibody positive and negative children. However, children who became anti-islet cell autoantibody positive showed significantly different distributions of eigenvector centrality in correlation based interaction networks of bacterial communities, and their networks consisted of more isolated nodes than those of children who remained autoantibody negative.

The strengths of our study lies in the relatively large number of stool samples analysed, the early and sequential sample collection, and the homogenous cohort of first degree relatives of individuals with type 1 diabetes with similar type 1 diabetes high-risk HLA DR-DQ genotypes. Furthermore, case and control children were matched by date of birth so that stool samples were collected within the same year, season, and under similar study conditions between groups. Although the control islet autoantibody negative children in our study are well matched to the case children, they are enriched for type 1 diabetes susceptibility genes and may therefore not be the most suitable controls. We have not examined the microbiome in children without an enriched genetic susceptibility, and it is

possible that our findings may not represent the microbiome status of children from the general population. The majority of the islet autoantibody positive children in our study developed multiple islet autoantibodies, a status that confers extreme risk for diabetes (47). However, findings may differ if only cases that subsequently developed diabetes were analyzed. Further limitations of the study include that samples were collected at home with overnight transport at room temperature, and data on other potential confounding factors such as infections or antibiotic therapy was not available for this analysis.

Two other studies from Finland have investigated the role of the gut microbiome in children who developed anti-islet cell autoantibodies (10, 20). One study investigating four children who seroconverted and an equivalent number who remained anti-islet cell autoantibody negative found a lower bacterial diversity and differences in the abundances of the phyla *Bacteroidetes* and *Firmicutes* in children with anti-islet cell autoantibodies before, at, and after islet autoantibody seroconversion (10). At the genus level, the same study found differences in *Eubacterium* and *Faecalibacterium* abundances, and reported that community composition was more similar in children who remained anti-islet cell autoantibody negative compared to children who became anti-islet cell autoantibody positive (10). A second study from Finland compared the gut microbiome of 18 children who developed anti-islet cell autoimmunity with 18 children who did not (20). In contrast to our study, the anti-islet cell autoantibody positive children in the study of Goffau et al. (20) were already autoantibody positive at the time of sampling and the probands were older. The authors reported significant differences in the abundance of the phylum *Bacteroidetes*, the genus *Bacteroides* and several bacteria on species level (20). In addition, a trend of increased bacterial diversity in anti-islet cell autoantibody negative children was reported (20). We did not find these reported differences between anti-islet cell antibody positive and negative children in our cohort. There was no single phylum that showed differences between anti-islet cell autoantibody positive and negative children. The reported differences in the abundances of the genera

Faecalibacterium, *Eubacterium* (10) and *Bacteroides* (20) were also not observed in the data presented here. The deviating results may be explained by differences in sample size (10), different study design used in (20), and/or geographical differences between the German and Finnish children. Finally, a recent study found differences in the abundances of several bacteria, including *Prevotella*, *Clostridium*, *Veillonella*, *Bifidobacterium*, *Lactobacillus* and *Bacteroides* in the microbiome of children with established diabetes as compared to healthy control children (48).

Nutrition has an important effect on the early microbial community. For example, exclusively breast-fed infants have different distributions of bacteria than formula fed infants (49). In line with these data, our analysis of microbial community composition revealed that early microbiome composition is associated with breast-feeding duration and the age of introduction of solid food. Caesarean section was also found to be associated with bacterial abundances at the genus level. These and other confounders should be considered when analyzing bacterial abundances in young children. We analyzed abundances with and without adjustment for such confounders. Although some differences in the genera were observed between children who did and who did not develop anti-islet cell autoantibodies, most of the significant genera had low abundances and none of the genera was significant after multiple testing corrections.

We hypothesized that instead of individual bacterial abundances, the interplay between bacteria might be compromised in children who became anti-islet cell autoantibody positive and that differences in the interaction between bacteria might be associated with the development of anti-islet cell autoimmunity. The estimation of microbial co-occurrence networks was recently successfully applied to a large microbiome dataset from different body sites (50). This encouraged us to use co-variation of microbial abundances as a surrogate for bacterial interaction and to compare the networks of children who became anti-islet cell

autoantibody positive with the networks of children who did not. An increased number of isolated nodes can cause a reduced number of possible communication paths and therefore impair the flexibility of the network and the adaptability of the bacterial community. Interestingly, we found significantly higher numbers of isolated nodes in children who became anti-islet cell autoantibody positive. In addition, we observed significant differences in the distributions of eigenvector centrality suggesting differences in the information flow between bacteria. Differences were observed at age 6 months and 2 years but not at age 1 year. Since many of the children changed from breast-feeding to solid food and we found the most pronounced shift of large-scale bacterial distributions between 6 months and 1 year, we suspect that strong nutritional effects may mask anti-islet cell autoimmunity associations with bacterial networks around age 1 year.

To our knowledge, this is the largest study of the early gut microbiome in children developing autoimmunity. Potentially relevant findings in relation to anti-islet cell autoantibodies did not appear to be focused on individual microbiota, but on their connectivity. Moreover, the gut microbiome at an early age was strongly influenced by factors such as delivery mode, fundamental changes in nutrition, and the shift to an adult like microbiome. We suggest that a systemic view is necessary to understand the complex relationship between the development of type 1 diabetes, the environment and the gut microbiome.

Author contribution: DE, WzC and MH performed data analysis and drafted the manuscript. PA and MP contributed to data collection, islet autoantibody measurement, interpretation and data analysis. EB contributed to study design and data analysis. AA, ADR, KAG, JRF, JCD, CTB, BK, and EWT performed 16-S sequencing and data analysis. DS and MA contributed to data analysis, data interpretation, and manuscript writing. AGZ is principal investigator of BABYDIET study and contributed to data collection, analysis, and manuscript writing. AGZ, EWT, EB, DS, and MA contributed to the design of this study. All authors critically reviewed and approved the manuscript.

Acknowledgement:

This work was supported by grants from the Juvenile Diabetes Research Foundation (17-2012-16 and 17-2011-266), Deutsche Forschungsgemeinschaft (DFG ZI-310/14-1 to -4) and the foundation “Children With Type 1 Diabetes” (Stiftung Das Zuckerkranke Kind). Ezio Bonifacio is supported by the DFG Research Center and Cluster of Excellence - Center for Regenerative Therapies Dresden (FZ 111). The authors thank Sandra Hummel (Helmholtz Zentrum München), Christiane Winkler (Helmholtz Zentrum München), Annette Knopff (Helmholtz Zentrum München), and Melanie Bunk (Helmholtz Zentrum München) for data management and clinical assistance. No conflicts of interest have to be disclosed. Anette-G. Ziegler is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Figure Legends

Figure 1: Association of bacterial diversity with age and the development of anti-islet cell autoantibodies. (**A**) Association of age with Shannon evenness calculated for all 452 genera. Each dot in the scatterplot represents one stool probe and the red line was estimated by a cubic spline regression. (**B**) Association of age with Chao richness calculated for all 452 genera. The red line was estimated by a cubic spline regression. (**C**) Comparison between 147 stool samples of children who became anti-islet cell autoantibody positive (anti-islet aAb+) and 151 stool samples of children who remained autoantibody negative (anti-islet aAb-) for Shannon evenness. (**D**) Comparison between 147 stool samples of children who became anti-islet cell autoantibody positive and 151 stool samples of children who remained autoantibody negative for Chao richness. P-values in (**C**) and (**D**) were obtained by two-sided Wilcoxon-Mann-Whitney tests. (**E**) Distribution of 21 phyla for children who became anti-islet cell autoantibody positive and children who remained autoantibody negative for three age groups.

Figure 2: Bacterial community composition for children who developed anti-islet cell autoantibodies and children who remained autoantibody negative. (**A-C**) Results of PCoA and univariable npMANOVA analyses comparing intra-group Bray-Curtis distances of bacterial abundances between case and control children for three age classes. Probes of children who became anti-islet cell autoantibody positive (anti-islet aAb+) are labeled in red and probes of children who remained autoantibody negative (anti-islet aAb-) are labeled in blue. For each child and each time interval at most one probe was used.

Figure 3: Bacterial interaction networks for children who developed anti-islet cell autoantibodies and children who remained autoantibody negative. Bacterial networks of genera are shown for children who became anti-islet cell autoantibody positive (anti-islet aAb+) (**A-C**) and children who remained autoantibody negative (anti-islet aAb-) (**D-F**) for three different age classes. Each node represents one of the 75 analyzed genera. Nodes with

high eigenvector centrality (EC≥0.5) are labeled in yellow, nodes with intermediate eigenvector centrality (0.05<EC<0.5) in green and nodes with low eigenvector centrality (EC≤0.05) in blue. Isolated nodes are labeled with red stars. (**G-I**) Comparison of cumulative eigenvector centrality distributions between children who became anti-islet aAb+ and children who remained anti-islet aAb- for three different age classes. P-Values to test for differences in cumulative distributions were obtained by one-sided Kolmogorov-Smirnov tests.

Figure 4: Percentage of isolated nodes for children who developed anti-islet cell autoantibodies and children who remained autoantibody negative. Percentage of isolated nodes for children who became anti-islet cell autoantibody positive (anti-islet aAb+) and children who remained autoantibody negative (anti-islet aAb-) for three different age classes. P-values were obtained by two-sided Fisher's exact tests.

References

1. Ziegler AG, Nepom GT. Prediction and pathogenesis in type 1 diabetes. Immunity, 2010. 32(4): p. 468-78.

2. Bonifacio E, Warncke K, Winkler C, Wallner M, Ziegler AG. Cesarean section and interferon-induced helicase gene polymorphisms combine to increase childhood type 1 diabetes risk. Diabetes, 2011. 60(12): p. 3300-6.

3. Ziegler AG, Bonifacio E, and B.-B.S. Group. Age-related islet autoantibody incidence in offspring of patients with type 1 diabetes. Diabetologia, 2012. 55(7): p. 1937-43.

4. Cerf-Bensussan N, Gaboriau-Routhiau V. The immune system and the gut microbiota: friends or foes? Nat Rev Immunol, 2010. 10(10): p. 735-44.

5. Hooper LV, Littman DR, Macpherson AJ. Interactions between the microbiota and the immune system. Science, 2012. 336(6086): p. 1268-73.

6. Kranich J, Maslowski KM, Mackay CR. Commensal flora and the regulation of inflammatory and autoimmune responses. Semin Immunol, 2011. 23(2): p. 139-45.

7. Maslowski KM, Mackay CR. Diet, gut microbiota and immune responses. Nat Immunol, 2011. 12(1): p. 5-9.

8. Atkinson MA, Chervonsky A. Does the gut microbiota have a role in type 1 diabetes? Early evidence from humans and animal models of the disease. Diabetologia, 2012. 55(11): p. 2868-77.

9. Brown CT, Davis-Richardson AG, Giongo A, Gano KA, Crabb DB, Mukherjee N, Casella G, Drew JC, Ilonen J, Knip M, Hyoty H, Veijola R, Simell T, Simell O, Neu J, Wasserfall CH, Schatz D, Atkinson MA, Triplett EW. Gut microbiome metagenomics analysis suggests a functional model for the development of autoimmunity for type 1 diabetes. PLoS One, 2011. 6(10): p. e25792.

10. Giongo A, Gano KA, Crabb DB, Mukherjee N, Novelo LL, Casella G, Drew JC, Ilonen J, Knip M, Hyoty H, Veijola R, Simell T, Simell O, Neu J, Wasserfall CH, Schatz D, Atkinson MA, Triplett EW. Toward defining the autoimmune microbiome for type 1 diabetes. ISME J, 2011. 5(1): p. 82-91.

11. Roesch LF, Lorca GL, Casella G, Giongo A, Naranjo A, Pionzio AM, Li N, Mai V, Wasserfall CH, Schatz D, Atkinson MA, Neu J, Triplett EW. Culture-independent identification of gut bacteria correlated with the onset of diabetes in a rat model. ISME J, 2009. 3(5): p. 536-48.

12. Vaarala O. The gut as a regulator of early inflammation in type 1 diabetes. Curr Opin Endocrinol Diabetes Obes, 2011. 18(4): p. 241-7.

13. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial community variation in human body habitats across space and time. Science, 2009. 326(5960): p. 1694-7.

14. Savage DC. Microbial ecology of the gastrointestinal tract. Annu Rev Microbiol, 1977. 31: p. 107-33.

15. Ley RE, Peterson DA, Gordon JI. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell, 2006. 124(4): p. 837-48.

16. Walter J, Ley R. The human gut microbiome: ecology and recent evolutionary changes. Annu Rev Microbiol, 2011. 65: p. 411-29.

17. Dominguez-Bello MG, Blaser MJ, Ley RE, Knight R. Development of the human gastrointestinal microbiota and insights from high-throughput sequencing. Gastroenterology, 2011. 140(6): p. 1713-9.

18. Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. Development of the human infant intestinal microbiota. PLoS Biol, 2007. 5(7): p. e177.

19. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI. Human gut microbiome viewed across age and geography. Nature, 2012. 486(7402): p. 222-7.

20. de Goffau MC, Luopajarvi K, Knip M, Ilonen J, Ruohtula T, Harkonen T, Orivuori L, Hakala S, Welling GW, Harmsen HJ, Vaarala O. Fecal microbiota composition differs between children with beta-cell autoimmunity and those without. Diabetes, 2013. 62(4): p. 1238-44.

21. Hummel S, Pfluger M, Hummel M, Bonifacio E, Ziegler AG. Primary dietary intervention study to reduce the risk of islet autoimmunity in children at increased risk for type 1 diabetes: the BABYDIET study. Diabetes Care, 2011. 34(6): p. 1301-5.

22. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. 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: p. 4516-22.

23. Fagen JR, Giongo A, Brown CT, Davis-Richardson AG, Gano KA, Triplett EW. Characterization of the Relative Abundance of the Citrus Pathogen Ca. Liberibacter Asiaticus in the Microbiome of Its Insect Vector, Diaphorina citri, using High Throughput 16S rRNA Sequencing. Open Microbiol J, 2012. 6: p. 29-33.

24. Huang X, Wang J, Aluru S, Yang SP, Hillier L. PCAP: a whole-genome assembly program. Genome Res, 2003. 13(9): p. 2164-70.

25. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res, 2009. 37(Database issue): p. D141-5.

26. Giongo A, Richardson AGD, Crabb DB, Triplett EW. TaxCollector: modifying current 16S rRNA databases for the rapid classification at six taxonomic levels. Diversity, 2010. 2(7): p. 1015-1025.

27. Hong SH, Bunge J, Jeon SO, Epstein SS. *Predicting microbial species richness.* Proc Natl Acad Sci U S A, 2006. 103(1): p. 117-22.

28. Mulder CPH, Bazeley-White E, Dimitrakopoulo, PG, Hector A, Scherer-Lorenzen M, Schmid B. Species evenness and productivity in experimental plant communities. Oikos, 2004. 107(1): p. 50-63.

29. Chao A. Non-parametric estimation of the number of classes in a population. Scandinavian Journal of Statistics, 1984. 11: p. 265-270.

30. Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control 1974. 19(6): p. 716–723.

31. Kloke J, McKean J. Rfit: Rank Estimation for Linear Models. http://CRAN.R-project.org/package=RFit, 2013.

32. Furrer R, Nychka D, Sain S. fields: Tools for spatial data. http://CRAN.R-project.org/package=fields, 2013.

33. Bray JR, Curtis JT. An ordination of upland forest communities of southern Wisconsin. Ecological Monographs, 1957. 27: p. 325-349.

34. Legendre P, Legendre L. *Numerical Ecology.* Elsevier, Amsterdam., 2012. Third edition.

35. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MH, Wagner H. vegan: Community Ecology Package. http://CRAN.R-project.org/package=vegan, 2013.

36. Brunner E, Munzel U. The nonparametric Behrens-Fisher problem: asymptotic theory and a small sample approximation. Biometrical J, 2000. 42: p. 17-25.

37. Bartlett MS. Properties of sufficiency and statistical tests. Proceedings of the Royal Statistical Society, 1937. Series A 160: p. 268-282.

38. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, 1995. Series B 57(1): p. 289–300.

39. Fruchterman TMJ, Reingold EM. Graph Drawing by Force-Directed Placement. Software – Practice & Experience, 1991. 21(11): p. 1129–1164.

40. Gould PR. On the geographical interpretation of eigenvectors. Transactions of the Institute of British Geographers, 1967. 42: p. 53-86.

41. Kolmogorov A. Sulla determinazione empirica di una legge di distribuzione. G. Ist. Ital. Attuari, 1933. 4: p. 83.

42. Smirnov N. Tables for estimating the goodness of fit of empirical distributions. Annals of Mathematical Statistics, 1948. 19: p. 279.

43. Joyce KE, Laurienti PJ, Burdette JH, Hayasaka S. A new measure of centrality for brain networks. PLoS One, 2010. 5(8): p. e12200.

44. Matsuzaki T, Nagata Y, Kado S, Uchida K, Kato I, Hashimoto S, Yokokura T. Prevention of onset in an insulin-dependent diabetes mellitus model, NOD mice, by oral feeding of Lactobacillus casei. APMIS, 1997. 105(8): p. 643-9.

45. Yadav H, Jain S, Sinha PR. Antidiabetic effect of probiotic dahi containing Lactobacillus acidophilus and Lactobacillus casei in high fructose fed rats. Nutrition, 2007. 23(1): p. 62-8.

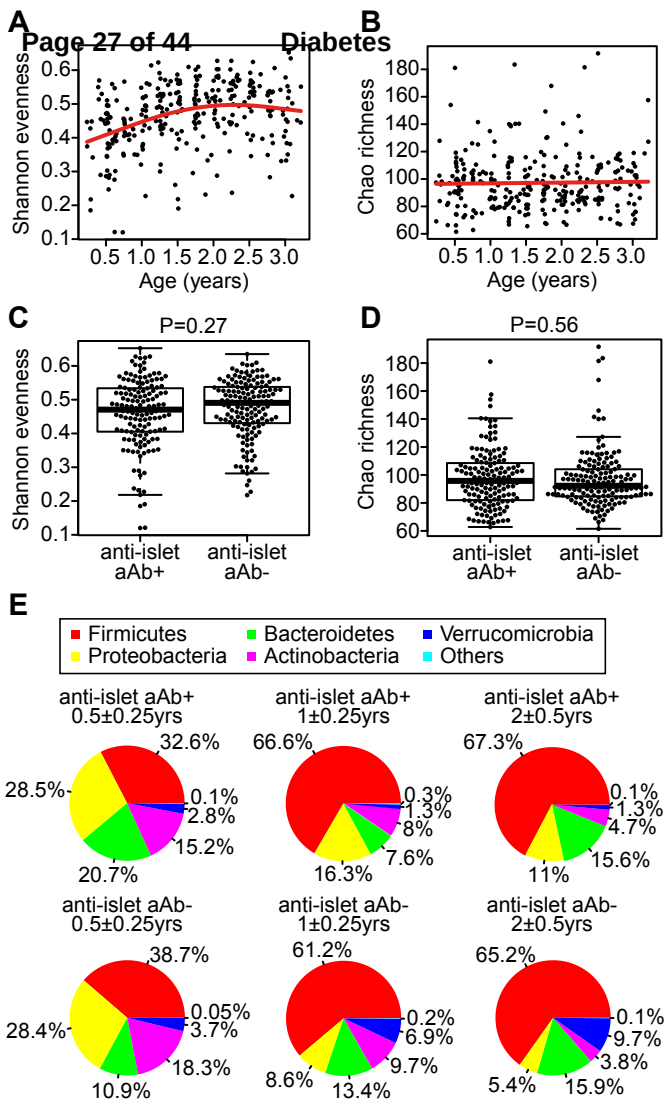
46. Calcinaro F, Dionisi S, Marinaro M, Candeloro P, Bonato V, Marzotti S, Corneli RB, Ferretti E, Gulino A, Grasso F, De Simone C, Di Mario U, Falorni A, Boirivant M, Dotta F. Oral probiotic administration induces interleukin-10 production and prevents spontaneous autoimmune diabetes in the non-obese diabetic mouse. Diabetologia, 2005. 48(8): p. 1565-75.

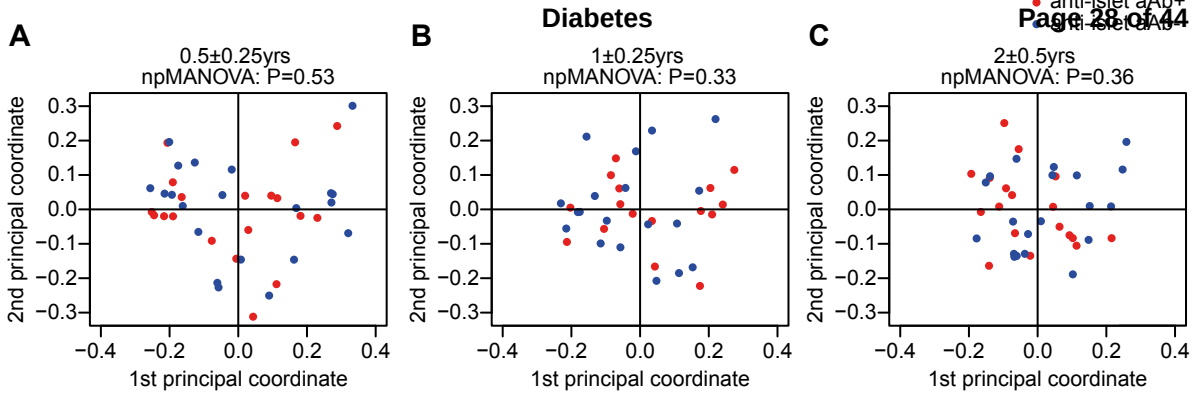
47. Ziegler AG, Rewers M, Simell O, Simell T, Lempainen J, Steck A, Winkler C, Ilonen J, Veijola R, Knip M, Bonifacio E, Eisenbarth GS. Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. JAMA, 2013;309(23):2473-9

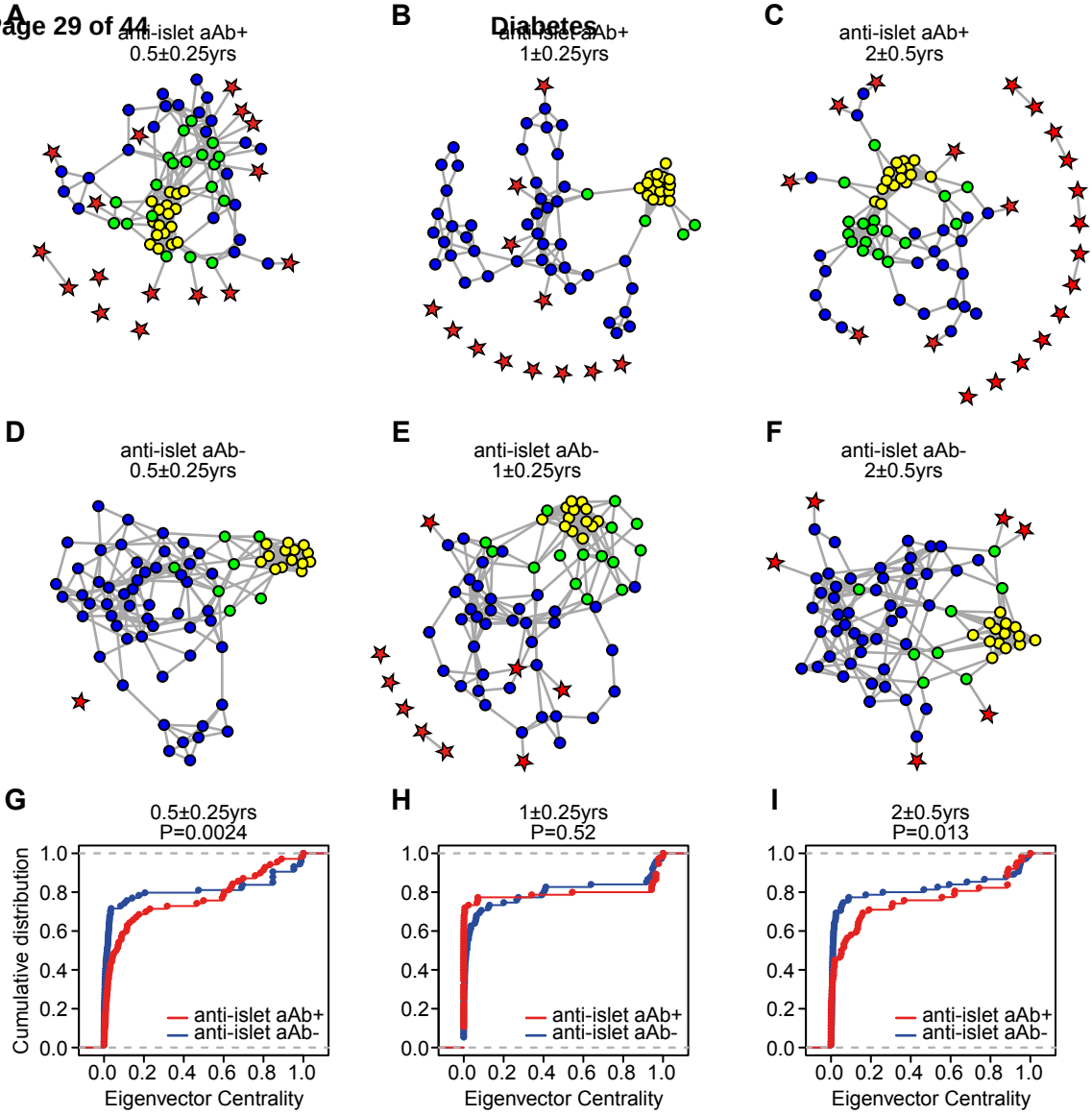
48. Murri M, Leiva I, Gomez-Zumaquero JM, Tinahones FJ, Cardona F, Soriguer F, Queipo-Ortuño MI. Gut microbiota in children with type 1 diabetes differs from that in healthy children: a case-control study. BMC Med, 2013. 11(46).

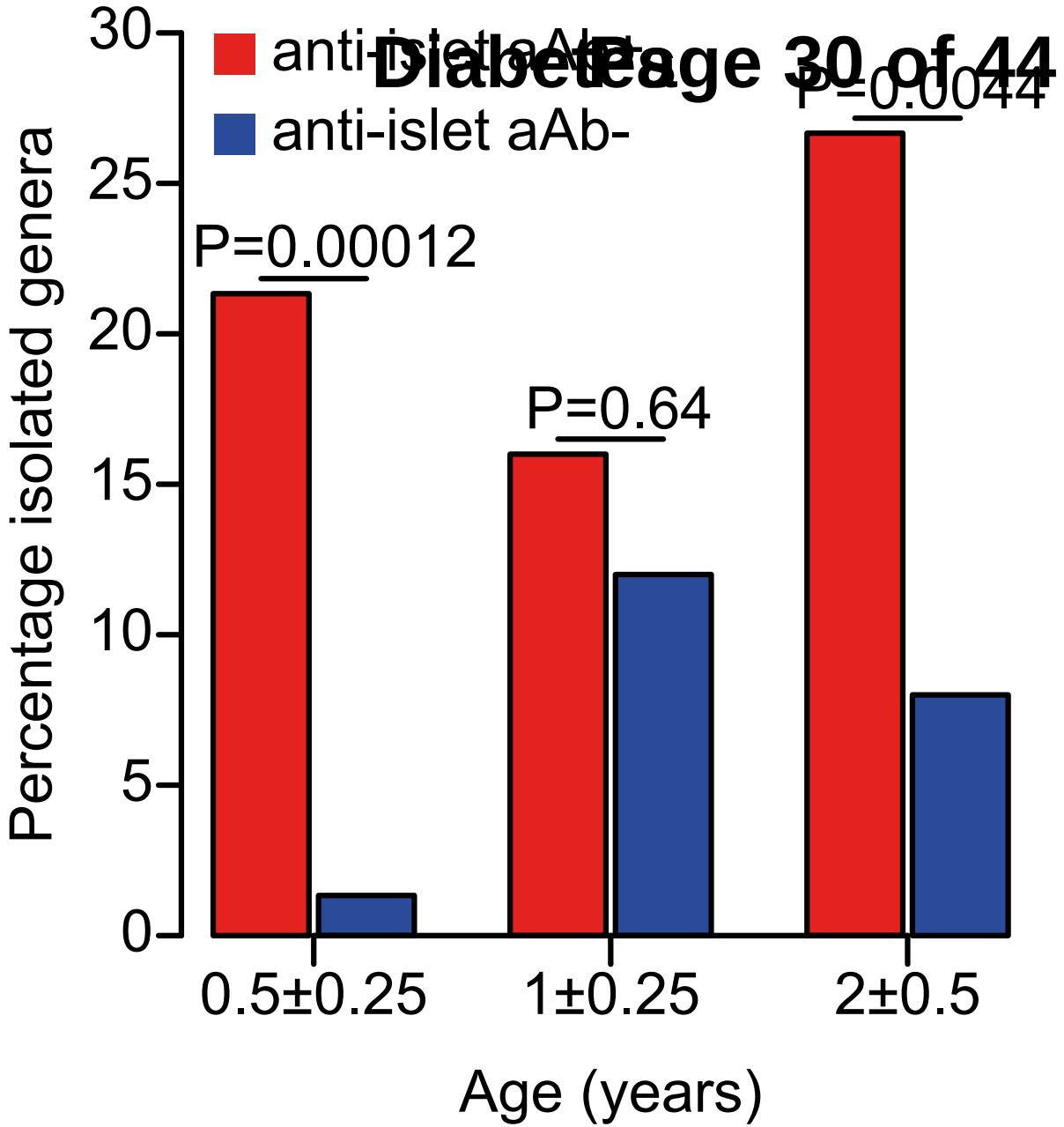
49. Penders J, Thijs C, Vink C, Stelma FF, Snijders B, Kummeling I, van den Brandt PA, Stobberingh EE. Factors influencing the composition of the intestinal microbiota in early infancy. Pediatrics, 2006. 118(2): p. 511-21.

50. Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, Huttenhower C. Microbial co-occurrence relationships in the human microbiome. PLoS Comput Biol, 2012. 8(7): p. e1002606.

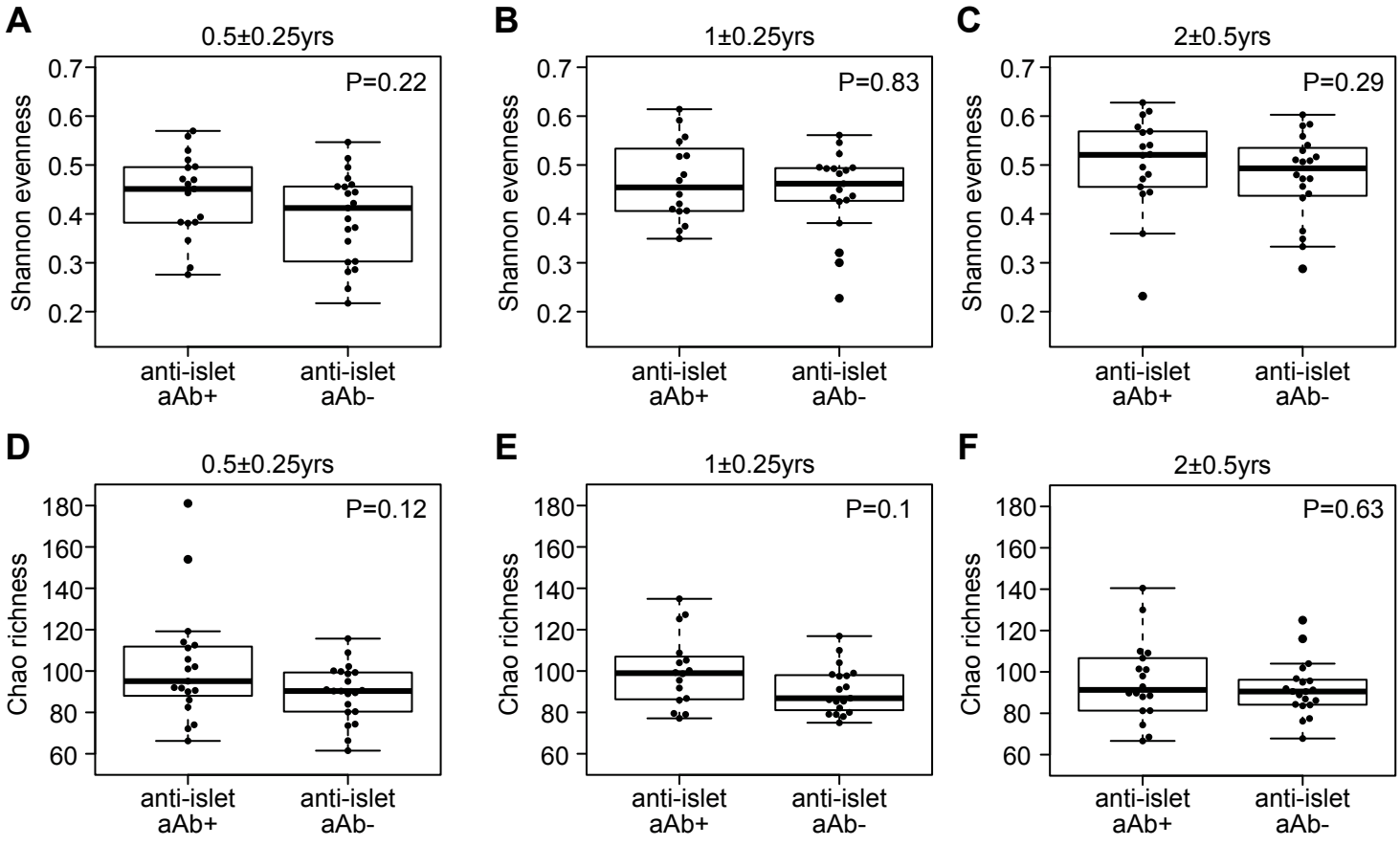

A
Shannon evenness
0.6
0.5
0.4
0.3
0.2
0.1
0.5 1.0 1.5 2.0 2.5 3.0
Age (years)
B
Chao richness
180
160
140
120
100
80
60
0.5 1.0 1.5 2.0 2.5 3.0
Age (years)
C
P=0.27
Shannon evenness
anti-islet aAb+
anti-islet aAb-
D
P=0.56
Chao richness
anti-islet aAb+
anti-islet aAb-
E
Firmicutes
Bacteroidetes
Verrucomicrobia
Proteobacteria
Actinobacteria
Others
anti-islet aAb+ 0.5±0.25yrs
32.6%
28.5%
0.1%
2.8%
15.2%
20.7%
anti-islet aAb+ 1±0.25yrs
66.6%
0.3%
1.3%
8%
7.6%
16.3%
anti-islet aAb+ 2±0.5yrs
67.3%
0.1%
1.3%
4.7%
15.6%
11%
anti-islet aAb- 0.5±0.25yrs
38.7%
28.4%
0.05%
3.7%
18.3%
10.9%
anti-islet aAb- 1±0.25yrs
61.2%
0.2%
6.9%
9.7%
8.6%
13.4%
anti-islet aAb- 2±0.5yrs
65.2%
0.1%
9.7%
3.8%
5.4%
15.9%


Diabetes
anti-islet aAb+
anti-islet aAb−
A
0.5±0.25yrs
npMANOVA: P=0.53
B
1±0.25yrs
npMANOVA: P=0.33
C
2±0.5yrs
npMANOVA: P=0.36
2nd principal coordinate
1st principal coordinate
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.2
0.0
0.2
0.4

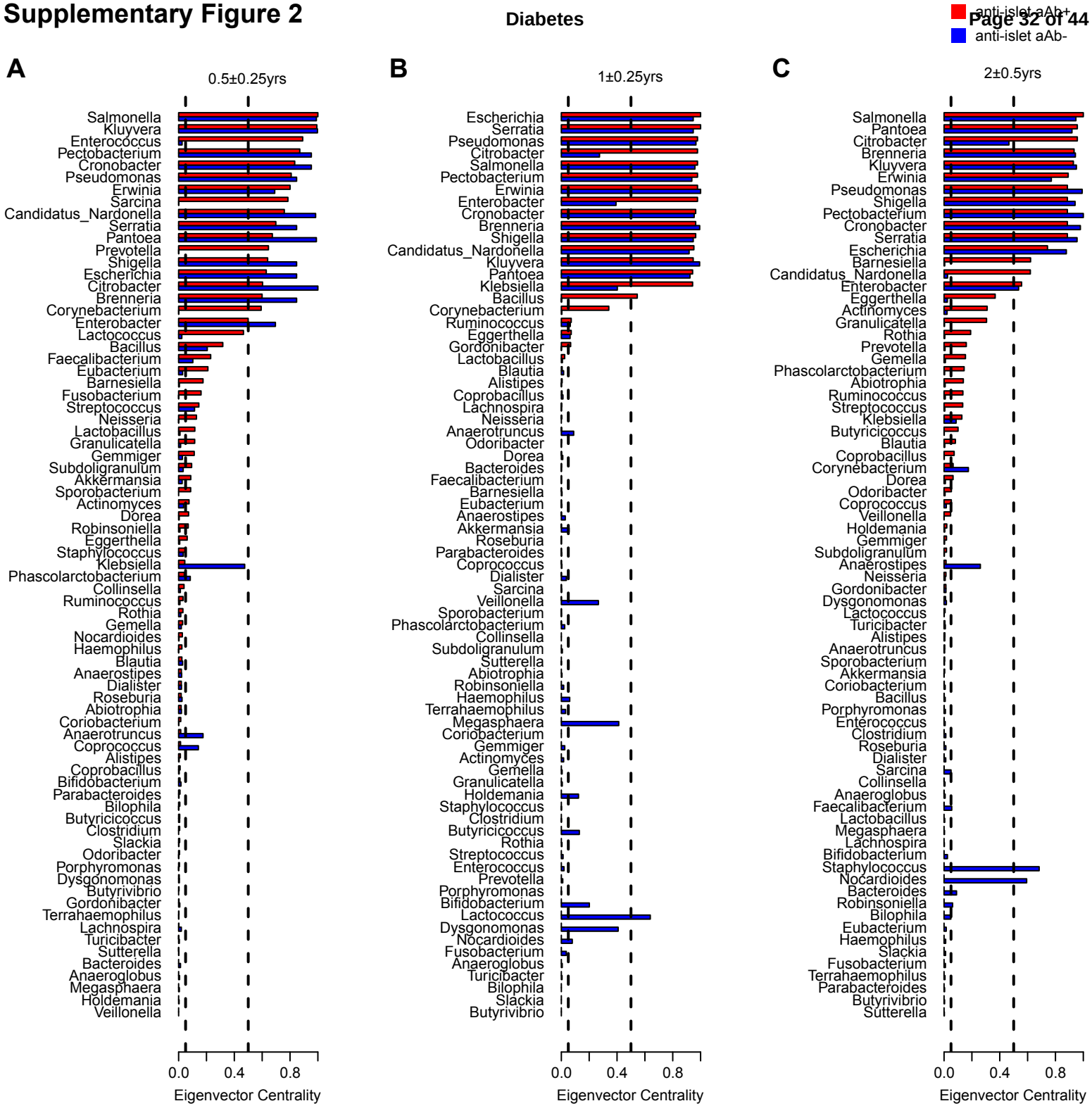

A
anti-islet aAb+
0.5±0.25yrs
B
anti-islet aAb+
1±0.25yrs
C
anti-islet aAb+
2±0.5yrs
D
anti-islet aAb-
0.5±0.25yrs
E
anti-islet aAb-
1±0.25yrs
F
anti-islet aAb-
2±0.5yrs
G
0.5±0.25yrs
P=0.0024
H
1±0.25yrs
P=0.52
I
2±0.5yrs
P=0.013
Cumulative distribution
0.0
0.2
0.4
0.6
0.8
1.0
anti-islet aAb+
anti-islet aAb-
Eigenvector Centrality


anti-islet aAb+
anti-islet aAb-
Percentage isolated genera
30
25
20
15
10
5
0
P=0.00012
P=0.64
P=0.0044
0.5±0.25
1±0.25
2±0.5
Age (years)

Supplementary Figure 1

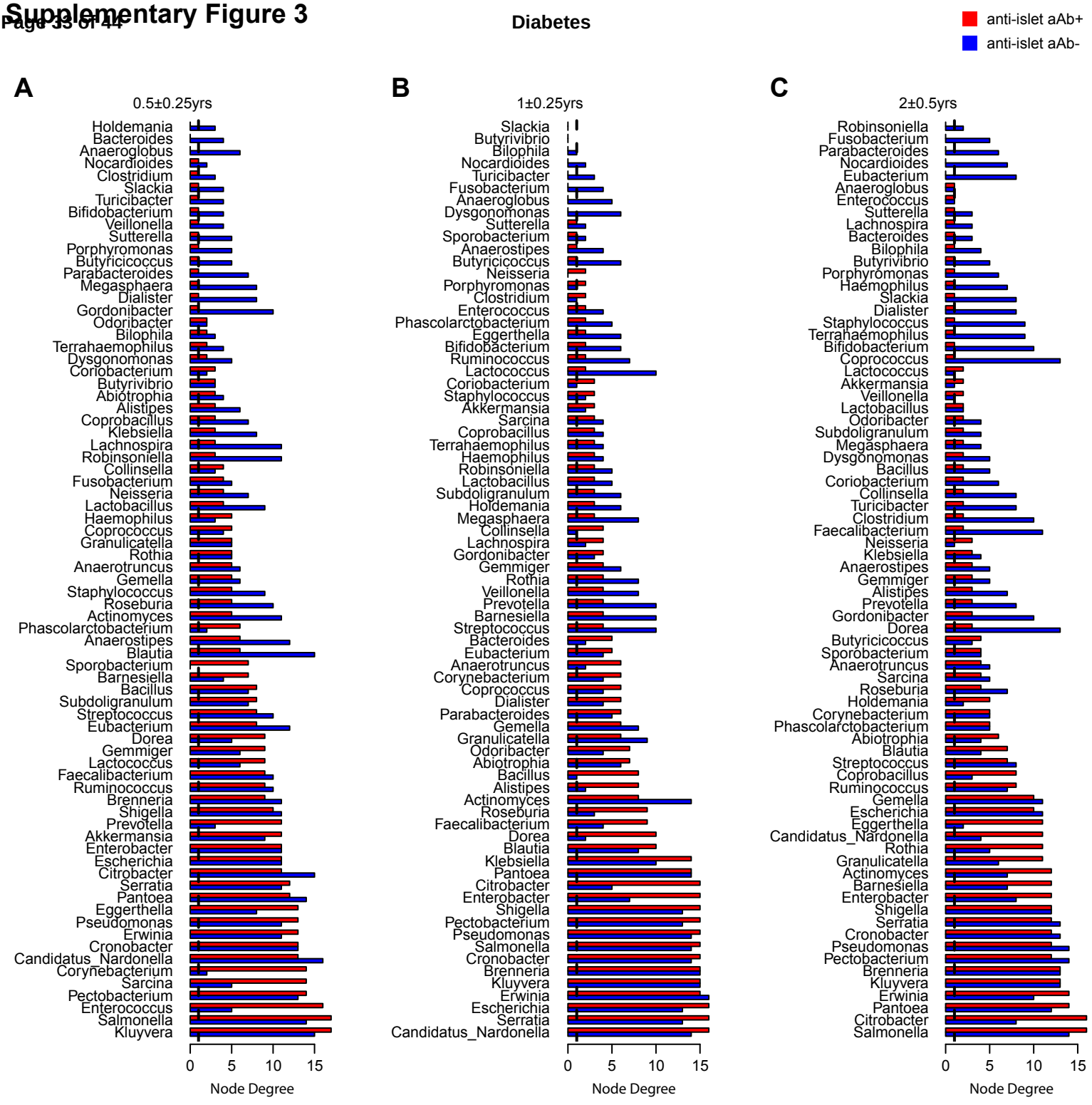


Supplementary Figure 1. (**A-C**) Comparison of Shannon evenness between children who became anti-islet cell autoantibody positive (anti-islet aAb+) and children who remained autoantibody negative (anti-islet aAb-) for three age classes. (**D-F**) Comparison of Chao richness between children who became anti-islet aAb+ and children who remained anti-islet aAb- for three age classes. P-values were obtained by Wilcoxon-Mann-Whitney tests.



Supplementary Figure 2: Eigenvector centrality of genera for bacterial interaction networks for children who became anti-islet cell autoantibody positive (anti-islet aAb+) and children who remained autoantibody negative (anti-islet aAb-) for age 0.5±0.25 years (**A**), age 1±0.25 years (**B**) and age 2±0.5 years (**C**). The rows of each plot are ordered by decreasing eigenvector centrality values of bacterial networks of children who became anti-islet aAb+. Vertical lines indicate eigenvector centrality values of 0.05 and 0.5, respectively.

Supplementary Figure 3



Supplementary Figure 3: Node degree of genera for bacterial interaction networks for children who became anti-islet cell autoantibody positive (anti-islet aAb+) and children who remained autoantibody negative (anti-islet aAb-) for age 0.5±0.25 years (**A**), age 1±0.25 years (**B**) and age 2±0.5 years (**C**). The rows of each plot are ordered by increasing node degree values of children who became anti-islet aAb+. Vertical lines indicate a node degree of 1.

Supplementary Table 1: Descriptive summary of the cohort for anti-islet cell autoantibody positive (anti-islet aAb+) and negative children (anti-islet aAb-).

	Children who became anti-islet aAb+ (N=22)	Children who remained anti-islet aAb- (N=22)	P-Value (*)
Median number stool samples per child (range)	7 (2-11)	7 (3-10)	-
Median age at anti-islet aAb seroconversion in years (range)	1.54 (0.59-4.76)	-	-
Number of children delivered via Caesarean section	10	12	0.76
Median duration of breast feeding in months (range)	7.1 (0-16.4)	7.6 (0-18.7)	0.73
Median age of solid food introduction in months (range)	5.16 (3.1-7.0)	5.76 (1.6-8.2)	0.69
Median age of first gluten exposure in months (range)	11.0 (4.6-12.6)	9.1 (1.6-12.6)	0.9

(*) For binary variables P-values were obtained by Fisher's Exact tests and for continuous variables P-values were obtained by Wilcoxon-Mann-Whitney tests

Supplementary Table 2: Community composition compared by multivariable npMANOVA. The table shows the p-values of multivariable npMANOVA analyses for three different age intervals. Significant results are labelled with asterisks and bold font.

	Age	C-section (yes/no)	Breast feeding at sampling time (yes/no)	Duration since solid food introduction	Duration since gluten introduction	Anti-islet cell autoantibody status (positive/negative)
0.5±0.25 years	0.099	0.19	**0.002****	**0.001****	0.084	0.20
1±0.25 years	0.23	0.52	0.39	**0.049***	0.13	0.38
2±0.5 years	0.18	0.24	---	0.56	0.91	0.18

Supplementary Table 3: Comparsion of the abundances of genera between children who became anti-islet cell autoantibody positive and children who remained anti-islet cell autoantibody negative for age 0.5 years. Percentages represent the fraction of the total number of reads of each genus in the particular anti-islet cell autoantibody group. Associated covariates were detected via stepwise AIC multiple regression models. Wilcoxon-Mann-Whitney or Brunner-Munzel test P-values were obtained by testing for differences in the abundances of genera between children who became anti-islet aAb+ and children who remained anti-islet aAb-.

Bacterial genera	Children who developed anti-islet cell autoantibodies (%)	Children who remained anti-islet cell autoantibody negative (%)	P-Value	Covariate adjusted P-Value	Associated covariates (*)
Dorea	0.62	0.0068	0.003	0.7	C-section; solid food
Barnesiella	0.0016	0.014	0.035	0.035	none
Haemophilus	0.11	0.21	0.058	0.058	none
Bacteroides	25	14	0.061	0.088	C-section; breast feeding; age
Sporobacterium	0.0087	0.00038	0.074	0.074	none
Veillonella	3	10	0.076	0.0098	age
Anaeroglobus	0	0.0027	0.083	0.013	C-section; solid food
Citrobacter	0.093	0.11	0.12	0.093	breast feeding
Enterococcus	3.3	0.83	0.13	0.00011	breast feeding; age; gluten
Blautia	1.4	0.78	0.15	0.57	breast feeding; solid food; age
Terrahaemophilus	0.017	0.032	0.17	0.17	none
Butyricicoccus	0.058	4.00E-04	0.17	0.17	none
Slackia	0.063	0.031	0.19	0.1	solid food
Clostridium	8.6	7.3	0.25	0.12	C-section
Granulicatella	0.11	0.15	0.26	0.21	C-section; age
Anaerostipes	0.079	0.11	0.26	0.66	breast feeding; gluten
Erwinia	0.22	0.12	0.27	0.27	none
Lactococcus	0.013	0.34	0.27	0.27	none
Gemella	0.14	0.42	0.29	0.35	age

Enterobacter	0.46	0.49	0.31	0.31	none
Turicibacter	0.17	0.034	0.31	0.52	solid food
Neisseria	0.0039	0.0083	0.31	0.49	C-section
Sutterella	0.095	0.11	0.32	0.32	none
Gemmiger	0.087	0.078	0.33	0.28	C-section; breast feeding
Roseburia	0.24	0.7	0.36	0.58	solid food; gluten
Sarcina	0.0029	0.002	0.39	0.39	none
Bilophila	0.0027	0.0012	0.39	0.78	C-section; breast feeding; gluten
Eubacterium	0.25	0.11	0.42	0.42	none
Dysgonomonas	0.00072	0.0027	0.44	0.86	C-section; age
Pantoea	0.91	0.21	0.49	0.43	breast feeding
Corynebacterium	0.012	0.0041	0.49	0.092	solid food
Salmonella	0.097	0.19	0.52	0.5	breast feeding
Candidatus_Nardonella	0.019	0.042	0.53	0.8	breast feeding; solid food
Klebsiella	0.14	0.5	0.56	0.56	none
Dialister	0.048	0.051	0.57	0.57	none
Ruminococcus	5.8	5.5	0.57	0.87	breast feeding; solid food; age
Lactobacillus	1.4	2.1	0.61	0.61	none
Prevotella	0.016	0.045	0.62	0.62	none
Robinsoniella	0.065	1.1	0.62	0.62	none
Bacillus	0.027	0.0079	0.62	0.33	C-section; age
Subdoligranulum	0.019	0.02	0.63	0.63	none
Streptococcus	6	7.8	0.67	0.67	none
Rothia	0.29	0.38	0.67	0.81	solid food
Alistipes	0.12	0.035	0.67	0.71	breast feeding; solid food
Faecalibacterium	0.23	0.15	0.69	0.94	solid food
Coprococcus	0.0036	0.0039	0.7	0.7	none
Cronobacter	1.1	0.51	0.76	0.36	breast feeding; solid food
Anaerotruncus	0.014	0.0056	0.76	0.97	C-section

Nocardioides	0.14	0.0025	0.76	0.76	none
Kluyvera	0.052	0.083	0.78	0.43	breast feeding; solid food
Abiotrophia	0.0066	0.0078	0.78	0.55	age
Pseudomonas	0.11	0.13	0.79	0.7	breast feeding; solid food
Coriobacterium	0.092	0.015	0.82	0.69	solid food
Akkermansia	4.3	4.4	0.83	0.92	C-section; gluten
Coprobacillus	0.25	0.034	0.83	0.83	none
Actinomyces	0.18	0.24	0.85	0.98	age
Butyrivibrio	0.0026	0.00035	0.85	0.85	none
Lachnospira	0.29	0.43	0.86	0.66	solid food; gluten
Pectobacterium	0.081	0.11	0.87	0.76	breast feeding; solid food
Holdemania	0.024	0.0056	0.87	0.71	age
Eggerthella	0.97	1.3	0.87	0.98	breast feeding
Shigella	1.3	1.6	0.87	0.78	breast feeding; solid food
Odoribacter	0.001	0.0025	0.88	0.85	breast feeding
Parabacteroides	0.004	0.021	0.9	0.84	breast feeding
Collinsella	0.28	0.038	0.91	0.43	solid food
Porphyromonas	0.0024	0.0047	0.92	0.92	none
Gordonibacter	0.013	0.0036	0.93	0.34	solid food
Brenneria	0.045	0.048	0.94	0.68	breast feeding; solid food
Phascolarctobacterium	0.0085	0.0044	0.95	0.22	C-section; breast feeding; gluten
Serratia	2.7	3.4	0.98	0.68	breast feeding; solid food
Escherichia	9.5	12	0.98	0.55	breast feeding; solid food
Fusobacterium	0.026	0.011	0.99	0.99	none
Staphylococcus	0.14	0.027	0.99	0.99	none
Bifidobacterium	18	21	1	1	none
Megasphaera	0.17	0.26	1	1	none

(*) gluten: duration since first gluten introduction
solid food: duration since first solid food introducton
breast: breast feeding at sampling time (yes, no)
age: age of the children

Supplementary Table 4: Comparsion of the abundances of genera between children who became anti-islet cell autoantibody positive and children who remained anti-islet cell autoantibody negative for age 1 years. Percentages represent the fraction of the total number of reads of each genus in the particular anti-islet cell autoantibody group. Associated covariates were detected via stepwise AIC multiple regression models. Wilcoxon-Mann-Whitney or Brunner-Munzel test P-values were obtained by testing for differences in the abundances of genera between children who became anti-islet aAb+ and children who remained anti-islet aAb-.

Bacterial genera	Children who developed anti-islet cell autoantibodies (%)	Children who remained anti-islet cell autoantibody negative (%)	P-Value	Covariate adjusted P-Value	Associated covariates (*)
Candidatus_Nardonella	0.013	0.0079	0.031	0.31	solid food
Holdemania	0.014	0.064	0.061	0.46	solid food; age; gluten
Salmonella	0.068	0.037	0.09	0.09	none
Faecalibacterium	2.4	9	0.093	0.65	age
Terrahaemophilus	0.022	0.0049	0.095	0.56	solid food
Pantoea	0.061	0.031	0.12	0.12	none
Rothia	0.095	0.056	0.15	0.46	solid food
Bacillus	0.0043	0.0025	0.17	0.17	none
Brenneria	0.021	0.012	0.18	0.25	gluten
Megasphaera	0.89	0.34	0.2	0.2	none
Clostridium	16	14	0.21	0.27	gluten
Shigella	0.73	0.37	0.21	0.2	gluten
Klebsiella	0.15	0.081	0.22	0.47	breast feeding
Escherichia	4.3	3	0.23	0.27	gluten
Dysgonomonas	0.031	0.00047	0.23	0.23	none
Gemmiger	6.6	1.1	0.24	0.24	none
Collinsella	0.18	0.21	0.25	0.25	none
Robinsoniella	1	0.02	0.25	0.55	age
Anaerotruncus	0.041	0.0058	0.26	0.0011	solid food
Haemophilus	0.14	0.058	0.27	0.79	C-section; solid food
Lactobacillus	0.17	0.63	0.27	0.53	solid food; age

Turicibacter	0.18	0.11	0.29	0.29	none
Bacteroides	9.8	17	0.29	0.29	none
Parabacteroides	0.012	0.011	0.3	0.54	C-section
Cronobacter	0.23	0.13	0.3	0.23	gluten
Serratia	1.6	0.84	0.32	0.29	gluten
Blautia	4.7	3.6	0.33	0.79	C-section; breast feeding; gluten
Slackia	0.00014	0.11	0.33	0.85	C-section; solid food; gluten
Enterobacter	0.2	0.096	0.35	0.35	none
Odoribacter	0.045	0.005	0.36	0.0069	solid food
Enterococcus	0.19	0.3	0.37	0.37	none
Pseudomonas	0.063	0.034	0.37	0.35	gluten
Bifidobacterium	7.9	11	0.38	0.38	none
Granulicatella	0.084	0.14	0.4	0.4	none
Pectobacterium	0.044	0.028	0.41	0.47	gluten
Bilophila	0.0054	0.0035	0.41	0.68	C-section; gluten
Citrobacter	0.046	0.037	0.45	0.96	breast feeding
Gemella	0.06	0.066	0.46	0.46	none
Lactococcus	0.05	0.019	0.46	0.51	C-section; solid food; age
Lachnospira	0.24	1.5	0.49	0.33	gluten
Streptococcus	3.7	6.9	0.52	0.55	C-section
Dialister	2.7	0.18	0.53	0.53	none
Kluyvera	0.036	0.02	0.54	0.47	gluten
Ruminococcus	13	11	0.54	0.13	breast feeding; gluten
Sutterella	0.1	0.0085	0.54	0.22	gluten
Anaeroglobus	0.11	0.023	0.55	0.083	age; gluten
Dorea	0.12	0.36	0.56	0.86	C-section; solid food
Fusobacterium	0.1	0.038	0.56	0.56	none
Neisseria	0.065	0.011	0.57	0.57	none
Coriobacterium	0.055	0.07	0.58	0.58	none
Anaerostipes	0.24	0.18	0.58	1	gluten

Porphyromonas	0.014	0.0043	0.61	0.61	none
Erwinia	0.046	0.05	0.63	0.63	none
Staphylococcus	0.012	0.0068	0.64	0.15	breast feeding; gluten
Eggerthella	2.5	1.3	0.66	0.66	none
Eubacterium	1.9	2.1	0.66	0.66	none
Akkermansia	1.6	9.3	0.68	0.68	none
Roseburia	0.75	0.66	0.68	0.68	none
Nocardioides	0.003	0.00027	0.71	0.71	none
Coprobacillus	0.081	0.071	0.76	0.76	none
Abiotrophia	0.012	0.027	0.77	0.46	age
Corynebacterium	0.0016	0.0044	0.78	0.79	C-section; age
Barnesiella	0.091	0.024	0.84	0.19	solid food
Gordonibacter	0.016	0.008	0.84	0.15	solid food
Prevotella	0.11	0.02	0.85	0.85	none
Alistipes	1.2	0.34	0.85	0.015	age
Sporobacterium	0.0049	0.0048	0.87	0.87	none
Subdoligranulum	0.61	0.12	0.88	0.88	none
Butyricicoccus	0.11	0.017	0.89	0.89	none
Butyrivibrio	0.0012	0.0019	0.89	0.2	solid food
Veillonella	11	3	0.9	0.29	C-section; breast feeding; gluten
Actinomyces	0.11	0.095	0.93	0.93	none
Sarcina	0.05	0.016	0.93	0.77	C-section
Coprococcus	0.032	0.026	0.96	0.97	gluten
Phascolarctobacterium	0.049	0.02	0.97	0.96	C-section; breast feeding; gluten

(*) gluten: duration since first gluten introduction
solid food: duration since first solid food introducton
breast: breast feeding at sampling time (yes. no)
age: age of the children

Supplementary Table 5: Comparsion of the abundances of genera between children who became anti-islet cell autoantibody positive and children who remained anti-islet cell autoantibody negative for age 2 years. Percentages represent the fraction of the total number of reads of each genus in the particular anti-islet cell autoantibody group. Associated covariates were detected via stepwise AIC multiple regression models. Wilcoxon-Mann-Whitney or Brunner-Munzel test P-values were obtained by testing for differences in the abundances of genera between children who became anti-islet aAb+ and children who remained anti-islet aAb-.

Bacterial genera	Children who developed anti-islet cell autoantibodies (%)	Children who remained anti-islet cell autoantibody negative (%)	P-Value	Covariate adjusted P-Value	Associated covariates (*)
Erwinia	0.028	0.0059	0.024	0.024	none
Enterobacter	0.14	0.015	0.045	0.023	gluten
Abiotrophia	0.02	0.026	0.053	0.093	age
Blautia	6.3	2.9	0.061	0.061	none
Staphylococcus	0.0043	0.0031	0.077	0.077	none
Bacillus	0.003	0.01	0.1	0.1	none
Bifidobacterium	5.7	4.4	0.11	0.059	C-section
Corynebacterium	0.018	0.014	0.12	0.12	none
Ruminococcus	17	12	0.16	0.16	none
Granulicatella	0.093	0.08	0.17	0.17	none
Haemophilus	0.044	0.046	0.18	0.35	C-section; solid food
Turicibacter	0.3	0.28	0.19	0.41	solid food
Lactobacillus	0.12	0.23	0.19	0.47	solid food
Lachnospira	0.15	0.095	0.19	0.19	none
Prevotella	0.29	0.03	0.2	0.2	none
Streptococcus	2.2	2.2	0.22	0.22	none
Enterococcus	0.64	0.031	0.24	0.15	gluten
Klebsiella	0.06	0.0098	0.24	0.24	none
Pantoea	0.03	0.013	0.25	0.055	gluten
Anaerotruncus	0.025	0.015	0.29	0.29	none
Kluyvera	0.02	0.0082	0.31	0.31	none
Holdemania	0.23	0.12	0.31	0.32	C-section

Nocardioides	2.00E-04	0.0061	0.33	0.33	none
Akkermansia	2	12	0.35	0.35	none
Megasphaera	0.47	0.074	0.36	0.36	none
Alistipes	0.74	0.71	0.38	0.38	none
Subdoligranulum	0.53	3.8	0.42	0.053	C-section
Dorea	0.53	0.83	0.43	0.51	solid food; age
Eubacterium	4.5	2.9	0.43	0.45	solid food; age
Neisseria	0.0042	0.0012	0.47	0.47	none
Coprococcus	0.35	0.14	0.48	0.48	none
Dysgonomonas	0.21	0.00069	0.48	0.48	none
Porphyromonas	0.0062	0.027	0.49	0.46	C-section
Slackia	0.0071	0.024	0.5	0.95	solid food; age
Candidatus_Nardonella	0.0072	0.003	0.52	0.29	C-section; age; gluten
Pseudomonas	0.029	0.016	0.53	0.39	C-section
Veillonella	0.42	2.8	0.55	0.37	solid food
Faecalibacterium	7.3	9.6	0.55	0.57	age
Brenneria	0.013	0.0052	0.56	0.27	C-section
Sarcina	0.15	0.057	0.56	0.41	gluten
Bacteroides	20	21	0.57	0.57	none
Terrahaemophilus	0.0056	0.0055	0.57	0.71	solid food
Anaeroglobus	0.024	0.0022	0.57	0.84	solid food; gluten
Robinsoniella	0.0099	0.0079	0.59	0.75	C-section
Gordonibacter	0.032	0.024	0.59	0.97	gluten
Serratia	1	0.39	0.61	0.25	C-section
Lactococcus	0.078	0.072	0.61	0.61	none
Barnesiella	0.23	0.095	0.64	0.64	none
Phascolarctobacterium	0.37	0.008	0.64	0.64	none
Gemmiger	5	3.2	0.65	0.65	none
Coprobacillus	0.092	0.038	0.67	0.92	solid food; age; gluten
Butyricicoccus	0.0085	0.0056	0.72	0.71	solid food
Pectobacterium	0.037	0.012	0.73	0.45	C-section

Anaerostipes	0.045	0.056	0.76	0.99	age; gluten
Parabacteroides	0.23	0.22	0.77	0.77	none
Dialister	2.7	3	0.78	0.65	C-section; age; gluten
Rothia	0.095	0.06	0.8	0.87	age
Bilophila	0.045	0.016	0.8	0.92	solid food; gluten
Roseburia	2.9	2.4	0.81	0.65	age
Sutterella	0.029	0.024	0.82	0.92	solid food; age
Citrobacter	0.023	0.0056	0.85	0.21	C-section; gluten
Coriobacterium	0.13	0.14	0.85	1	age
Gemella	0.063	0.043	0.85	0.85	none
Escherichia	3.5	1.5	0.89	0.33	C-section
Cronobacter	0.14	0.067	0.91	0.25	C-section
Salmonella	0.043	0.012	0.91	0.24	C-section; gluten
Collinsella	0.36	0.38	0.92	0.97	age
Actinomyces	0.073	0.067	0.92	0.92	none
Odoribacter	0.042	0.034	0.94	0.94	none
Butyrivibrio	0.0022	0.076	0.95	0.95	none
Eggerthella	1	0.56	0.96	0.98	age
Shigella	0.45	0.2	0.96	0.41	C-section
Fusobacterium	0.0067	0.013	0.96	0.96	none
Clostridium	10	10	0.99	0.99	none
Sporobacterium	0.0034	0.0066	0.99	1	solid food

(*) gluten: duration since first gluten introduction
solid food: duration since first solid food introducton
breast: breast feeding at sampling time (yes, no)
age: age of the children