

ORIGINAL ARTICLE

Toward defining the autoimmune microbiome for type 1 diabetes

Adriana Giongo¹, Kelsey A Gano¹, David B Crabb¹, Nabanita Mukherjee², Luis L Novelo², George Casella², Jennifer C Drew¹, Jorma Ilonen^{3,4,5}, Mikael Knip^{6,7}, Heikki Hyöty⁸, Riitta Veijola⁹, Tuula Simell¹⁰, Olli Simell¹⁰, Josef Neu¹¹, Clive H Wasserfall¹², Desmond Schatz¹¹, Mark A Atkinson¹² and Eric W Triplett¹

¹Department of Microbiology and Cell Science, University of Florida, Gainesville, FL, USA; ²Department of Statistics, University of Florida, Gainesville, FL, USA; ³Department of Clinical Microbiology, University of Kuopio, Kuopio, Finland; ⁴Immunogenetics Laboratory, University of Turku, Turku, Finland; ⁵The Juvenile Diabetes Research Foundation (JDRF), Center of Prevention of Type 1 Diabetes, Turku, Finland; ⁶Hospital for Children and Adolescents, University of Helsinki, Helsinki, Finland; ⁷Department of Pediatrics and Research Unit, Tampere University Hospital, Tampere, Finland; ⁸Department of Virology, Medical School and Center for Laboratory Medicine, Tampere University Hospital, University of Tampere, Tampere, Finland; ⁹Department of Pediatrics, University of Oulu, Oulu, Finland; ¹⁰Department of Pediatrics, Turku University Hospital, Turku, Finland; ¹¹Department of Pediatrics, University of Florida, Gainesville, FL, USA and ¹²Department of Pathology, Immunology, and Laboratory Medicine, University of Florida, Gainesville, FL, USA

Several studies have shown that gut bacteria have a role in diabetes in murine models. Specific bacteria have been correlated with the onset of diabetes in a rat model. However, it is unknown whether human intestinal microbes have a role in the development of autoimmunity that often leads to type 1 diabetes (T1D), an autoimmune disorder in which insulin-secreting pancreatic islet cells are destroyed. High-throughput, culture-independent approaches identified bacteria that correlate with the development of T1D-associated autoimmunity in young children who are at high genetic risk for this disorder. The level of bacterial diversity diminishes overtime in these autoimmune subjects relative to that of age-matched, genotype-matched, nonautoimmune individuals. A single species, *Bacteroides ovatus*, comprised nearly 24% of the total increase in the phylum Bacteroidetes in cases compared with controls. Conversely, another species in controls, represented by the human firmicute strain CO19, represented nearly 20% of the increase in Firmicutes compared with cases overtime. Three lines of evidence are presented that support the notion that, as healthy infants approach the toddler stage, their microbiomes become healthier and more stable, whereas, children who are destined for autoimmunity develop a microbiome that is less diverse and stable. Hence, the autoimmune microbiome for T1D may be distinctly different from that found in healthy children. These data also suggest bacterial markers for the early diagnosis of T1D. In addition, bacteria that negatively correlated with the autoimmune state may prove to be useful in the prevention of autoimmunity development in high-risk children.

The ISME Journal (2010) 0, 000–000; doi:10.1038/ismej.2010.92

Subject Category: microbe-microbe and microbe-host interactions

Keywords: Bacteroidetes; Firmicutes; gut microbiota; seroconversion

Introduction

A recent review by Vaarala *et al.* (2008) describes a trio of factors that create a perfect storm of events leading to autoimmunity in type 1 diabetes (T1D). These factors include an aberrant intestinal microbiota, a leaky intestinal mucosal barrier and an altered intestinal immune responsiveness. The

interplay of these factors seems to have a crucial role in the onset of several allergenic and autoimmune diseases, including Crohn's disease, celiac disease, T1D and multiple sclerosis (Frank *et al.*, 2007; Wen *et al.*, 2008; Willing *et al.*, 2009).

Insulin-dependent T1D is relatively common, with no female dominance. Chronic and autoimmune diseases are usually diagnosed in young children and are caused by T-cell-mediated destruction of insulin-producing pancreatic β cells in the islets of the pancreas (Harrison *et al.*, 2008). A role for bacteria in the onset of diabetes has been shown in two murine models. For example, feeding probiotic bacterial strains, usually lactic acid bacteria,

Correspondence: EW Triplett, Department of Microbiology and Cell Science, University of Florida, 1052 Museum Road, Gainesville, FL 32611-0700, USA.

E-mail: ewt@ufl.edu

Received 24 February 2010; revised 25 May 2010; accepted 28 May 2010

to nonobese diabetic mice or biobreeding diabetes-prone (BB-DP) rats can delay or prevent diabetes (Matsuzaki *et al.*, 1997; Calcinaro *et al.*, 2005; Yadav *et al.*, 2007). Feeding antibiotics to nonobese diabetic mice or BB-DP rats can also increase survival in these models (Brugman *et al.*, 2006; Schwartz *et al.*, 2007). In addition, pathogen-free nonobese diabetic mice lacking an adaptor protein for multiple toll-like receptors known to bind to bacterial ligands fail to develop diabetes (Wen *et al.*, 2008).

Roesch *et al.* (2009b) conducted a culture-independent analysis of gut bacteria in BB-DP and biobreeding diabetes-resistant (BB-DR) rats and showed that, at the time of diabetes onset, the bacterial communities in these two rat strains differed significantly. Stool from BB-DR rats contained much higher populations of probiotic-like bacteria, such as *Lactobacillus* and *Bifidobacterium*, whereas BB-DP rats had higher numbers of *Bacteroides*, *Eubacterium* and *Ruminococcus*. A total of 24 bacterial species were found to differ significantly in abundance between the BB-DP and BB-DR rat samples. Five species of *Clostridium* were higher in BB-DP rats, whereas *Clostridium hylemonae* was higher in BB-DR rats. In addition, hundreds of bacterial taxa that could not be classified to genus level were also found to differ. Many Lachnospiraceae were in higher abundance in BB-DP rats, whereas many unclassified Clostridiaceae were more common in BB-DR rats. The differences in *Lactobacillus* and *Bifidobacterium* observed by pyrosequencing were confirmed by quantitative PCR.

All these results from Roesch *et al.* (2009b) are consistent with the notion that beneficial bacteria seem to provide a protective effect in rodent models by delaying or preventing the onset of diabetes. As BB-DP rats have lower populations of species that contain known probiotic strains than do BB-DR rats, potentially beneficial bacteria may be necessary for the maintenance of a healthy microbiome essential in preventing a leaky gut. A *Lactobacillus johnsonii* strain has since been isolated from the stool of the same set of BB-DR rats as was used in Roesch *et al.* (2009b). This strain of *L. johnsonii* prevents diabetes when fed to BB-DP rats (Valladares *et al.* 2010).

These results encouraged a close examination of gut bacteria in humans who are at high risk for autoimmunity and T1D. Human stool samples for such an analysis have been collected by the Diabetes Prediction and Prevention study in Finland (Nejentsev *et al.*, 1999; Kupila *et al.*, 2001). The Diabetes Prediction and Prevention study has been collecting stool samples from children on the basis of their genotype since 1994. At birth, the HLA-DQ genotypes of babies are determined. Those infants who possess specific HLA genotypes are considered to be at high risk for autoimmunity and progression toward T1D early in life. When children enter the study, stool samples are collected every 3 months and blood samples are collected and assayed for the presence of specific autoantibodies. Once two autoantibodies are detected,

the child is diagnosed as having seroconverted to autoimmunity for T1D.

Materials and methods

The samples used in this study came from a total of eight Finnish children, each represented by three stool samples collected at three time points, in a total of 24 separate samples. The case children all developed autoimmunity and eventually T1D overtime (Table 1). Autoimmunity was diagnosed by the appearance of at least two autoantibodies. Each of these cases is matched with three samples from a child of the same age and HLA-DQ genotype who did not become autoimmune during the study.

Stool samples were collected by parents at home and delivered to the repository for frozen storage within 48 h. Recent work has shown that storage at room temperature for up to 72 h has a minimal effect on stool bacterial community structure (Roesch *et al.*, 2009a). The case children became autoimmune at or near the time of the third sampling (Table 1). Autoimmunity in T1D is defined as the appearance of two autoantibodies in the serum as described by The Environmental Determinants of Diabetes in The TEDDY Study Group, (2007). Each case subject was matched with a healthy (that is, non-T1D-associated autoantibody positive, nondiabetic) child of the same genotype and of approximately the same age.

DNA extraction, 16S rRNA amplification and pyrosequencing were performed as described previously (Roesch *et al.*, 2009a). An average of 15 709 sequences were obtained for each of the 24 samples (Table 2). The barcodes used in this study to differentiate the samples are described in Supplementary Table 1. The original sequences and the corresponding quality scores are submitted to GenBank as study accession number SRP002359.1.

The sequences were processed and analyzed to determine differences at all taxonomic and

Table 1 Age (days after birth) of each subject at the three time points of collection

Subject	Days after birth				
	Sample 1	Sample 2	Sample 3	Autoimmunity	T1D
Case 1	125	351	512	513	513
Control 1	244	370	599		
Case 2	131	447	659	902	1430
Control 2	139	445	655		
Case 3	123	552	832	298	731
Control 3	149	611	1093		
Case 4	147	242	696	470	1921
Control 4	147	238	693		
Average	150.6	407	717.4		

For each case subject, the ages of diagnosis for autoimmunity and type 1 diabetes (T1D) are also listed.

Table 2 Characteristics of samples collected from the DIPP study

Reads	Number
Total number of reads	406 979
Number of useable reads	390 759
Average of read length	264
Average number of reads/sample	15 709
Minimum number of reads/sample	9 025
Maximum number of reads/sample	42 942

Abbreviation: DIPP, Diabetes Prediction and Prevention study.

Q4 community levels using PANGEA (Giongo *et al.*, 2010a). In PANGEA, small sequences are discarded, poor quality ends are trimmed, 16S rRNA sequences are separated by barcode, the closest cultured relative of each sequence is identified using Megablast, data are collated into stables, statistically significant differences in the abundance of taxa are determined and the Shannon diversity index (Shannon and Weaver, 1949) of each community is calculated. **Q5** Classification was performed using an up-to-date RDP-II database modified using TaxCollector (Giongo *et al.*, 2010b). The Shannon diversity index was chosen as it considers both the presence and abundance of OTU. **Q6** The significant differences between taxa are determined in a manner that includes a false discovery rate determination and is carried out using a modified χ^2 -test.

To test whether the microbial communities from controls are more similar to each other than those from cases, UniFrac (Lozupone *et al.*, 2006) pairwise distances in between all pairs of cases and all pairs of controls were determined. Specifically, for each one of the six possible pairs of individuals in the same group, a corresponding phylogenetic tree was generated using MUSCLE (Edgar, 2004). **Q7** The sum of the UniFrac distances of the trees in the case group was computed, as well as the sum of the six UniFrac differences in the control group. **Q8** The difference between these sums (D_observed) was used to calculate the difference between cases and controls. A permutation test was calculated to verify whether the difference was statistically significant.

To assess the significance of the UniFrac distances described above, we ran a permutation test as follows (R code is available from the authors). For each phylogenetic tree, we randomly permute the labels of the individuals (children) and calculate D for the data with permuted labels. This is repeated 1000 times to obtain a sample of differences (D_permuted), and to create a histogram that estimates this distribution. As the labels are permuted, the populations are now equivalent, and the average value of D should be closer to zero.

If the observed distance, D_observed, is in the tail of the sample of D_permuted, pairwise distances in the case group are significantly greater than those of the control. In other words, the microbial commu-

Table 3 Mean percent of total reads for all phyla identified in the case and control samples

Phylum	Cases			Controls		
	1	2	3	1	2	3
Bacteroidetes	53.27	62.19	69.17	76.13	60.07	54.65
Firmicutes	43.41	25.60	20.66	21.78	22.44	25.89
Proteobacteria	1.92	6.24	1.74	0.37	0.39	0.65
Fusobacteria	0.02	0.40	0	0	0.05	0.37

Phyla were identified by clustering at the 80% level of similarity using the Megablast results. The number following the case-control designation refers to time point of collection. The time point 3 samples were collected at the time of autoimmunity diagnosis for the case children. Those phyla that are statistically significantly higher ($P \leq 0.01$) in cases or controls are depicted in bold.

nities in healthy children are more similar to each other.

To test whether the microbial communities from controls are more similar to each other than those of cases during the time course, samples from the same time point were grouped before the clustering using cluster database at high identity with tolerance (Li and Godzik, 2006) and then submitted to principal coordinate analysis using weighted UniFrac (Lozupone *et al.*, 2006). **Q9**

Results

Stool samples were obtained from eight children in the Diabetes Prediction and Prevention study in Finland. These children represented four matched case-control pairs in which the samples of each autoimmune child were paired with samples from a nonautoimmune child of approximately the same age and HLA genotype (Table 1). Samples from each individual were taken at three time points (Table 1). The first sample was taken between 4 and 8 months of age before any child was found to have autoantibodies. In two of the four case subjects, the first autoantibody appeared about 6 months before the second collection point. In all four case subjects, the second autoantibody was detected within several months of the third collection point. Ultimately, all four cases were diagnosed with T1D.

After trimming the 16S rRNA sequences for low-quality sequences and bases, a total of 390 759 sequences were useful in the analysis of the 24 stool samples examined for this case-control experiment (Table 2). This represents an average of 15 709 sequences per sample. The number of sequences identified at six taxonomic levels shows that the number of species identified in all samples was 377, with the species defined at the 99% level of similarity. The classification of sequences at each of the six phylogenetic levels for all cases and controls at each collection time point was determined (Tables 3 and 4; Supplementary Tables 2–5).

Table 4 Mean percent of total reads for those species identified that showed significant differences in abundance between cases and controls at any time point of collection

Species (or OTU)	Cases			Controls		
	1	2	3	1	2	3
<i>Alistipes onderdonkii</i>	0	0	0.305	0.005	0.022	0.017
Bacterium mpn-isolate group 18	0.019	0.144	1.078	0	0.034	0.057
<i>Bacteroides caccae</i>	0.226	0.398	1.278	0.390	0.137	0.813
<i>Bacteroides dorei</i>	13.50	6.34	13.01	25.44	10.24	12.69
<i>Bacteroides fragilis</i>	3.95	1.10	3.80	12.42	11.28	6.99
<i>Bacteroides galacturonicus</i>	0	0.337	0	0	0.345	0
<i>Bacteroides ovatus</i>	1.53	9.82	9.94	1.75	1.10	0.62
<i>Bacteroides</i> sp. B2	1.17	1.25	0.09	1.46	0.02	0.04
<i>Bacteroides</i> sp. CJ78	2.95	0.84	5.10	2.57	0.69	0.33
<i>Bacteroides thetaiotaomicron</i>	0.34	3.65	8.29	0.35	0.71	1.95
<i>Bacteroides uniformis</i>	0.57	9.35	2.55	0.02	0.02	1.31
<i>Bacteroides vulgatus</i>	22.17	9.53	0.001	12.14	1.77	4.54
<i>Citrobacter</i> sp. CK3	0.015	0.009	0.009	0.002	0.006	0.010
Clostridiaceae DJF VR07	0.246	0.006	0.019	0.315	0.069	0.004
Clostridiales 80/3	0	0	0.485	0	0.231	0.019
Clostridiales 80/4	0.053	0.083	0.103	0.023	0.043	0.042
<i>Clostridium aldenense</i>	0	0.091	0.049	0.142	0.014	0.066
<i>Clostridium hathewayi</i>	0	0.288	0.050	0	0.007	0.058
<i>Clostridium innocuum</i>	0.004	0.040	0.260	0	0.007	0.007
<i>Clostridium orbiscindens</i>	0.082	0.071	0.249	0.103	0.029	0.116
<i>Clostridium</i> sp.	0	0.219	0.149	0	0.017	0.040
<i>Clostridium</i> sp. CJ67	0	0.006	0.011	0	0.001	0.002
<i>Dialister</i> sp. E2 20	0	1.296	0.187	0	0.002	0
<i>Enterobacter</i> sp. NII-26	0.041	0.026	0.019	0.016	0.015	0.013
<i>Escherichia coli</i>	1.195	4.094	0.689	0.149	0.087	0.078
<i>Eubacterium eligens</i>	0	0.582	0.483	0	1.114	1.480
<i>Eubacterium rectale</i>	0	0	0.706	0	0.716	2.859
<i>Faecalibacterium prausnitzii</i>	0	7.40	0.79	0	1.18	3.32
<i>Faecalibacterium</i> sp. DJF VR20	0	0.101	0.013	0	0.267	0.882
<i>Granulicatella para-adiacens</i>	0.09	0	0.003	0.005	0.007	0
<i>Haemophilus parainfluenzae</i>	0.141	0.005	0.016	0.044	0.148	0.082
Human intestinal firmicute CB47	0	0	0.002	0	0.443	1.066
Human intestinal firmicute CO19	0	0.749	1.800	0	6.316	7.114
Human intestinal firmicute CO35	0	0.050	0.125	0	1.342	0.554
Lachnospiraceae DJF RP14	0.03	0.166	0.194	0	0.042	0.105
<i>Lactobacillus rhamnosus</i>	0.737	0	0.012	0.046	0	0.019
<i>Odoribacter splanchnicus</i>	0.004	0.138	0.035	0.288	0.006	0.075
<i>Parabacteroides distasonis</i>	0.97	0.77	0.04	4.09	0.49	1.07
<i>Parabacteroides</i> sp. DJF B086	0.04	0	0	0.12	0.04	0.02
<i>Parasutterella excrementihominis</i>	0	0.60	0.29	0	0.01	0.01
<i>Ruminococcus bromii</i>	0	0.016	0.085	0	2.516	0.037
<i>Ruminococcus gnavus</i>	12.02	0.39	0.54	12.61	0.55	0.14
<i>Ruminococcus</i> sp. CB3	0.026	0.162	0.703	0	0.067	0.120
<i>Ruminococcus</i> sp. CO27	0	0	0.456	0	0.037	0.128
<i>Ruminococcus</i> sp. ID8	0	0.025	0.051	0	0.097	0.023
<i>Ruminococcus</i> sp. M10	0	0.011	0.009	0	0.022	0.015
<i>Streptococcus salivarius</i>	0.27	0.03	0.08	0.01	0.28	0.05
Swine fecal bacterium RF3E-Xyl1	0.008	0.187	0.037	0	0	0.013
<i>Veillonella atypica</i>	2.31	0.10	0.05	0.31	0.08	0.02
<i>Veillonella dispar</i>	5.49	0.06	0.19	0.99	0.21	0.02
<i>Veillonella</i> sp. ADV 3107.03	1.841	0.022	0.054	0.011	0.013	0.002

Species/OTUs were identified by clustering at the 99% level of similarity using the Megablast results. The number following the case/control designation refers to time point of collection. The time point 3 samples were collected at the time when the case children were positive for at least two autoantibodies. Those genera that are statistically significantly higher ($P \leq 0.01$) in cases or controls are depicted in bold. Species that represent at least 1% of the total reads in either cases or controls at the time of autoimmunity are highlighted in grey.

Q10

Phylum level

By far, the two most striking differences between the healthy (nonautoimmune) and autoimmune stool microbial communities were the differences within the two most abundant phyla, the Bacteroidetes and the Firmicutes (Table 3). In case samples, the Bacteroidetes sequences increased from 53.27% of

all sequences at the first collection point to 69.17% of all sequences at the third collection point, whereas in control samples, Bacteroidetes sequences decreased overtime from 76.13% to 54.65% of all sequences (Table 3). In contrast, the second most abundant phylum, the Firmicutes, expressed an inverse pattern. The Firmicutes sequences declined in case samples

overtime from 43.41% to 20.66% of all sequences, whereas they increased in control samples from 21.78% to 25.89% of all sequences (Table 3). The differences in abundance of both Bacteroidetes and Firmicutes observed between cases and controls were significant at all three time points.

Class and order levels

Similar trends occurred at the class level for the two most abundant classes, the Bacteroidetes and the Clostridia (Supplementary Table 2). As the Bacteroidetes become more abundant in the case samples overtime, they became less abundant in the control samples overtime. Conversely, as the abundance of Clostridia decreased overtime in the case samples, the Clostridia sequences increased in control samples. These trends repeated at the order level: the Bacteroidales sequences increased in case samples, whereas they decreased in control samples and the Clostriales sequences increased in control samples, whereas they decreased in case samples. Four other orders differed at one time point or another between cases and controls, but no clear trends emerged overtime (Supplementary Table 3).

Family level

Analysis at the family level showed that three families among the Firmicutes, the Ruminococcaceae, the Lachnospiraceae and the Eubacteriaceae, were significantly more abundant in controls than in cases at time point 3. In contrast, Veillonellaceae, also in the Firmicutes, were present at significantly higher levels in cases than in controls at all time points, but decline rapidly overtime in cases. Bacteroidaceae was the dominant family of the Bacteroidetes phylum and was significantly more abundant in cases than in controls (Supplementary Table 4). Among Bacteroidetes, the Porphyromonadaceae sequences were present at greater levels in controls than in cases at all time points. The Rikenellaceae family was also significantly more abundant in cases compared with controls at the second and third time points.

Genus level

At all time points and in both cases and controls, more than 92% and 81% of all case and control sequences, respectively, could be classified to the genus level. Hence, the majority of organisms in these samples are well known to science, and their physiology and morphology are understood. The *Bacteroides* is by far the dominant genus in these samples, representing over two-thirds of all sequences early in controls, and at the third time point in cases. In cases, the number of *Bacteroides* increases overtime and is significantly higher than that in controls at time points 2 and 3. In controls, however, *Bacteroides* declines from 66.47% to

38.63% of the bacteria present in stool samples. Two genera in the Firmicutes, *Eubacterium* and *Faecalibacterium*, increase dramatically overtime in controls compared with cases, and together represent more than 13% of the total population of bacteria in autoimmune children at time point 3 (Supplementary Table 5).

Species level

At the 99% identity level, 59% and 49% of case and control sequences, respectively, were classified to named and cultured species. Of the 377 bacterial species identified in these samples, 51 species statistically differed in abundance between cases and controls with a *P*-value <0.01 in at least 1 time point (Table 4). At the species level, specific taxa make large contributions to the overall differences between cases and controls. Perhaps the most striking example is *Bacteroides ovatus*. Nearly one-fourth of the difference between cases and controls within the phylum Bacteroidetes can be explained by this single species (Figure 1a). At the first time point, there are slightly more *B. ovatus* sequences in control samples than in case samples, but, at the time of autoimmunity, there are 16-fold more *B. ovatus* sequences in cases than in controls (Table 4).

Most other *Bacteroides* species are also present in greater proportions in cases than in controls. However, two *Bacteroides* species, *B. vulgatus* and *B. fragilis*, are observed at much higher levels in controls than in cases at the third time point and represent more than 11% of total sequences in the control samples.

Another bacterium of interest is the human firmicute CO19 (Table 4). Although not sufficiently characterized to be given a species name, this organism has been cultured (Hayashi *et al.*, 2002). At the 99% similarity level, more than 7% of all control sequences at the third time point clustered with human firmicute CO19. Although increasing in both cases and controls overtime, the abundance of the human firmicute CO19 at the third time point seems to be nearly fourfold greater in controls than in cases.

At the third collection time point, just 37 of 377 species differed significantly between cases and controls (Table 4). The 15 species that are higher in abundance in controls represent 30% of all control sequences at time point 3. Similarly, the 22 species that are much higher in cases than in controls represent nearly half of all sequences present in cases in autoimmune children.

The unclassified sequences

At all taxonomic levels, the sequences that cannot be classified to known taxa are significantly more abundant in control samples than in cases at the third time point (Figure 1). As expected, the number of unclassified sequences increases as the phylo-

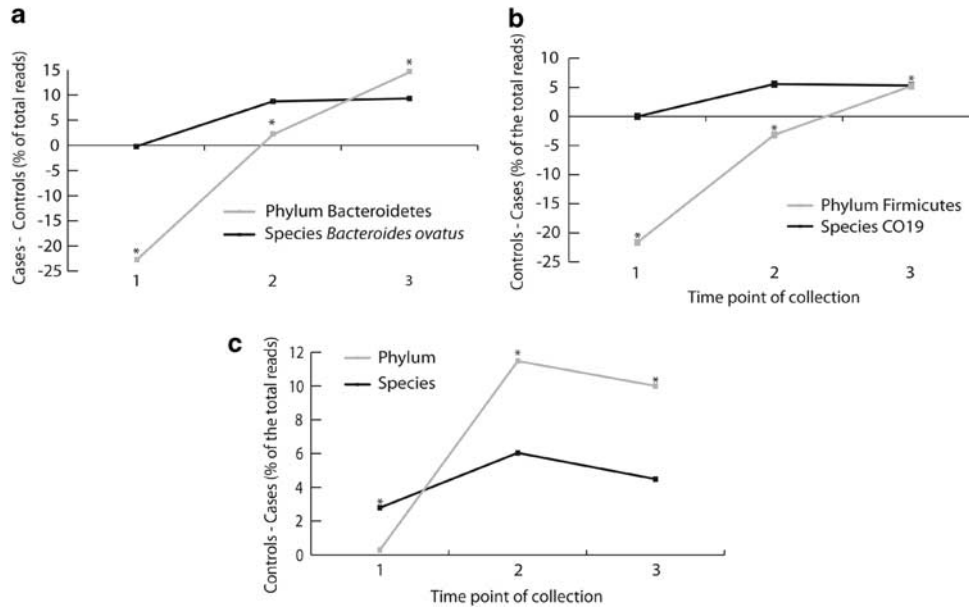


Figure 1 Significant differences in taxa between cases (autoimmune) and controls (healthy). Samples were collected approximately 4 months, 1 year and 2 years after birth, represented, respectively, as time points 1, 2 and 3: (a) increasing numbers of Bacteroidetes in cases overtime compared with controls; (b) increasing numbers of Firmicutes in controls overtime compared with cases; and (c) higher proportion of unclassified sequences in controls compared with cases. Significant differences between cases and controls are designated by a star ($P \leq 0.002$).

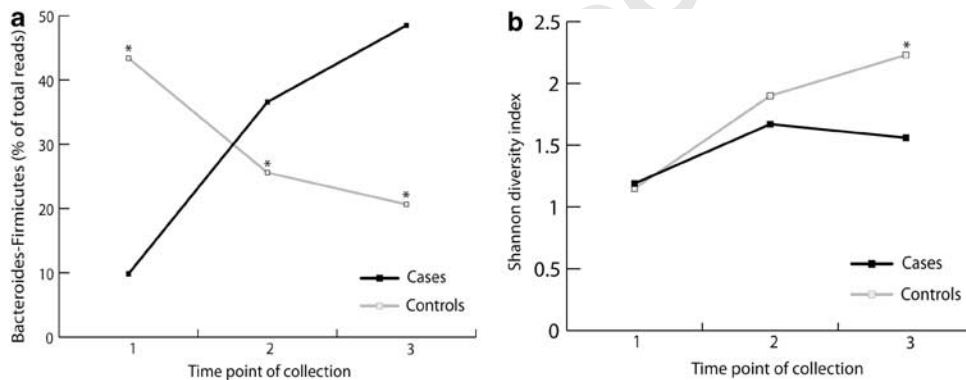


Figure 2 Bacterial community differences between cases and controls during autoimmunity development in cases; (a) significant increase in Bacteroidetes with concomitant decrease in Firmicutes in cases compared with controls (P -value ≤ 0.01 at all time points; (b) significantly higher ($P < 0.05$) Shannon diversity index in controls compared with cases in time point 3. Significant differences between cases and controls are designated by a star. The P -values for time points 1, 2 and 3 are (a) 0.0000, 0.0000 and 0.0000, and (b) 0.80, 0.33 and 0.03, respectively.

genetic classification becomes more restrictive. Hence, the proportion of unclassified sequences at the phylum level is $< 1\%$ of all sequences, whereas the proportion of unclassified sequences at the species level is over 30% in some control samples.

Community diversity indices

At the genus level, bacterial diversity, as measured through the Shannon index, increases overtime in control samples (Figure 2). At the third time point, the diversity index is significantly higher in control communities than in case communities (P -value < 0.05).

In addition, the analysis of microbial communities using principal coordinate analysis shows that the bacterial communities in control samples are more similar to each other than are the bacterial communities in case samples (Figures 3 and 4). No differences in community diversity were observed at time point 1 at the 10% confidence interval. However, the average distance between any pair of cases was significantly higher than between any pair of controls at the second and third collection points at the 5% and 10% level of confidence, respectively. Although these confidence intervals are relatively high, they are remarkable, given that only six pairwise comparisons were available in cases and control.

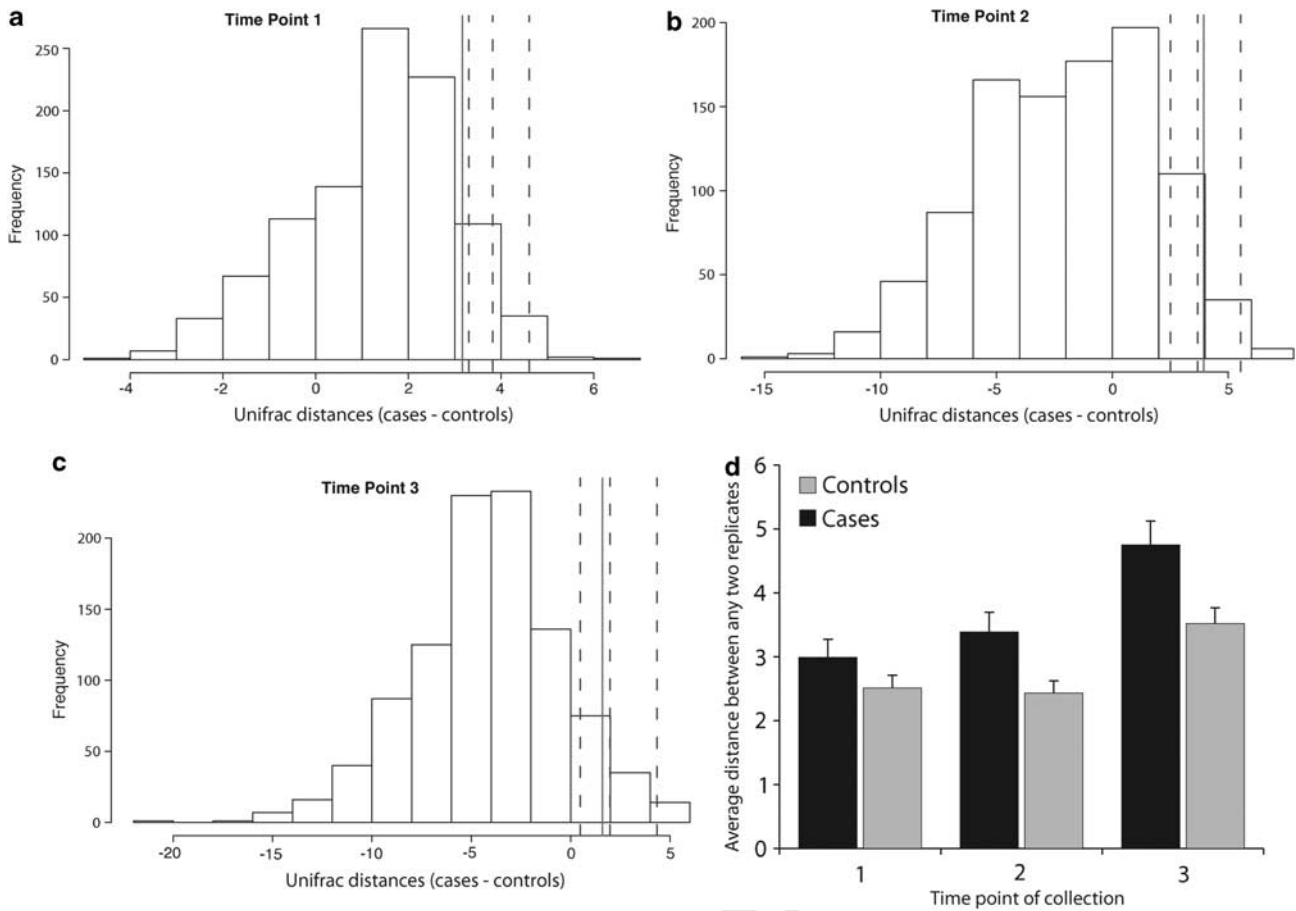


Figure 3 Histograms showing the permutation test based on the UniFrac significance obtained from the three time points (a, b and c). Dashed blue lines represent the 0.10, 0.05 and 0.01 quantiles, and the red line indicates the value of the observed difference. No differences in community diversity were observed at time point 1 at the 10% confidence interval. However, the average distance between any pair of cases was significantly higher than that between any pair of controls at the second and third collection points at the 5% and 10% level of confidence, respectively. A summary of data over all time points is shown in (d). (The color version of this figure is available in online version only).

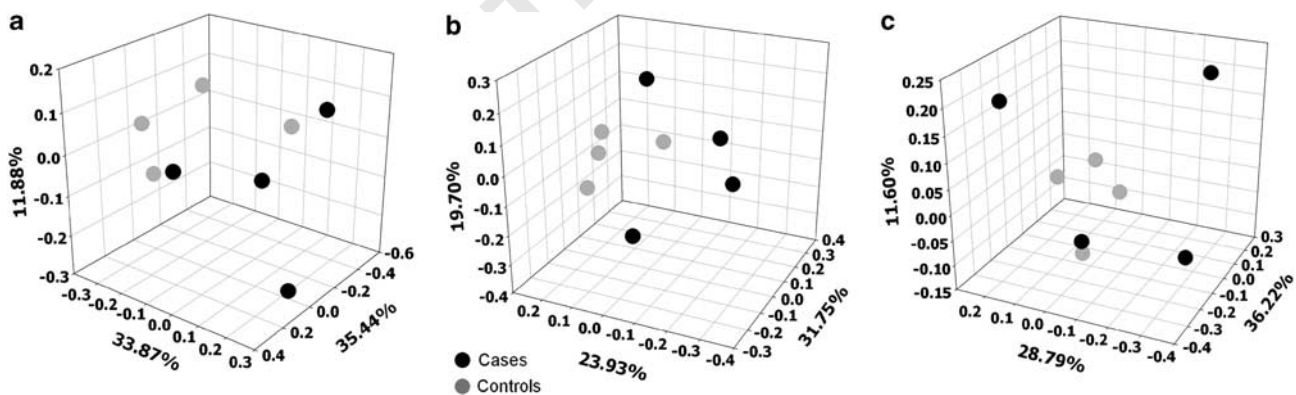


Figure 4 Principal coordinate analysis for the case and control communities at time points 1 (a), 2 (b) and 3 (c).

Discussion

The data presented here were analyzed at various taxonomic levels and at the community level to identify specific taxa and community characteristics that differ between microbiomes in healthy children

and autoimmune microbiomes in children who develop T1D. One striking result of this analysis is the decline in Firmicutes and increase in Bacteroidetes in the gut microbiome overtime as children become autoimmune; in contrast, Firmicutes increase as Bacteroidetes declines in healthy

children (Figures 1a, b and 2a). These two phyla comprise more than 80% of the sequences at all time points in both cases and controls. Phylogenetic analysis revealed that virtually all changes that occurred overtime between cases and controls within the Bacteroidetes phylum were attributed to a single genus, *Bacteroides*. More than one-fifth of the changes that occurred within the genus *Bacteroides* can be ascribed to a single species, *B. ovatus* (Figure 1a). Nearly all of the increased abundance in Firmicutes in control samples compared with cases can be attributed to a single order, Clostridiales (Supplementary Table 3). Over 17% of that increase is ascribed to a single species represented in the literature by a single strain, human intestinal firmicute CO19 (Figure 1b), which was isolated from the intestine of a healthy human subject (Hayashi *et al.*, 2002).

The dysbiosis observed at phylum level between Bacteroidetes and Firmicutes in the human gut has been described in several human disorders. As described in previous studies, the ratio between Firmicutes and Bacteroidetes in human type 2 diabetes declines compared with controls (Larsen *et al.*, 2010). In Crohn's disease, both Bacteroidetes and Firmicutes seem to decline, whereas Proteobacteria increase (Frank *et al.*, 2007; Willing *et al.*, 2009). The reverse was observed in obesity, the imbalance is observed by the reduction in the *Bacteroides* proportion in obese human subjects, with a corresponding increase in the Firmicutes/Bacteroidetes ratio. Among Firmicutes, *Lactobacillus* numbers seem to increase in obese patients (Turnbaugh *et al.*, 2006; Armougom *et al.*, 2009).

In all four pairs of cases and controls at the third time point, 22 bacterial species were significantly more abundant ($P < 0.01$) in cases than in controls. Of those 22 species, 5 species (bacterium mpn-isolate group 18, *B. ovatus*, *Bacteroides* sp. CJ78, *B. thetaiotaomicron* and *B. uniformis*) represented more than 1% of all sequences each. Similarly, 15 bacterial species were found to be significantly more abundant ($P < 0.01$) in controls compared with cases in all four pairs of children at the third time point. Of those species (*Bacteroides fragilis*, *B. vulgatus*, *Eubacterium eligens*, *E. rectale*, *Faecalibacterium prausnitzii*, human intestinal firmicute CB47 and human intestinal firmicute CO19), each represented at least 1% of all sequences. Thus, this study identified highly abundant bacteria in the gut microbiomes that are either negatively or positively correlated with the development of autoimmunity in children who are at high risk for the onset of T1D.

Three lines of evidence suggest that, overtime, nonautoimmune children are able to build a healthy and stable gut microbiome, whereas the microbiomes of autoimmune children are less diverse and unstable. Hence, we now refer to this unhealthy, unstable microbiome as the autoimmune microbiome for T1D. The evidence presented here does not prove that the microbiomes of case children are

less healthy than those of control children. These data, however, do build a framework on which specific questions can be asked regarding the autoimmune microbiome. These three lines of evidence are as follows:

First, as children become older, changes in the composition of the intestinal microbiota take place depending on the feeding stage of childhood (Favier *et al.*, 2002). Near 2 years of age, the microbiota of children become more similar to the adult and present large changes thereafter (Tiihonen *et al.*, 2009). In this study, healthy children contain a significantly higher number of poorly known, unclassified microorganisms than do autoimmune children overtime (Figure 1c). The presence of a greater proportion of unclassified organisms in microbiomes of healthy subjects suggests that these individuals may host a greater proportion of non-pathogenic organisms than do autoimmune subjects. Historically, microbiology has focused on the characterization of microorganisms that cause disease, as well as their mechanisms of pathogenesis (de Kruif, 1926; Lechevalier and Solotorovsky, 1974; Sapp, 2009). Thus, if a bacterium is found that causes disease, the scientific community studies this organism immediately and with great intensity. As a result, if a bacterium is a pathogen, it is far more likely to be classified and characterized than if it is not pathogenic. Thus, if the autoimmune microbiome possesses more organisms with a pathogenic potential than the healthy microbiome, the number of unclassified sequences would be expected to be lower in autoimmune subjects compared with healthy ones.

Second, bacterial diversity increased in controls, but remained constant in cases (Figure 2b). The same phenomenon was described in Crohn's disease, in which decreased diversity of fecal microbiota of cases in comparison with controls (Manichanh *et al.*, 2006; Sartor, 2008). A decrease in microbial diversity in the gut was also observed in obese individuals in relation with lean individuals (Turnbaugh *et al.*, 2009). This result may be caused by a disturbance, likely an immunological response, in autoimmune children that reduces the microbial diversity in these case subjects. In natural systems, microbial communities are very sensitive to disturbance and are not sufficiently resilient to return to their original states (Allison and Martiny, 2009). Reduced community composition decreases the set of ecosystem processes available in this community. In the human intestine, limited diversity may lead to a reduced capacity to digest a diverse diet, leading to lower energy levels in affected individuals. Hence, a gut microbial community with less diversity may lead to less healthy individuals or may be an indicator of an unhealthy state.

Finally, the mean molecular distance between any two case samples is greater than the molecular distance between any two control samples collected

at any of the time points (Figures 3 and 4). Thus, although control communities are more diverse than case communities, overall, control communities are much more similar to each other and show a community stability that is lacking in case communities. This community difference between any two case samples suggests that the autoimmune communities are much less stable.

Of all the differences observed between autoimmune and healthy microbiomes, there are two indicators of the autoimmune microbiome that can be observed at the earliest time points, before the first appearance of autoantibodies in case subjects. The first of these characteristics is the instability of the autoimmune microbiome. The second characteristic is the high ratio of Firmicutes to Bacteroidetes in cases, which is observed within the first 6 months after birth, compared with the low ratio found in controls. Both these characteristics may be early diagnostic markers of pending autoimmunity that can be observed before the first appearance of autoantibodies in serum. This knowledge may allow early interventions that can delay or prevent autoimmunity in children by alerting the intestinal microbiome or by other means.

Taken collectively, these findings support the concept of an autoimmune microbiome for T1D. Although the number of samples used here is too low to make any firm conclusions, these data suggest that the autoimmune microbiome for this disease tends to have more classified members but decreased diversity and reduced stability when compared with a healthy microbiome. The forces at work that result in the autoimmune microbiome are unknown, but may be related to the breeching of the epithelial layer combined with abnormal immune responsiveness. This work provides a model of the autoimmune microbiome for T1D and delineates four features that can serve as springboards for further analysis. The defining characteristics of the stable, healthy microbiome and the unstable, unhealthy autoimmune microbiomes are reminiscent of the famous first line from Tolstoy's Anna Karenina: 'All happy families are alike; each unhappy family is unhappy in its own way.'

Acknowledgements

This research was supported by the Florida Agricultural Experiment Station. We thank Professor Vassiliki Betty Smocovitis of the University of Florida for her advice on the history of microbiology.

References

Allison SD, Martiny JBH. (2009). Resistance, resilience, and redundancy in microbial communities. *Proc Natl Acad Sci USA* **105**: 11512–11519.

- Armougom F, Henry M, Vialettes B, Raccach D, Raoult D. (2009). Monitoring bacterial community of human gut microbiota reveals an increase in *Lactobacillus* in obese patients and *Methanogens* in anorexic patients. *PLOS One* **4**: e7125.
- Brugman S, Klatter FA, Visser JTT, Wildeboer-Veloo ACM, Harmsen HJM, Rozing J *et al.* (2006). Antibiotic treatment partially protects against type 1 diabetes in the Bio-Breeding diabetes-prone rat. Is the gut flora involved in the development of type 1 diabetes? *Diabetologia* **49**: 2105–2108.
- Calcinaro F, Dionisi S, Marinaro M, Candeloro P, Bonato V, Marzotti S *et al.* (2005). Oral probiotic administration induces interleukin-10 production and prevents spontaneous autoimmune diabetes in the non-obese diabetic mouse. *Diabetologia* **48**: 1565–1575.
- de Kruif P. (1926). *Microbe Hunters*. Harcourt, Brace and Co.: New York.
- Dicksved J, Halfvarson J, Rosenquist M, Jarnerot G, Tysk C, Apajalahti J *et al.* (2008). Molecular analysis of the gut microbiota of identical twins with Crohn's disease. *ISME J.* **2**: 716–727.
- Edgar RC. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *Bioinformatics* **5**: 1–19.
- Favier CF, Vaughan EE, De Vos WM, Akkermans ADL. (2002). Molecular monitoring of succession of bacterial communities in human neonates. *Appl Environ Microb* **68**: 219–226.
- Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR *et al.* (2007). Molecular-phylogenetic characterization of microbial community imbalance in human inflammatory bowel diseases. *Proc Natl Acad Sci USA* **104**: 13780–13785.
- Giongo A, Crabb DB, Davis-Richardson AG, Chauliac D, Moberley JM, Gano KA *et al.* (2010a). PANGEA: pipeline for analysis of next generation amplicons. *ISME J* (in press).
- Giongo A, Davis-Richardson AG, Crabb DB, Triplett EW. (2010b). TaxCollector: tools to modify existing 16S rRNA databases for the rapid classification at six taxonomic levels. Submitted.
- Harrison LC, Honeyman MC, Morahan G, Wentworth JM, Elkassaby S, Colman PG *et al.* (2008). Type 1 diabetes: lessons for other autoimmune diseases? *J Autoimmun* **31**: 306–310.
- Hayashi H, Sakamoto M, Benno Y. (2002). Phylogenetic analysis of the human gut microbiota using 16S rDNA clone libraries and strictly anaerobic culture-based methods. *Microbiol Immunol* **46**: 535–548.
- Hong S-H, Bunge J, Jeon S-O, Epstein SS. (2006). Predicting microbial species richness. *Proc Natl Acad Sci USA* **103**: 117–122.
- Kupila A, Muona P, Simell T, Arvilommi P, Savolainen H, Hamalainen AM *et al.* (2001). Feasibility of genetic and immunological prediction of Type I diabetes in a population-based birth cohort. *Diabetologia* **44**: 290–297.
- Larsen N, Vogensen FK, van den Beg FWL, Nielsen DS, Andreassen AS, Pedersen BK *et al.* (2010). Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS One* **5**: e9085.
- Lechevalier HA, Solotorovsky M. (1974). *Three Centuries of Microbiology*. Dover Publications: New York.
- Ley RE, Turnbaugh PJ, Klein S, Gordon JL. (2006). Microbial ecology—Human gut microbes associated with obesity. *Nature* **444**: 1022–1023.

Q11

Q20

Q12

Q13

- Li W, Godzik A. (2006). CD-HIT: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658–1659.
- Lozupone C, Hamady M, Knight R. (2006). UniFrac—An online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics* **7**: 371.
- Manichanh C, Rigottier-Gois L, Bonnaud E, Gloux J, Pelletier E, Frangeul L *et al.* (2006). Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* **55**: 205–211.
- Matsuzaki T, Nagata Y, Kado S, Uchida K, Kato I, Hashimoto S *et al.* (1997). Prevention of onset in an insulin-dependent diabetes mellitus model, NOD mice, by oral feeding of *Lactobacillus casei*. *APMIS* **105**: 643–649.
- Nejentsev S, Sjorros M, Soukka T, Knip M, Simell O, Lovgren T *et al.* (1999). Population-based genetic screening for the estimation of Type 1 diabetes mellitus risk in Finland: selective genotyping of markers in the HLA-DQB1, HLA-DQA1 and HLA-DRB1 loci. *Diabetic Med* **16**: 985–992.
- Ochoa-Reparaz J, Mielcarz DW, Ditrio LE, Burroughs AR, Foureau DM, Haque-Begum S *et al.* (2009). Role of gut commensal microflora in the development of experimental autoimmune encephalomyelitis. *J Immunol* **183**: 6041–6050.
- Ou GW, Hedberg M, Horstedt P, Baranov V, Forsberg G, Drobni M *et al.* (2009). Proximal small intestinal microbiota and identification of rod-shaped bacteria associated with childhood celiac disease. *Am J Gastroenterol* **104**: 3058–3067.
- Roesch RFW, Casella G, Simell O, Krischer J, Wasserfall CH, Schatz D *et al.* (2009a). Influence of sample storage on bacterial community diversity in fecal samples. *Open Microbiol J* **3**: 40–46.
- Roesch RFW, Lorca GL, Casella G, Giongo A, Naranjo A, Pionzo AM *et al.* (2009b). Culture-independent identification of gut bacteria correlated with the onset of diabetes in a rat model. *ISME J* **3**: 536–548.
- Sapp J. (2009). *The new foundations of evolution: on the tree of life*. Oxford University Press: New York.
- Sartor RB. (2008). Microbial influences in inflammatory bowel diseases. *Gastroenterology* **134**: 577–594.
- Schwartz RF, Neu J, Schatz D, Atkinson MA, Wasserfall C. (2007). Comment on: Brugman S *et al.* (2006). Antibiotic treatment partially protects against type 1 diabetes in the Bio-Breeding diabetes-prone rat. Is the gut flora involved in the development of type 1 diabetes? *Diabetologia* **49**: 2105–2108. *Diabetologia* **50**: 220–221.
- Shannon CE, Weaver W. (1949). *The mathematical theory of communication*. Urbana IL: University of Illinois Press.
- Tamura K, Dudley J, Nei M, Kumar S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molec Biol Evol* **24**: 1596–1599.
- Tiihonen K, Ouwehand AC, Rautonen N. (2009). Human intestinal microbiota and healthy ageing. *Ageing Res Rev*, doi:10.1016/j.arr.2009.10.004.
- Tjellstrom B, Stenhammar L, Hogberg L, Faith-Magnusson K, Magnusson KE, Midtvedt T *et al.* (2005). Gut microflora associated characteristics in children with celiac disease. *Am J Gastroenterol* **100**: 2784–2788.
- The TEDDY Study Group (2007). The Environmental Determinants of Diabetes in the Young (TEDDY) study: study design. *Pediatr Diabetes* **8**: 286–298.
- Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JL. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**: 1027–1031.
- Turnbaugh PJ, Hamady M, Yatsunencko T, Cantarel BL, Duncan A, Ley RE *et al.* (2009). A core gut microbiome in obese and lean twins. *Nature* **457**: 480–485.
- Vaarala O, Atkinson MA, Neu J. (2008). The 'Perfect Storm' for type 1 diabetes. The complex interplay between intestinal microbiota, gut permeability, and mucosal immunity. *Diabetes* **57**: 2555–2562.
- Valladares R, Sankar D, Li N, Williams E, Mahmou ASA, Lai K-K *et al.* (2010). *Lactobacillus johnsonii* N6.2 mitigates the development of type 1 diabetes in BB-DP rats. *PLoS One* (in press).
- Wen L, Ley RE, Volchkov PY, Stranges PB, Avanesyan L, Stonebraker AC *et al.* (2008). Innate immunity and intestinal microbiota in the development of Type 1 diabetes. *Nature* **455**: 1109–1113.
- Willing B, Halfvarson J, Dicksveld J, Rosenquist M, Jarnerot G, Engstrand L *et al.* (2009). Twin studies reveal specific imbalances in the mucosa-associated microbiota of patients with ileal Crohn's disease. *Inflamm Bowel Dis* **15**: 653–660.
- Yadav H, Jain S, Sinha PR. (2007). Antidiabetic effect of probiotic dahi containing *Lactobacillus acidophilus* and *Lactobacillus casei* in high fructose fed rats. *Nutrition* **23**: 62–68.
- Zhang Z, Schwartz S, Wagner L, Miller W. (2000). A greedy algorithm for aligning DNA sequences. *J Comput Biol* **7**: 203–214.

Q16

Q17

Q18

Q21

Q19

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)