GASTROINTESTINAL MICROBIOME SIGNATURES OF PEDIATRIC PATIENTS WITH IRRITABLE BOWEL SYNDROME

Delphine M. Saulnier[1, 6]*, Kevin Riehle[3,5]*, Toni-Ann Mistretta[1, 6], Maria-Alejandra Diaz[1,6], Debasmita Mandal[7], Sabeen Raza[1,6], Erica M. Weidler[4], Xiang Qin[8], Cristian Coarfa[3,5], Aleksandar Milosavljevic[3,5], Joseph F. Petrosino[2,8], Sarah Highlander[2,8], Richard Gibbs[8], Susan V. Lynch[7], Robert J. Shulman[4, 9], James Versalovic[1,2, 3, 4, 5, 6]

Departments of [1]Pathology & Immunology, [2]Molecular Virology & Microbiology, [3]Molecular & Human Genetics, and [4]Pediatrics, Baylor College of Medicine, Houston, TX; [5]Bioinformatics Research Laboratory, Baylor College of Medicine, Houston, TX; [6]Department of Pathology, Texas Children's Hospital, Houston, TX; [7]Department of Medicine, UCSF, San Francisco, CA; [8]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; [9]Children's Nutrition Research Center, Houston, TX

SUPPLEMENTARY APPENDIX

This appendix has been provided by the authors to give readers additional information about their work.

SUPPLEMENTARY METHODS

16S rRNA sequence-based surveys of the distal gut microbiota

DNA samples extracted were used for 16S rRNA sequence-based survey. Prior to sequencing, amplicons from the individual PCR samples were quantified using the Quant-iT PicoGreen double-stranded DNA assay (Invitrogen, Carsbald, CA). High Fidelity PCR reaction was carried out in 96 well plates. 16 µL of master mix composed of 13.85 µL RNAse/DNAse free water, 2 µL 10X AccuPrime PCR Buffer II, and 0.15 µL AccuPrime™ Taq DNA Polymerase (Invitrogen) were mixed into individual wells in the 96 well reaction plate. The covered plate was spun and centrifuged at 2000 rpm to collect sample at the bottom of the wells. For the initial reaction, a 2 µL sample of DNA diluted 1:1 in water was added to the reaction wells. Barcoded primers (2 µM) from the primer plate were added to corresponding wells in the 96 well PCR plate. Two different PCRs were set up separately with a set of barcoded primers targeting the V3-V1 region and V3-V5 regions. The V1-V3 and V3-V5 regions of the 16S rRNA gene were amplified by PCR using bar-coded universal primers 27F and 534R (V1-V3) or 357F and 936R (V3-V5) containing the A and B sequencing adaptors (454 Life Sciences, Branford, CT) obtained from Operon. The sequence of the forward primer B-27F was 5'-cctatcccctgtgtgccttggcagtctcaGAGAGTTTGATCCTGGCTCAG-3, and that of primer B-354F was 5'-cctatcccctgtgtgccttggcagtctcaGCCTACGGGAGGCAGCAG-3'
(the sequence of the B adaptor is shown in lowercase letters). The sequence of the reverse primer A-534R was 5'-ccatctcatccctgcgtgtctccgactcagNNNNNATTACCGCGGCTGCTGG-3's and that of

primer A-926R was

5’ccatctcatccctgcgtgtctccgactcagNNNNNCCGTCAATTCMTTTRAGT

(the sequence of the A adaptor is shown in lowercase letters, and N represents a bar code that is unique for each sample). The plate was securely sealed with clear adhesive plate seal and vortexed vigorously, before being centrifuged briefly at 2000 rpm. The plate was then placed in a thermocycler. Cycling conditions were 95°C for 2 min, followed by 30 cycles at 94°C for

20 s, 50°C for V1V3 primers sets or 56°C for V3V5 primer sets for 30 sec, and 72°C for 5 min, with a final extension period of 20 min at 4°C. PCR products were cleaned using AmPure Beads Agencourt (Beckman Coulter, Beverly, MA) using 1.8x volume beads. Beads were eluted with 25 µL 1x low TE, pH 8.0 and transferred to a new 96-well plate. PCR products were quantified using Quant-IT ds DNA high sensitivity assay (Invitrogen) according to the manufacturer’s specifications. All samples were diluted according to the sample that had the lowest concentration. Equal volumes of each (5-10 µL) sample were pooled and then concentrated using MinElute columns (Qiagen, Valencia, CA). Samples were then transferred for the 454 library completion. DNA amplification and 454 sequencing were performed at the Human Genome Sequencing Center (HGSC) in Houston, TX.

454 Sequence Analysis

Samples were isolated and quality filtered from each multiplexed Standard Flowgram Format (SFF) file. Individual samples are passed through a quality filtering algorithm that follows the following guidelines: >= 200 nt, >= 20 average quality score,

match exact bar code and exact proximal (A) primer, cut at first 'N' or 'n', and cut at matched distal (B) primer (allowing up to 4 mismatches).

A local copy of RDP (Ribosomal Database Project) Classifier[1] was utilized to classify sequences from each sample using an 80% bootstrap cut off. Classified samples were separated based on taxonomic depth, combined based on taxonomic name, and normalized to account for variable sequencing depth. Taxonomically binned data were used as an input for abundance and diversity measurements.

OTU (Operational Taxonomic Unit) tables were produced from a local copy of QIIME 1.1.0[2] using a chained multi-step OTU picking algorithm. De-convoluted sequences were concatenated into a single fasta file and a mapping file was created to contain the sample names and clinical metadata. The first part of the multi-step OTU picking algorithm is to pick OTUs from the sequence file using a rough, fast method (i.e. PrefixSuffix). PrefixSuffix collapses sequences that are identical in the first and last bases at the default length of 50 nt. Representative sequences were then chosen from the PrefixSuffix output and then a slow, more rigorous OTU picker, CD-HIT [3], was used. CD-HIT applies a longest-sequence-first list removal algorithm to remove and cluster sequences above a certain identity threshold (i.e. 97%). The resultant OTU maps from the fast and slow method are merged and we can pick our final representative set. Finally, we can produce an OTU Table based on the abundance of each representative sequence in relation to the individual samples.

A local copy of RDP classifier was used to assign taxonomy and label each element in the OTU Table. Representative sequences were aligned to the

Greengenes[4] core gene set using PyNAST[3] which does not allow new sequences to introduce new gap characters into the template database. The resultant alignment is a NAST[5] aligned file that contains sequences aligned to the core gene set (at a max length of 7,682). Next, the alignment is filtered using the Lane mask[6] which limits calculations to 1,287 conserved columns of aligned characters. Finally, we can construct a phylogenetic tree from the filtered NAST alignment using FastTree[7]. FastTree infers approximate maximum likelihood phylogenetic trees using the Jukes-Cantor[8] model of nucleotide evolution and incorporates the "CAT" approximation for varying rates of evolution across sites.

Alpha diversity measurements were calculated with Biodiversity[9], an R[10] package that incorporates a GUI to perform biodiversity and community ecology analysis (i.e. richness, evenness, rank abundance, Chao1[11], Shannon[12], and Renyi's diversity indices[13]. Beta diversity calculations were produced for both phylogenetic and non-phylogenetic metrics. Phylogenetic beta diversity calculations were produced for weighted and unweighted UniFrac[14] distance metrics. UniFrac distance metric incorporates evolutionary distances of sequences and allows for the phylogenetic comparison of multiple microbial communities. QIIME was used to generate plots for beta diversity metrics via principal coordinates analysis (PCoA) plots[15] and a custom mapping file was used to color data points based on clinical meta data. PCoA plots can be analyzed to determine if any significant clusters result among the clinical metadata coloring, signifying differences in the microbial communities.

Machine learning algorithms are useful in determining the strength of meta data clusters (bagging, binning, etc.) as well as listing the most important variables involved

in discriminating two groups of samples (feature selection). The algorithm randomForest enables one to classify groups of samples by constructing a classification tree, randomly sampling the predictors, choosing the best splitting variables, and predicting new data by combining the predictions from all the trees. Random Forest allows us to produce an estimate of the error rate ["out-of-bag" (OOB)] and list the highest performing variables. We utilized an R package called Boruta[16] due to the fact that randomForest does not explicitly perform feature selection[17]. Boruta is a feature selection algorithm that implements random Forest and finds all relevant variables in a data set. Boruta either confirms or rejects variables as being relevant based on their classification performance versus random probes.

To evaluate the stability of the pediatric microbiome in health and disease, OTU sequences were analyzed with respect to collection times for 8 IBS samples (3x), 2 IBS (2x) and 6 Healthy subjects (3x). A non-parametric method, Mann-Whitney or Wilcoxon Rank Sum Test Hollander, M. and Wolfe, D. A. (1999). *Nonparametric Statistical Methods* (2nd Ed.), was applied to normalized OTU counts to compare Healthy samples from IBS samples or RDP normalized values to investigate the stability of the microbiomes in children with IBS and Healthy Subjects during 2-3 month intervals.

G2 Phylochip analysis

For G2 Phylochip analysis, data was normalized, and taxa deemed present if they exhibited pf (positive fraction) value ≥0.95 (95% of all probes in a given probe set for an individual taxon report fluorescence). The dataset was filtered using this cut-off;

fluorescence intensity values for those taxa deemed present in at least one sample, were log-transformed prior to subsequent statistical analyses. Metrics describing bacterial community richness (number of taxa detected), evenness (relative distribution of taxa within a community) and diversity (gross community metric based on richness and evenness) were calculated using XLSTAT Version 2009.4.07 (Addinsoft, New York, NY). Analysis of compositional patterns or relationships within the microbiomes in pediatric IBS and healthy subjects was analyzed using the Pearson correlation index ([17]). Taxon-level changes in relative abundance between healthy and IBS groups were determined using Mann Whitney test with multiple correction testing. Comparative analysis of 4 IBS patients reporting the highest severity of pain and 4 healthy subjects with no pain, was performed using T-Test.

PCR validation of bacterial taxa detected only by G2 Phylochip

To confirm that specific members of the community identified by the microarray as more abundant in IBS subjects, but not detected by 454, were truly present, two primer sets based on G2 Phylochip probes were designed to detect *Nitrospiraceae* (otu_0984; G2 taxonomy; http://greengenes.lbl.gov), using NTF1 5'-CAATGGGCGAAAGCCTGACGCAGC-3'& NTR1 5'-TACGCCCAGTAAATCCGAACAAC GC-3', and NTF2 5'-ACGGGCCAGACTCCTACGGGAGG-3' & NTR2 5'-TACGCCCAGTAAATCCGAACAACG-3'. Additionally, two sets of primers were designed to detect *Coxiellaceae* (otu_7893) using CXF1 5'-GTAATACAGAGGGTGCGAGCGTTAA-3' & CXR1 5'-ATTTCACCGCTACACCGGAAATTCC-5', and CXF2 5'-

TCGTGTTGTGAGATGTTGGGTTAAG-3' & CXR2 5'-TGCAGACTCCAATCCGGACTACGAG-3'. Gradient PCR (50-66°C) was performed to determine optimal conditions for nucleic acid amplification. PCR was performed with a subset of samples from IBS and healthy subjects with sufficient remaining DNA using FastStart Master Mix (Roche, Indianapolis, IN) and amplified using the following program 10 min at 95°C, followed by 35 cycles of 15s at 95°C, 20s at 60°C, 30s at 72°C. Amplicons were cleaned after PCR using QIAquick kit (Qiagen) were then submitted for bi-directional sequencing (DeWalch, Houston, TX), and sequence identities were assigned by comparing reads to the RDP [31] and Greengenes databases [25]. Additionally, amplification for the Nitrospira phylum was performed by using primers 27f and Nspira705r, or NSR1113f and NSR1264r as previously described [32],[33].

Universal and Targeted Bacterial qPCR

Total bacterial load was evaluated by real-time PCR with universal primers targeting the 16S rRNA gene (338F 5'-ACTCCTACGGGAGGCAGCAG and Bact-518R 5'-ATTACCGCGGCTGCTGG)[18]. Real-time PCR reactions were performed in triplicate using "Brilliant® II SYBR® Green QPCR Master Mix" (Stratagene, La Jolla, CA), with a Mx-3005 QPCR thermal cycler and real-time DNA detection system (Stratagene). Data were quantified using software in the MxPro-3005P Multiplex Quantitative PCR System (Stratagene) and normalized against wet weight of the fecal samples.

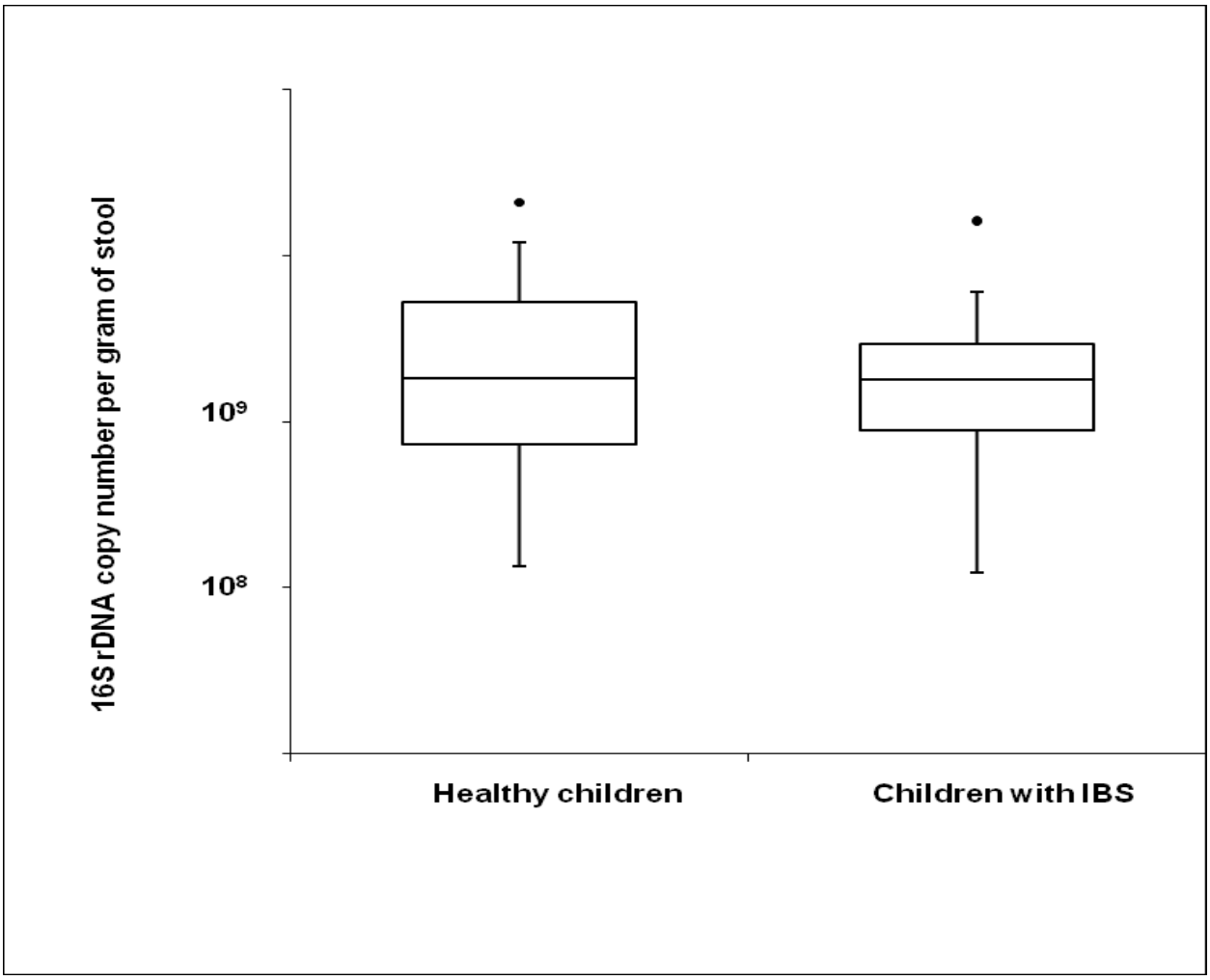
To confirm that specific members of the community identified by the microarray as more abundant in IBS subjects, but not detected by 454, were truly present, two primer sets

based on PhyloChip probes were designed to detect ***Nitrospiraceae*** (otu_0984; G2 taxonomy; http://greengenes.lbl.gov), and Coxiellaceae (otu_7893). The method used for PCR validation is described in greater detail in the supplementary methods.

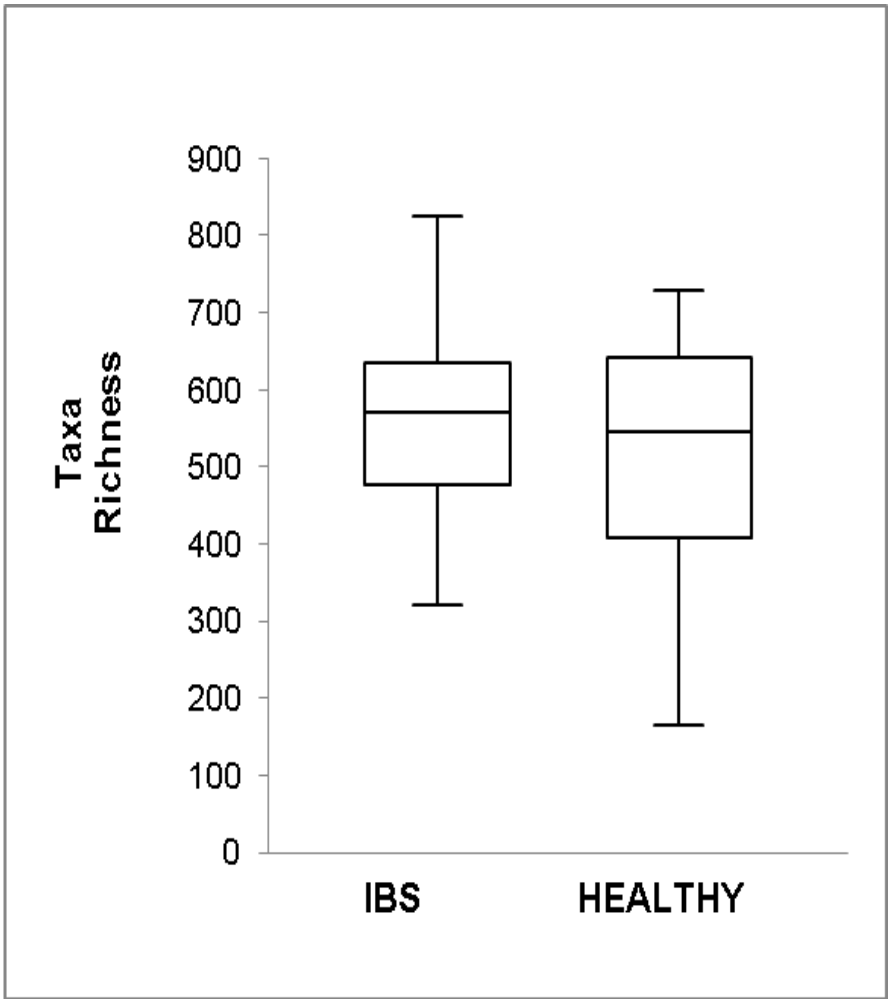
Evaluation of Human Biomarker of Inflammation

Calprotectin was measured by enzyme-linked immunoassay (Phical™ Fecal Calprotectin immunoassay, Genova Diagnostics, Inc, Asheville, NC) in order to assess gastrointestinal inflammation as previously described[19].

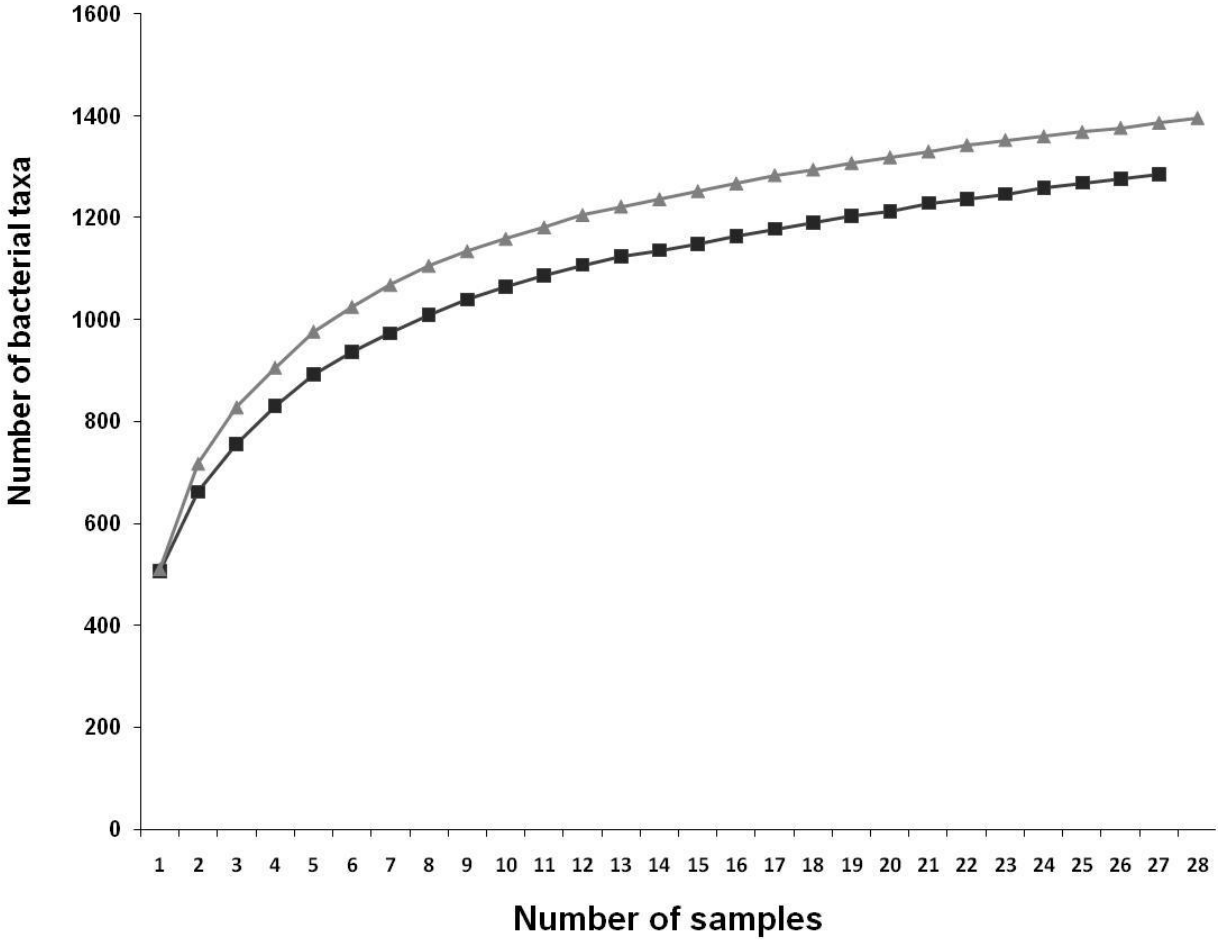
SUPPLEMENTARY FIGURES



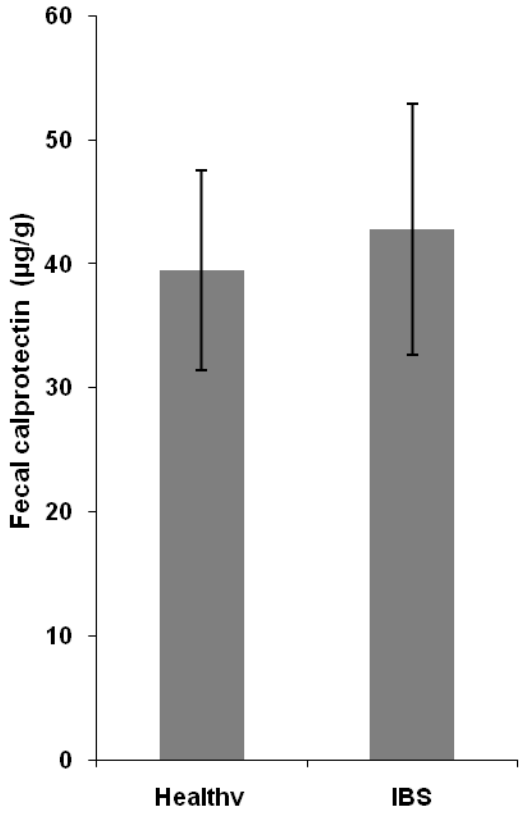
Supplementary Figure 1. Similar quantities of fecal bacteria are present in healthy children and children with IBS. No significant differences in total bacterial load (16S rRNA copy per g of feces) between healthy subjects (n=29 samples from 22 subjects) and IBS patients (n=42 samples from 22 patients) by real-time PCR using SYBR Green (T-Test, $P < 0.05$).



Supplementary Figure 2 Bacterial richness was not significantly different between the specimens from children with IBS (28 samples from 17 subjects) or specimens from healthy children (27 samples from 21 subjects) by G2 Phylochip hybridization (Median Richness: IBS=570 taxa, healthy=545 taxa).



Supplementary Figure 3 Greater OTU diversity in bacterial taxa identified using the G2 Phylochip is observed in the IBS group (28 samples from 17 subjects) compared to the healthy group (27 samples from 21 subjects). ◆ Samples from IBS subjects; ● Samples from healthy children. OTUs were defined by having 97% identity. Curves were generated using the EstimateS program (Chao 1).



Supplementary Figure 4. Children with IBS exhibit no evidence of increased intestinal inflammation. Fecal calprotectin concentrations (average ± standard error of mean) are similar between healthy children (n=19 samples) and children with IBS (n=19 samples) from the first collection. An outlier in the healthy group (fecal calprotectin =685 µg/g) was removed from the analysis.

Supplementary Table 1A Taxa significantly present in higher abundance (% total bacteria) in IBS samples (N=28) compared to healthy samples (N=27) on the G2 Phylochip. Unpaired T-Test (*P*<.05)

OTU STRING_SUBFAMILY	Fold-change IBS/Healthy	p-value
984_Nitrospiraceae	19.6	0.003
7893_Coxiellaceae	19.0	0.003
4265_Clostridiaceae	14.2	0.019
17_Peptostreptococcaceae	14.0	0.032
2339_Unclassified	13.5	0.013
9800_Desulfobacteraceae	12.7	0.028
9105_Unclassified	12.5	0.027
6259_Prevotellaceae	12.3	0.014
8245_Unclassified	12.1	0.013
8554_Enterobacteriaceae	11.8	0.032
Escherichia_coliG2_st	10.6	0.050
8741_Thiotrichaceae	9.0	0.024
10489_Unclassified	8.8	0.036
1450_Cellulomonadaceae	8.7	0.014
9628_Pasteurellaceae	8.6	0.011
3869_Streptococcaceae	8.2	0.026
9135_Enterobacteriaceae	7.8	0.016
5946_Prevotellaceae	7.1	0.047
8876_Pasteurellaceae	6.9	0.028
3028_Lachnospiraceae	6.6	0.039
2941_Lachnospiraceae	6.5	0.012
8952_Pasteurellaceae	6.4	0.012
10298_Polyangiaceae	6.4	0.012
9288_Alteromonadaceae	6.3	0.012
9661_Syntrophobacteraceae	5.8	0.032
8532_Unclassified	5.7	0.012
8861_Pasteurellaceae	5.3	0.050
8228_Pasteurellaceae	5.1	0.046
8174_Alteromonadaceae	4.9	0.024
6414_Acidobacteriaceae	4.9	0.024
9218_Alteromonadaceae	4.8	0.024
1671_Promicromonosporaceae	4.7	0.024
9282_Unclassified	4.5	0.024

6426_Unclassified	3.6	0.046
9205_Alteromonadaceae	3.5	0.046
9640_Alteromonadaceae	3.5	0.046
9143_Alteromonadaceae	3.5	0.046
8863_Alteromonadaceae	3.5	0.046
10189_Desulfovibrionaceae	3.4	0.045
8336_Alteromonadaceae	3.4	0.046
3579_Bacillaceae	3.4	0.046
59_Peptococc/Acidaminococc	3.4	0.046
9294_Aeromonadaceae	3.4	0.046
1066_Eubacteriaceae	2.7	0.046

Supplementary Table 1B Taxa significantly present in lower abundance (% total bacteria) in IBS samples (n=28) compared to healthy samples (n=27) on the G2 Phylochip. Unpaired T-Test (*P<.05*)

OTU STRING_SUBFAMILY	**Fold-change IBS/Healthy**	**p-value**
6168_Unclassified Bacteroidetes	-31.3	0.006
5264_Bacteroidaceae	-26.7	0.001
5489_Bacteroidaceae	-23.6	0.008
Bacteroides_vulgatus	-16.7	0.021
5192_Chloroplasts	-16.0	0.005
2893_Lachnospiraceae	-8.8	0.055
2364_Unclassified	-8.4	0.025
5510_Porphyromonadaceae	-8.0	0.010
3685_Streptococcaceae	-7.2	0.050
3630_Paenibacillaceae	-5.9	0.051
1923_Cellulomonadaceae	-5.4	0.034
1217_Unclassified	-4.1	0.031

Supplementary Table 2 OTUs associated with IBS-C when comparing IBS samples as described by randomForest Analysis[•].

OTU# 454	G2 Phylochip_chip tax string	DNAML id*	U[§]	Z°	Directional change (IBS-C) [¶]
5428	Bacteroidetes; Bacteroidetes; Bacteroidales; Bacteroidaceae; sf_12; otu_6309	0.95	144	-4.45	DOWN
6692	Bacteroidetes; Bacteroidetes; Bacteroidales; Bacteroidaceae; sf_12; otu_6309	0.96	158	-4.73	DOWN
2264	Bacteroidetes; Bacteroidetes; Bacteroidales; Bacteroidaceae; sf_12; otu_6064	0.95	165	-4.12	UP
4403	Bacteroidetes; Bacteroidetes; Bacteroidales; Bacteroidaceae; sf_12; otu_6309	0.95	166	-4.17	DOWN
6254	Firmicutes; Mollicutes; Anaeroplasmatales; Erysipelotrichaceae; sf_3; otu_4076	0.90	170	-4.06	DOWN
2785	Bacteroidetes; Bacteroidetes; Bacteroidales; Unclassified; sf_15; otu_5472	0.92	180	-3.98	UP
8493	Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; sf_5; otu_4255	0.97	180	-3.91	DOWN
1865	Firmicutes; Clostridia; Clostridiales; Peptococc/Acidaminococc; sf_11; otu_59	0.97	186.5	-3.97	UP
4457	Bacteroidetes; Bacteroidetes; Bacteroidales; Bacteroidaceae; sf_12; otu_5424	0.91	188	-3.79	UP
5157	Bacteroidetes; Bacteroidetes; Bacteroidales; Bacteroidaceae; sf_12; otu_5424	0.90	188.5	-3.80	UP
3086	Firmicutes; Clostridia; Clostridiales; Peptococc/Acidaminococc; sf_11; otu_59	0.96	194	-3.81	UP
5859	Bacteroidetes; Bacteroidetes; Bacteroidales; Bacteroidaceae; sf_12; otu_6309	0.95	200	-3.66	DOWN
2580	Bacteroidetes; Bacteroidetes; Bacteroidales; Bacteroidaceae; sf_12; otu_5329	0.96	206.5	-3.55	UP
8195	Firmicutes; Clostridia; Clostridiales; Clostridiaceae; sf_12; otu_4570	0.95	207	-3.67	UP

2739	Bacteroidetes; Bacteroidetes; Bacteroidales; Bacteroidaceae; sf_12; otu_5329	0.96	211	-3.49	UP
3990	Bacteroidetes; Flavobacteria; Flavobacteriales; Flavobacteriaceae; sf_1; otu_5401i	0.86	212.5	-3.85	UP
7615	Actinobacteria; Actinobacteria; Coriobacteriales; Coriobacteriaceae; sf_1; otu_1790	0.94	213	-3.52	DOWN
8276	Firmicutes; Clostridia; Clostridiales; Peptococc/Acidaminococc; sf_11; otu_181	0.96	213	-3.46	UP
3479	Bacteroidetes; Bacteroidetes; Bacteroidales; Bacteroidaceae; sf_12; otu_5329	0.96	215	-3.40	UP
6613	Firmicutes; gut clone group; Unclassified; Unclassified; sf_1; otu_4616	0.90	215.5	-3.40	DOWN
3381	Bacteroidetes; Bacteroidetes; Bacteroidales; Prevotellaceae; sf_1; otu_5905	0.99	217	-3.96	DOWN
6271	Firmicutes; Clostridia; Clostridiales; Clostridiaceae; sf_12; otu_4198	0.95	229	-3.58	UP
6327	Firmicutes; Clostridia; Clostridiales; Clostridiaceae; sf_12; otu_4593	0.90	230	-3.35	UP
2032	Bacteroidetes; Bacteroidetes; Bacteroidales; Bacteroidaceae; sf_12; otu_6064	0.96	232.5	-3.16	UP
2570	Bacteroidetes; Bacteroidetes; Bacteroidales; Prevotellaceae; sf_1; otu_5905	0.99	233	-3.69	DOWN
3458	Bacteroidetes; Bacteroidetes; Bacteroidales; Bacteroidaceae; sf_12; otu_5329	0.96	235	-3.12	UP
7763	Firmicutes; Mollicutes; Anaeroplasmatales; Erysipelotrichaceae; sf_3; otu_4076	0.93	238	-3.12	UP
4835	Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; sf_5; otu_4276	0.88	254.5	-3.09	DOWN
7361	Firmicutes; Clostridia; Clostridiales; Peptococc/Acidaminococc; sf_11; otu_167	0.91	261	-4.00	DOWN
4306	Bacteroidetes; Bacteroidetes; Bacteroidales; Unclassified; sf_15; otu_5890	0.82	261	-2.74	UP
5925	Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; sf_5; otu_2777	0.96	273	-2.66	DOWN

5564	Firmicutes; gut clone group; Unclassified; Unclassified; sf_1; otu_4616	0.94	277.5	-2.59	DOWN
4371	Firmicutes; Mollicutes; Anaeroplasmatales; Erysipelotrichaceae; sf_3; otu_4076	0.91	284	-2.55	DOWN
1318	Bacteroidetes; Bacteroidetes; Bacteroidales; Prevotellaceae; sf_1; otu_5949	0.89	287	-4.09	DOWN
3289	Bacteroidetes; Bacteroidetes; Bacteroidales; Rikenellaceae; sf_5; otu_5452	0.82	287	-4.09	DOWN
8708	Firmicutes; Clostridia; Clostridiales; Peptococc/Acidaminococc; sf_11; otu_534	0.89	287	-4.09	DOWN
2027	Firmicutes; Mollicutes; Anaeroplasmatales; Erysipelotrichaceae; sf_3; otu_4007	0.93	287	-4.09	DOWN
4083	Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; sf_5; otu_2777	0.93	287.5	-2.36	DOWN
2555	Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; sf_5; otu_4344	0.82	294.5	-3.71	DOWN
5191	Bacteroidetes; Bacteroidetes; Bacteroidales; Bacteroidaceae; sf_12; otu_5424	0.85	303	-2.14	UP
6754	Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; sf_5; otu_4571	0.99	312	-2.43	DOWN
2396	Bacteroidetes; Bacteroidetes; Bacteroidales; Bacteroidaceae; sf_12; otu_5885	0.97	312	-2.07	DOWN
5075	Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; sf_5; otu_2756	0.99	326	-1.80	DOWN
2465	Bacteroidetes; Bacteroidetes; Bacteroidales; Porphyromonadaceae; sf_1; otu_5664	0.86	327	-1.79	UP
7896	Firmicutes; Clostridia; Clostridiales; Clostridiaceae; sf_12; otu_4570	0.97	344	-1.59	DOWN
6755	Firmicutes; Clostridia; Clostridiales; Clostridiaceae; sf_12; otu_4198	0.96	353.5	-1.41	DOWN
7714	Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; sf_5; otu_4571	0.99	355	-1.38	DOWN
6458	Bacteroidetes; Flavobacteria; Flavobacteriales; Flavobacteriaceae; sf_1; otu_6248	0.85	360	-1.48	UP

3757	Bacteroidetes; Bacteroidetes; Bacteroidales; Bacteroidaceae; sf_12; otu_5980	0.95	381.5	-1.04	DOWN
7165	Bacteroidetes; Flavobacteria; Flavobacteriales; Flavobacteriaceae; sf_3; otu_6034	0.87	386.5	-0.94	DOWN
7980	Bacteroidetes; Flavobacteria; Flavobacteriales; Flavobacteriaceae; sf_3; otu_6034	0.88	393	-0.84	DOWN
4922	Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; sf_5; otu_4276	0.89	397.5	-0.77	DOWN
3440	Bacteroidetes; Bacteroidetes; Bacteroidales; Bacteroidaceae; sf_12; otu_5980	0.95	418	-0.48	DOWN
2781	Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; sf_1; otu_8033	0.88	424	-0.40	DOWN

•Classification of the different IBS subtypes based on these OTUs had a success rate of 98.5 % for IBS-C and were confirmed by using the Boruta algorithm. Discriminative taxa are ranked by decreasing importance for the classification. These sequences were compared to OTU 454 pyrosequencing data to generate this table using Greengenes (http://greengenes.lbl.gov/cgi-bin/JD_Tutorial/nph-Classification_Excel.cgi)

*DNAML id : The percent identity between the fasta sequence of the OTU clone (454 data) and the reference considering only conserved bases of the G2 PhyloChip taxonomy.

§U: Boruta Mann-Whitney total rank score;

°Z: Boruta Mann-Whitney z-score values.

¶UP: present in greater abundance in IBS-C, DOWN: present in lower abundance in IBS-C.

Supplementary Table 3: Comparisons of Percent Relative Abundance of Bacterial Genera between Healthy Children and Children with IBS.

Percent Relative Abundance * Genus Level				
	V1V3		V3V5	
Genus	Healthy	IBS	Healthy	IBS
Abiotrophia	0.00	0.02	0.00	0.01
Acetanaerobacterium	0.00	0.00	0.00	0.00
Acetivibrio	0.00	0.00	0.22	0.30
Acidaminococcus	0.01	0.00	0.03	0.06
Actinomycineae	0.01	0.08	0.03	0.02
Aggregatibacter	0.00	0.00	0.00	0.02
Akkermansia	4.82	5.87	2.83	1.70
Alistipes	10.16	10.65	10.12	8.84
Allisonella	0.01	0.02	0.00	0.00
Anaerostipes	0.00	0.00	0.52	0.30
Anaerotruncus	0.26	0.36	0.02	0.02

Percent Relative Abundance * Genus Level				
	V1V3		V3V5	
Genus	Healthy	IBS	Healthy	IBS
Anaerovorax	0.22	0.26	0.09	0.06
Bacteroides	39.83	35.04	44.58	41.92
Barnesiella	1.42	0.67	1.05	0.62
Bifidobacteriaceae	0.30	0.30	2.83	0.89
Blautia	2.55	3.48	1.92	3.93
Brachyspira	0.03	0.00	0.00	0.00
Butyricicoccus	0.31	0.89	0.58	0.70
Butyricimonas	0.17	0.10	0.00	0.22
Catenibacterium	0.00	0.23	0.00	0.04
Clostridiaceae 1	0.25	0.37	0.04	0.49
Coprobacillus	0.76	0.72	0.45	0.72
Coprococcus	0.60	0.43	0.76	1.47
Coriobacterineae	0.65	0.97	0.37	0.72

Percent Relative Abundance * Genus Level				
	V1V3		V3V5	
Genus	Healthy	IBS	Healthy	IBS
Desulfovibrio	0.01	0.02	0.00	0.00
Dialister	9.83	10.42	12.99	6.50
Dorea	0.86	0.89	0.87	1.79
Escherichia/Shigella	0.02	0.89	0.03	0.09
Ethanoligenens	0.00	0.00	0.03	0.01
Eubacterium	0.84	0.94	0.78	0.30
Faecalibacterium	5.02	3.97	2.15	4.62
Fusobacterium	0.00	0.00	0.00	0.01
Gemella	0.01	0.00	0.00	0.02
Granulicatella	0.00	0.06	0.00	0.00
Haemophilus	0.02	0.02	0.00	0.72
Holdemania	0.03	0.00	0.06	0.06
Howardella	0.01	0.02	0.02	0.03

Percent Relative Abundance * Genus Level				
	V1V3		V3V5	
Genus	Healthy	IBS	Healthy	IBS
Klebsiella	0.01	0.00	0.00	0.02
Lactobacillus	0.29	0.00	0.00	0.16
Lactococcus	0.00	0.03	0.02	0.03
Lawsonia	0.00	0.01	0.00	0.00
Megamonas	0.00	0.00	0.00	0.00
Megasphaera	0.04	0.51	0.01	0.11
Methanobrevibacter	0.00	0.00	0.00	0.02
Mogibacterium	0.01	0.01	0.00	0.00
Odoribacter	1.01	1.06	0.79	1.07
Oscillibacter	4.11	5.24	1.92	1.97
Oxalobacter	0.01	0.04	0.00	0.00
Papillibacter	0.00	0.00	0.18	0.10
Parabacteroides	3.56	4.76	3.07	4.17

Percent Relative Abundance * Genus Level				
	V1V3		V3V5	
Genus	Healthy	IBS	Healthy	IBS
Paraprevotella	0.06	0.05	0.00	0.50
Parasutterella	0.58	0.88	0.40	0.51
Peptococcaceae 1	0.00	0.00	0.00	0.04
Phascolarctobacterium	0.25	0.70	0.14	0.50
Porphyromonas	0.01	0.02	0.00	0.00
Prevotella	3.35	1.65	0.21	2.57
Pseudobutyrivibrio	0.02	0.02	0.00	0.00
Pyramidobacter	0.02	0.00	0.00	0.00
Rikenella	0.00	0.00	0.00	0.03
Roseburia	0.92	1.16	3.36	3.43
Ruminococcus	1.72	0.82	0.70	1.66
Sporacetigenium	0.44	0.19	0.19	0.61
Sporobacter	0.00	0.00	0.12	0.20

Percent Relative Abundance * Genus Level				
	V1V3		V3V5	
Genus	Healthy	IBS	Healthy	IBS
Streptococcus	0.45	0.32	0.74	0.39
Subdoligranulum	3.24	3.45	4.06	3.29
Succiniclasticum	0.02	0.40	0.00	0.14
Sutterella	0.20	0.37	0.01	0.13
Tannerella	0.00	0.00	0.00	0.00
Turicibacter	0.10	0.03	0.18	0.19
Veillonella	0.11	0.09	0.00	0.37
* RDP classified genus level data relative percent abundance data filtered at >0.01%				
Sample counts V1V3/V3V5 (n): Healthy Children (69/68) Children with IBS (71/70)				

SUPPLEMENTARY REFERENCES

1. 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:D141-5.
2. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010;7:335-336.
3. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 2006;22:1658-9.
4. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 2006;72:5069-72.
5. DeSantis TZ, Jr., Hugenholtz P, Keller K, Brodie EL, Larsen N, Piceno YM, Phan R, Andersen GL. NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. Nucleic Acids Res 2006;34:W394-9.
6. Lane DJ. 16S/23S rRNA sequencing. In: Goodfellow ESM, ed. Nucleic Acid Techniques in Bacterial Systematic. New York: Wiley, 1991:115-175.
7. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 2009;26:1641-50.
8. Jukes TH, Cantor CR. Evolution of protein molecules. In: Munro HN, ed. Mammalian protein metabolism. New York: Academic Press, 1969:21-123.
9. Kindt R, Coe R. Tree diversity analysis: a manual and software for common statistical methods for ecological and biological studies, 2005.
10. Team RDC. A language and environment for statistical computing. In: Computing RFfS, ed. Vienna, 2008.
11. Chao A. Non-parametric estimation of the number of classes in a population. Scand J Statist 1984;11:265-70.
12. Shannon CE, Weaver WE. The Mathematical Theory of Communication. In: Press I, ed. Urbana, 1963.
13. Rheny A. On measure of information and entropy, In Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability, Berkeley, 1961.
14. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 2005;71:8228-35.
15. Manly B. Multivariate statistical method: A primer, 1986.
16. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. Journal of statistical software 2010;36:1-13.
17. Knights D, Costello EK, Knight R. Supervised classification of human microbiota. FEMS Microbiol Rev 2010.
18. Einen J, Thorseth IH, Ovreas L. Enumeration of Archaea and Bacteria in seafloor basalt using real-time quantitative PCR and fluorescence microscopy. FEMS Microbiol Lett 2008;282:182-7.
19. Carroccio A, Iacono G, Cottone M, Di Prima L, Cartabellotta F, Cavataio F, Scalici C, Montalto G, Di Fede G, Rini G, Notarbartolo A, Averna MR. Diagnostic accuracy of fecal calprotectin assay in distinguishing organic causes of chronic diarrhea from irritable bowel syndrome: a prospective study in adults and children. Clin Chem 2003;49:861-7.