Supporting Information

Hartman, et al. 10.1073/pnas.0904847106

SI Text

DNA Extraction. Effluent samples were transported on dry ice and stored at −80 °C until processed. A 400-μL sample of ileal effluent was combined with 400 μL of phenol and 300 μL of water. Then, 1 mL of 0.1-mm Glass beads (100g, Bio-Spec Products) were added and the sample blended in a Bead-Beater (Bio-Spec Products) for 3 min. The resulting slurry was transferred to a 2 mL of Phase Lock Gel Light tube (Eppendorf) and spun for 10 min at 13,500 × *g*. The supernatant was transferred to a new Phase Lock Gel Light tube, mixed with 700 μL of phenol-chloroform, briefly vortexed, and centrifuged for 10 min at 13,500 × *g*. The supernatant was precipitated in 2.5 vol of 100% ethanol and 0.1 vol of 3 M NaOAc. The precipitate was washed in 70% ethanol.

Sequencing of 16S rDNA. Ileal effluent samples were obtained ≈70 days posttransplant from each of nine patients randomly selected from the cohort of 17. DNA was extracted as described above. For each sample, separate PCR procedures were performed at 48 °C, 49 °C, and 50 °C and the products pooled. Full length bacterial 16S rDNA sequences were obtained using the universal reverse primer 1391R (5′-GACGGGCRGTGWGTRCA-3′) and the Bacteria domain-specific forward primer 27F (5′-AGAGTTTGATCCTGGCTCAG-3′). Each 25-μL reaction mixture contained 2.5 μL of 10× Takara ExTaq buffer, 1.0 μL of 5M Betaine, 1.0 μL of 10 μM 27F primer, 1.0 μL of 10 μM 1391R primer, 17.5 μL of water, and 2 μL of a 1:50 dilution of extracted DNA in DNase/RNase-free water. Thermocycler conditions (BioRad Alpha Unit Block on Dyad Disciple Peltier Thermal Cycler) were 94 °C for 2 min followed by 30 repeats of the following steps: 94 °C for 30 sec, a temperature gradient of 48–50 °C for 30 sec, and 72 °C for 1 min. Size of the resulting PCR products was confirmed by agarose gel electrophoresis and subsequently extracted from the gel using the Freeze 'n Squeeze DNA Gel Extraction Kit (Bio-Rad). The PCR product was then cloned using Invitrogen's TOPO PCR3 kit and sequenced at the Department of Energy's Joint Genome Institute (JGI). Using JGI's 16S GeneLib pipeline, 2993 high quality, full-length assembled sequences were selected; after alignment and chimera screens, 1892 sequences remained and were deposited in GenBank under Accession numbers FJ975781-FJ976038 for the OTU dataset and GQ154686-GQ156577 for all of the 1892 sequences. Taxonomic analysis was performed using the STAP package (1).

qPCR. ***Primer choice and optimization.*** First, several SYBR green master mixes were compared to test their amplification efficiency and reliability for these samples. It was determined that Takara's perfect real-time 2x SYBR Green master mix (Fisher Scientific) performed the best. The taxonomic groups quantified here were selected both on the taxonomic group's presence in the clone libraries and on previously known biology. Bacteroidales and Clostridiales were chosen because they had been shown to be the dominant organisms along the length of the GI tract (2). Enterobacteriales was chosen because this group has been shown to bloom in response to host inflammation (3). After sequencing results were examined (described below), Lactobacillales were also surveyed with qPCR. The panbacteria (total bacteria) qPCR primer set was chosen because it was the broadest (amplified the most bacteria) primer set available within our qPCR guidelines (below).

From extensive literature searches group-specific primers were then chosen (4, 5, 15) and confirmed for specificity and validity of these primers using the RDP's Probe Match website (6). During these searches, all potential primers were chosen based on their breadth, i.e., the percentage of the desired clade covered by the primer set and the primer specificity, i.e., whether the primers amplified nontarget groups. The Lactobacillales primers were designed specifically for this study based on the Lactobacillales sequences in the clone library because none of the existing primers were sufficiently broad. They were then verified using ProbeMatch. We also verified that the top two most numerically dominant ileal Clostridiales members according to Ahmed et al. (2), *Clostridium clostridioforme* and *Eubacterium rectale* should be amplifiable with our Clostridiales primers.

Other primer considerations were whether the predicted amplicon size was ≈150–300 bp and, we also attempted to minimize the degeneracy required within the oligomer. The primers that were ultimately chosen (Table S2) were the "best fit" of all these criteria.

As an additional technical consideration to increase qPCR efficiency, primers were ordered with HPLC purification at a 50 nM scale from Invitrogen. Primer pairs were then optimized for annealing temperature along an 8-degree temperature gradient flanking the predicted optimal. Note that annealing temperatures (Table S2) were optimized using Takara's ExTaq perfect real time 2x master mix (Fisher Scientific) and annealing temperatures might not be optimal for every other master mix.

Bacterial standard curves. For each primer set, a reference organism was chosen to create a 6-log-fold standard curve for direct quantification of all samples. All gDNA preparations used for standard curves were purchased from the America Type and Tissue Collection (ATCC). For total bacteria and Enterobacteriales assays, *Escherichia coli* ATCC 700926D-5 gDNA was used. For Bacteroidales assays, *Bacteroides fragilis* ATCC 25285 was used. For Clostridiales assays, *Ruminococcus productus* ATCC 27340 was used. For Lactobacillales assays, *Lactobacillus delbreuckii* ATCC 11842 was used. With the exception of *R. productus*, all standard curve organisms have a genome sequence available, and, therefore, the number of *rrn* operons and the size of the genome was known and used. To associate a $C(t)$ value (fluorescence threshold) with 16S rDNA copies, the reference organisms' genome size and 16S rDNA copy, and DNA concentration at the first serial dilution were used in the following formula to calculated copies per microliter (Table S2): (((((16S rDNA copies in the genome)/(Genome size))*(DNA concentration))/(molecular weight of DNA))*(Avogadro's number))/(ng/g conversion).

Finally, possible cross-reactivity (a primer set amplifying nontarget organisms) of each primer set against all standard curve organisms was done to confirm that, at least for these organisms, that there was no unexpected cross-reactivity of these primers.

Nod2 Genotyping. We took advantage of published work, in which we had Nod2 genotyped a large amount of patients in the SBT study at GUMC for other research purposes according to methods described in ref. 7. An average proportion of Nod2 wild type and mutant individuals (relative to all SBTs performed at GUMC) was included in this study, 12 WT: 5 Mutant. The patient's Nod2 status was then used in the statistical tests described below.

Statistical Analysis. All statistical analysis and graphing were done in the R 8.0 statistical software package. To compare the qPCR and sequencing methods to each other, the statistical package epi.ccc() was used. To robustly analyze the transplant effluent dataset (Fig. 1), the significant inter-patient differences in sampling and time scale first had to be normalized. To do this, time bins were created to cover sampling differences between all individuals for the first 125 days (Fig. 1 and Fig. S2) (16). This period was frequently sampled as well as a clinically critical period for graft healing and possible rejection or other complications. This time-binning method permitted a uniform comparison of these time series from one subject to another.

Briefly, time bins were made that allowed each individual to have at least one sample but no more than three in any given bin. The bins became increasingly large at later time sections because the frequency of sampling decreased with time posttransplant. Eight time bins were created that represent days 0–125 posttransplant. If there was more than one sample in a given time bin, those bacterial quantities were averaged together using geometric means. Time bin selection details are displayed in Fig. S2*A*.

The time bin delineation created 2 matrices: an 8×17 matrix for total bacterial quantities and an 8×68 (17 individuals for 4 bacterial assays for 8 time bins) on which all further statistical tests were done. Note that because of the interdependence of the 4 subgroup bacterial assays on the total bacteria qPCR assay, these 2 matrices were considered and analyzed independently. However, all following statistical tests were performed on both datasets.

To test for differences and similarities in these time series, first, principal component analysis was done using the pricom() function in R. Next, hierarchical clustering (using Euclidean, maximum, and minimum distance calculations) was done using the hclust() function. K-means clustering was then done using the kmeans() and clustplot() functions. Also, model-based clustering was done using the Mclust() function. Then, a more sophisticated machine-learning approach was attempted, using self-organizing-map clustering in the VisualGene software (version 1.01.0024; Visipoint, Kuopio, Finland), commonly used to group similarly expressed gene profiles. Linear and quadratic discriminant analyses were also performed with no supported separation along any of the clinical variables and implemented with the MASS package's lda() and qda(). Finally, population variability was determined according to Heath (8) and reprogrammed from matlab (code from the original publication) to R.

Metabolomics Network Graphs. Using PubChem, a pair wise matrix of tanimoto chemical similarity coefficients was calculated for molfile-encoded chemical structures. These structures were associated with compound identifiers of 139 known metabolites. Next, a pair wise matrix was calculated within Binbase between all known metabolites based on chemical similarities. This matrix was clustered using the hierarchical clustering algorithm in the TMEV software package (http://www.tm4.org/mev.html). The matrix was constrained to a high similarity threshold (0.7) using R and Microsoft Excel. Metabolites that did not pass the threshold were connected to the nearest possible neighbor in the resulting hierarchical tree and were displayed with a dashed line. The network was then formatted into Cytoscape's SIF format. Differential statistical results (P value <0.1) and dataset properties (e.g., preclosure and postclosure) were mapped onto different visual properties: node color = *t* test statistics, edge style = cutoff score. The final network was visualized using Cytoscape's organic layout (9).

SI Results

Attempts to Cluster and Statistically Separate the SBT Cohort Failed. We asked whether some individuals were more similar to each other than to the rest of the cohort or whether or not any recorded clinical variable associated with any observed clustering pattern. The founding hypothesis of this study was that a Nod2 mutant genotype should affect the bacterial populations of these individuals. Beyond Nod2 genotypes, it seemed intuitive that perhaps survival, graft decontamination and debridement, or pretransplant pathology could also affect these 17 individuals in a unique manner. These variables were analyzed and visually inspected as the clustering methods were applied.

Principal component analysis was performed to test whether the total variation in the sample would cluster some individuals together and whether or not these clusters reflected any known clinical variable (Fig. S3). Although, the first three principal components captured 55% of the variation in these samples, no clear clustering pattern was seen, suggesting that no individual or group of individuals was uniquely different from any other. No difference between bacterial populations associated with Nod2 mutant (red) and wild type patients (black) was observed. After no clear patterns were found with PCA, other clustering methods were used: hierarchical clustering, k-means clustering, model-based clustering and self-organizing maps. These methods also yielded no significant findings.

We also examined the possibility that microbial population stability (how much and how frequently a population changes through time) could correlate with subject outcome. To quantify stability, population variability (Fig. S3*C*), an ecological metric of stability (deviation from the difference between the maximum and minimum population levels) within a time series, was calculated. No correlation was detected with patient survival, Nod2 genotype, ileostomy closure (a measure of good patient outcome), or decontamination. The lack of statistical support for any clinical-microbial pattern was surprising because the Nod2 mutant phenotype is particularly devastating in terms of subject and graft survival (7).

Visually and descriptively the inter-individual temporal variability within SBT patients seems higher than studies of other disruptive gut events such as after antibiotic usage (10, 11) or in patients with Crohn's disease (12). Still the bacterial population stability in the transplanted small bowel is probably greater or more consistent than that observed in human skin, which seems to be a very transient microbial community (13). The stability of the SBT bacterial populations seems most similar to patterns observed in a time series study of infants, which showed high within-subject variability in the first 200 days of life (14) that eventually seemed to stabilize as their gut ecosystems equilibrated.

Metabolomics Statistical Analyses. Using principal component analysis (Fig. S5) the 2nd and 3rd axes of variation separate these two types of samples, although it is clear that the degree of inter-individual variation is stronger than simply the pre-/postclosure axis. Using partial least squares (PLS) regression where the clustering is based on the fit to a given variable, in this case preclosure or postclosure, the separation of these samples is significantly stronger. The wide distribution of the postclosure samples in Fig. S5 perhaps echoes the wider difference in microbial profiles seen in these postclosure samples.

1. Wu D, Hartman A, Ward N, Elsen JA (2008) An automated phylogenetic tree-based small subunit rRNA taxonomy and alignment pipeline (STAP). PLoS ONE 3:e2566.
2. Ahmed S, et al. (2007) Mucosa-associated bacterial diversity in relation to human terminal ileum and colonic biopsy samples. *Appl Environ Microbiol* 73:7435–7442.
3. Lupp C, et al. (2007) Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of Enterobacteriaceae. *Cell Host Microbe* 2:204.
4. Castillo M, et al. (2006) Quantification of total bacteria, enterobacteria and lactobacilli populations in pig digesta by real-time PCR. *Vet Microbiol* 114:165–170.

5. Rinttila T, Kassinen A, Malinen E, Krogius L, Palva A (2004) Development of an extensive set of 16S rDNA-targeted primers for quantification of pathogenic and indigenous bacteria in faecal samples by real-time PCR. *J Appl Microbiol* 97:1166–1177.
6. Cole JR, et al. (2009) The Ribosomal Database Project: Improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* 37:D141–145.
7. Fishbein T, et al. (2008) NOD2-expressing bone marrow-derived cells appear to regulate epithelial innate immunity of the transplanted human small intestine. *Gut* 57:323–330.
8. Heath JP (2006) Quantifying temporal variability in population abundances. *Oikos* 115:573–581.
9. Shannon P, et al. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504.
10. Dethlefsen L, Huse S, Sogin ML, Relman DA (2008) The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol* 6:e280.
11. Jernberg C, Lofmark S, Edlund C, Jansson JK (2007) Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. *ISME J* 1:56–66.
12. Scanlan PD, Shanahan F, O'Mahony C, Marchesi JR (2006) Culture-independent analyses of temporal variation of the dominant fecal microbiota and targeted bacterial subgroups in Crohn's disease. *J Clin Microbiol* 44:3980–3988.
13. Grice EA, et al. (2009) Topographical and temporal diversity of the human skin microbiome. *Science* 324:1190–1192.
14. Palmer C, Bik EM, Digiulio DB, Relman DA, Brown PO (2007) Development of the Human Infant Intestinal Microbiota. *PLoS Biol* 5:e177.
15. Amann RI, et al. (1990) Combinatjion of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* 56(6):1919–1925.
16. Lin Ll (1989) A concordance correlation-coefficient to evaluate reproducibility. *Biometrics* 45:255–268.

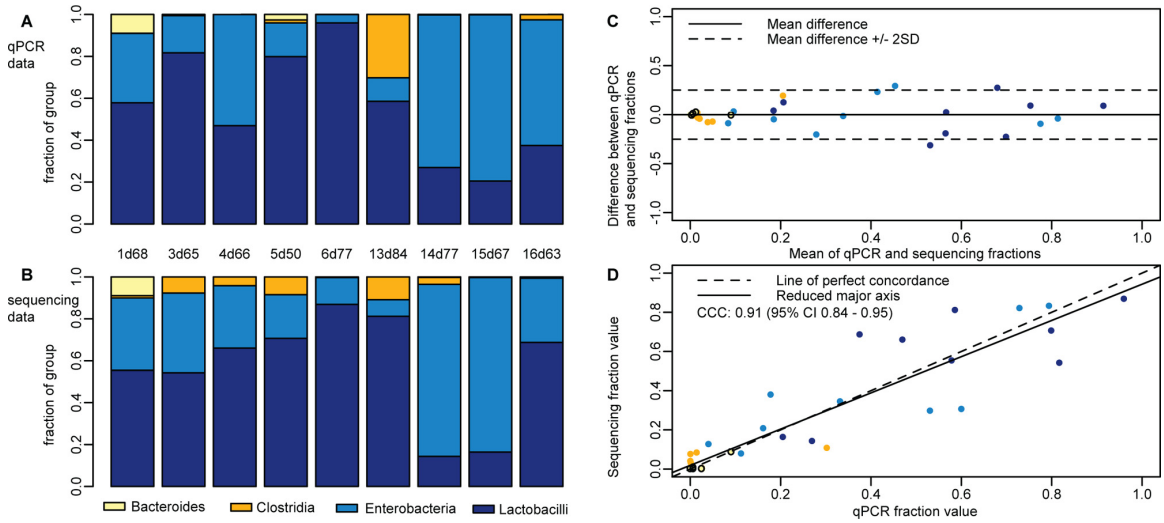

Fig. S1. High degree of concordance between qPCR and sequencing methods. The relative populations of four bacterial orders present in ileal effluent from 9 SBT patients sampled ≈70 days posttransplant. (*A*) qPCR data. (*B*) Full-length 16S rDNA sequence data. Bar labels: patient number followed by the number of days posttransplant. (*C*) The difference between the population fractions determined by the two methods is plotted against their mean. (*D*) Correlation between the qPCR and sequencing values determined by CCC. The dashed line illustrates the degree of concordance relative to the solid line of perfect method concordance; the correlation coefficient between the qPCR and sequencing methods had an R^2 value of 0.91.

A

Bin #	#days	days	Patients 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	8	0--8	0	0	5	0	0	6	6	0				0	0		0	0	
1	8	0--8	4				4		8	6	7	4	6	6	4	3			
	8	0--8	7			7	7							8	7				8
	8	9--17	11	15	9	10	10		12	14	10	10	9	14	12	10	13		12
2	8	9--17	15		16	14	17				14	13	13		15	17			
	8	9--17				17		17							18			18	
	11	18--29	19	21	23	21	21		19	19		24	19	22	22	24	23		
3	11	18--29	25	25					22	22			26		25			24	
	11	18--29	26	26			29	28	27		28			28	27		29	27	36
	11	30-40	32	34	31	30			30	33		31	33			31		31	
4	11	30-40				34	31		33		38	38		33	33	38	33		
	11	30-40	39		40		38	41					40	39			40		40
	16	41--56				48	42	61	43			41			42			35	
5	16	41--56	46	39	47				47		43	48			50	42	43	39	47
	16	41--56	56		51					48	46		50	56	56		54	46	54
6	16	57--72		46	65		50		57							52	61	63	
	16	57--72	68			66		77	64	61	59	61	57	60	64		67		69
	24	72--96							83							77			
7	24	72--96	81			73						69	75	74	71				
	24	72--96	85	95	95	93	77	91		92	85	73	89		84		90	80	110
	24	97--125							99					103					
8	24	97--125		104				97			101		106		102			112	
	24	97--125	116		115	125	154	103		117	115	153	114	116	112	224	180	118	187

B

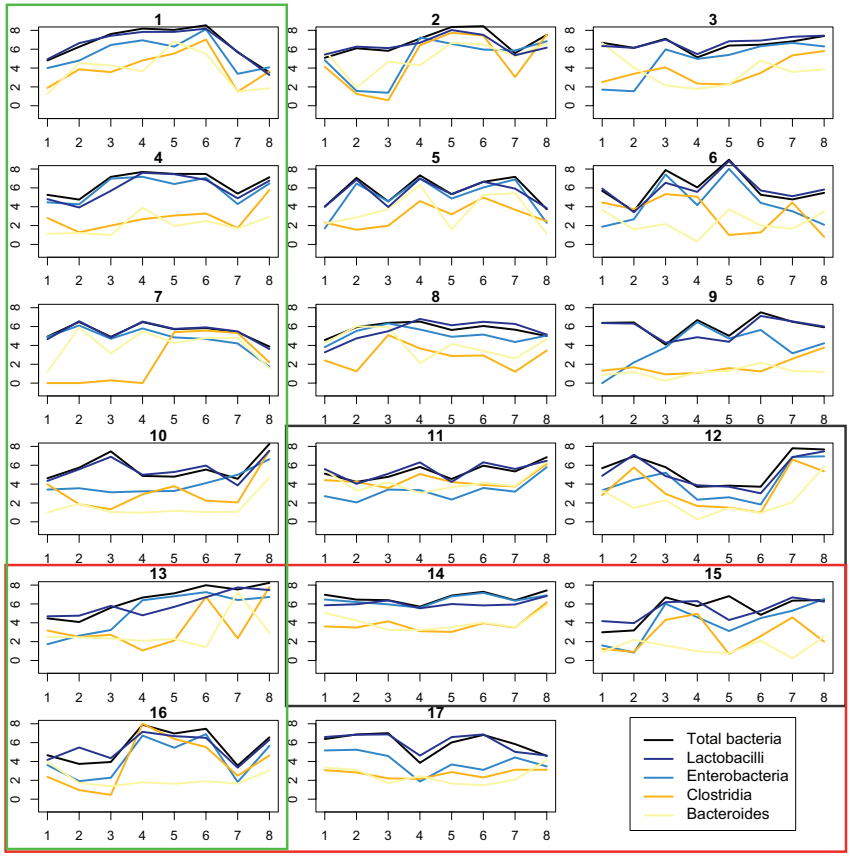

Fig. S2. Time-binning and graphs of samples in the day 0–125 time period. (*A*) Table of sample collection time points (in days) and their assignment to 8 time bins (each bin is a different color) within days 0–125. These time bins were used for all statistical analyses. A black box around a sample indicates that it is not a perfect fit into that time bin, but that no other sample was available. Multiple samples from one patient within a time bin were averaged together using geometric mean. (*B*) Graphical representation of the 17 patients' bacterial data after time-binning. In the graphs, the *y* axis is log scale. *x* axis represents the 8 time bins in chronological order (outlined in *A*). The layout of the patients 1–17 is the same as in Fig. 1 in the main body of the paper.

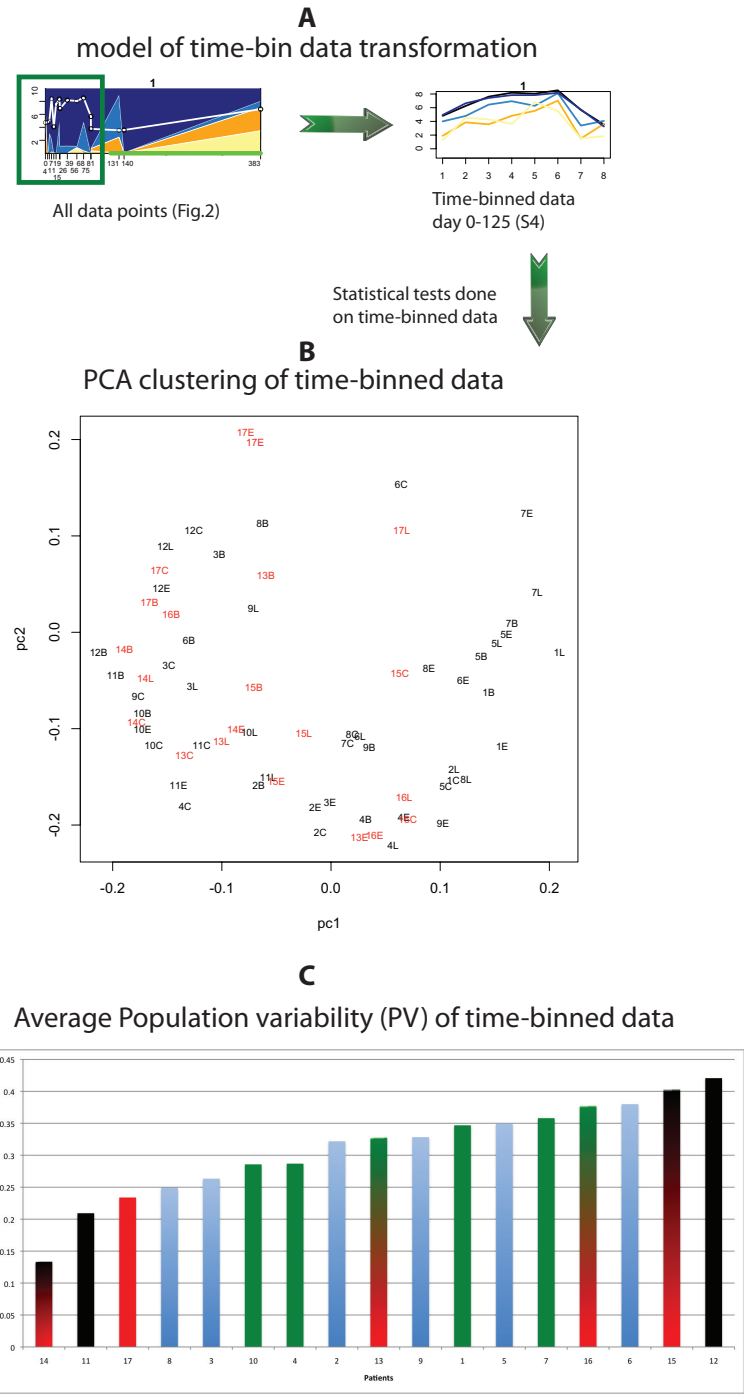

Fig. S3. Statistical analyses of the time-binned data reveal no association between clinical variables and bacterial populations. Because of the uneven sampling of the SBT cohort (Fig. 1), time points from the heavily sampled and clinically critical period of day 0–125 were transformed into 8 time bins for each patient such that every individual and each bin could be compared with all others for statistical purposes. In Panel A, an example from patient 1 is given. All subsequent statistical analyses were performed on time-binned data. Panel B shows principal component analysis for the time-binned data of the 4 bacterial subgroups surveyed and shows no correlation with Nod2 genotype or any other clinical variable. Panel C shows the mean time series population variability (PV) (a measure of bacterial population stability) of all five assays (Universal, Lactobacilli, Enterobacteria, Clostridia, Bacteroides). Black bars represent patients that died. Red bars represent Nod2 mutant patients. Green bars indicate postclosure patients. A gradient of two colors represents a patient had both variables. No correlation between clinical variables and PV was observed.

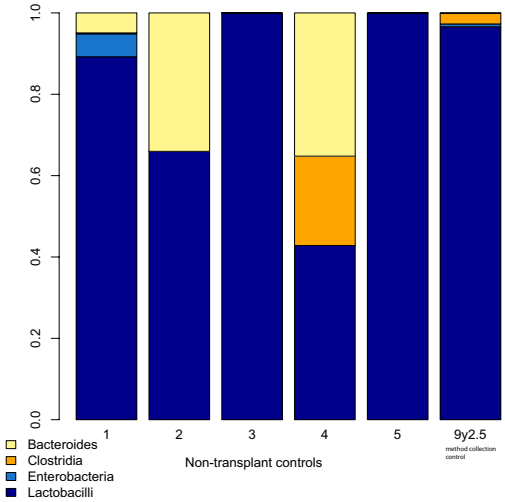

Fig. S4. Nontransplant ileostomy effluent is similar to transplant ileostomy effluent. Microbial populations in effluent taken from five individuals (1–5) with an ileostomy but without a small bowel transplant are similar to the ileostomy-open transplant samples shown in Fig. 1–2 and Fig. S2*B*. A control transplant sample was collected in a similar manner from SBT patient 9, 2.5 years after transplant (9y2.5).

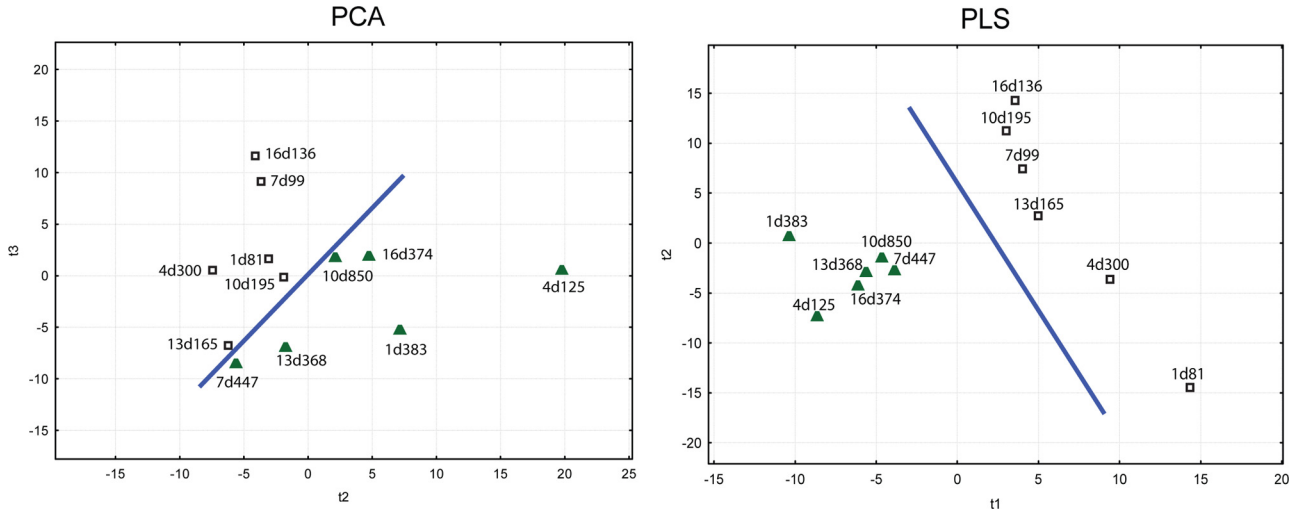

Fig. S5. Metabolomic statistical analyses: Metabolomic profiling suggests that there are significant metabolite differences between the six preclosure and six postileostomy-closure samples, suggesting that the metabolic repertoire of these two sets of samples is unique. In principal component analysis (an unsupervised method), preclosure and postclosure samples separate in the comparison of the 2nd and 3rd principal components, which account for 40% of the variation in the sample. In supervised PLS clustering, the separation is more distinct, although the after-closure samples show a large amount of variation. In both graphs, black squares are the last sample taken before ileostomy closure and green triangles are the last time point available after ileostomy closure. The naming of samples is the same as in Fig. S4: patient followed by ''d'' and the day of the sample relative to the transplant surgery, i.e., 16d136 is patient 16 taken on day 136 after transplant. The following pages are the box-and-whisker plots of the top 18 known metabolites associated with the preclosure and postclosure sample separation shown to be significantly different in an ANOVA analysis of metabolomic samples. These samples are the most significant contributors to the separation of the 12 samples seen in the PCA and PLS graph on the first page and are also found in Fig. 4. The *y* axis units for the box-and-whisker plots are peak height counts per spectrum for the quantification of ions of each individual compound.

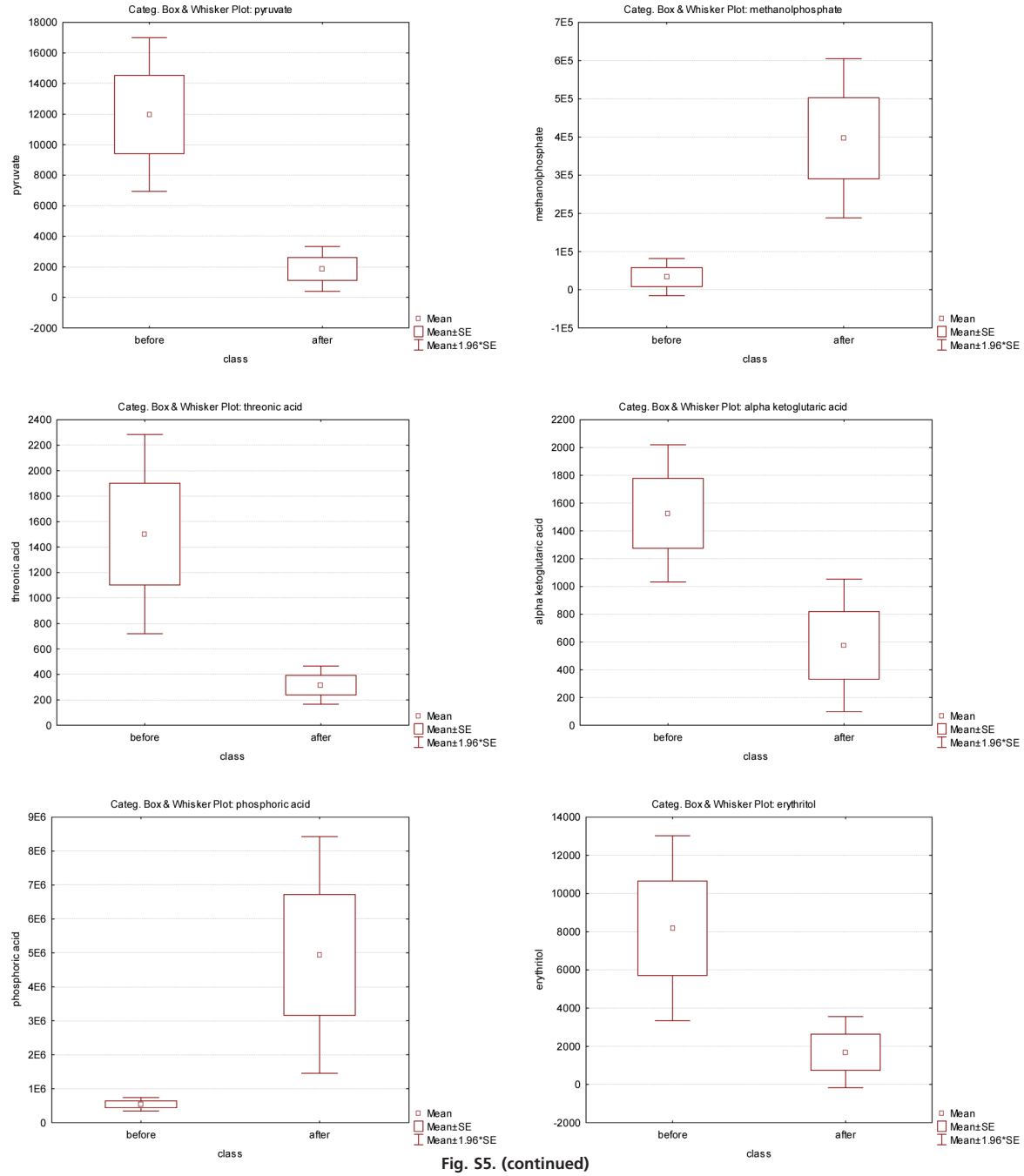

Fig. S5. (continued)

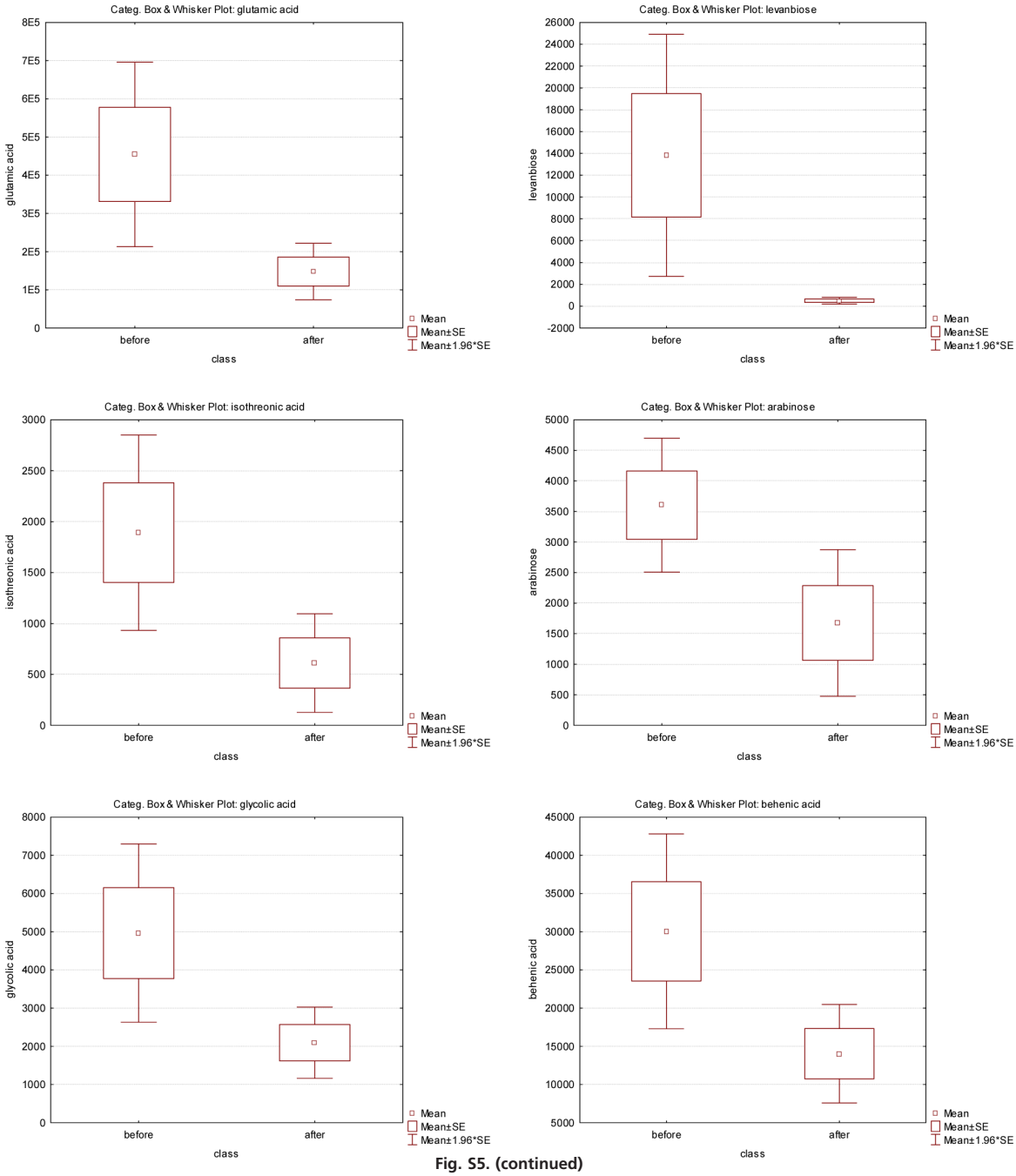

Fig. S5. (continued)

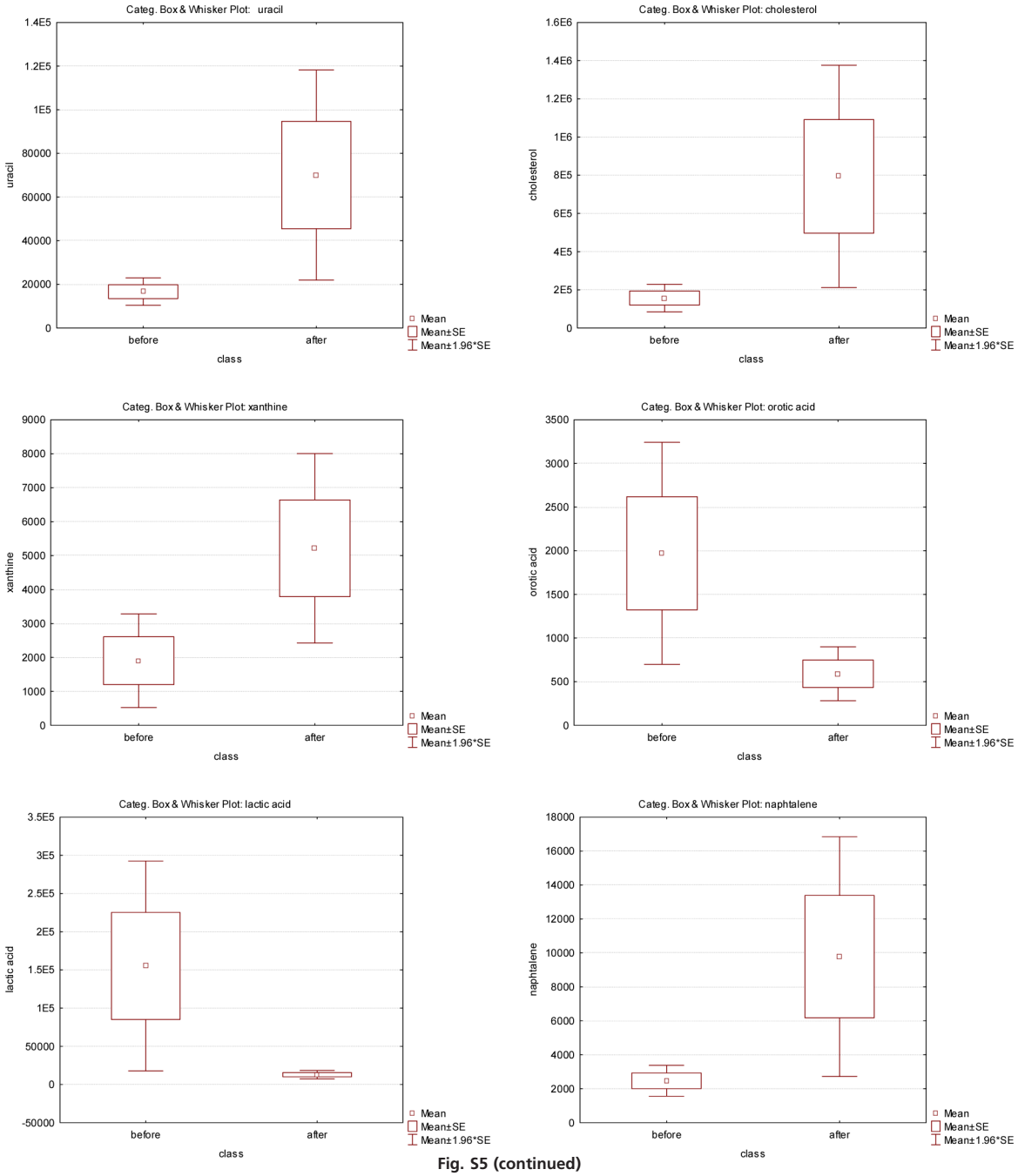

Fig. S5 (continued)

Table S1. Cohort of 17 adult SBT patients

Patient	Reason for transplant	Organs	Tx Date	Age	Status	*nod2*	Rej	PC	Dec	Ileo	N
1	Resection; motorcycle accident trauma	SB	3/30/07	39	Alive	WT	N	Y	N	S	16
2	Mass Resection	SB	8/25/07	41	Alive	WT	N	N	N	S	10
3	SGS	SB	7/29/06	56	Alive	WT	N	N	Y	S	13
4	SGS	SB	11/1/06	57	Alive	WT	N	Y	N	L	18
5	SGS	SB	2/27/07	49	Alive	WT	N	N	N	L	11
6	SGS; radiation enteritis	SB	9/2/06	64	Alive	WT	N	N	Y	S,L	11
7	Colon cancer; SB ischemic necrosis	SB	09/01/05	51	Alive	WT	Y	Y	Y	S	18
8	Gastric bypass hernia complication	SB,L,St	09/13/07	26	Alive	WT	Y	N	N	L	12
9	Desmoid Tumor; Gardners Syndrome	SB,L	04/18/06	32	Alive	WT	N	N	Y	E	16
10	Post viral dysmotility syndrome	SB,L,P,St	04/16/05	26	Alive	WT	N	Y	Y	S	15
11	Sporadic Colonic Pseudo-obstruction	SB*,L,P,St,C	06/17/06	45	Dead	WT	Y	N	Y	L	16
12	Portal vein thrombosis; SB malformation	SB,L,P	05/26/06	34	Dead	WT	N	N	Y	L	17
13	Malrotation, volvulus	SB*	2/23/07	41	Alive	1007-/-	N	Y	N	S	21
14	Crohn's Disease	SB	04/12/05	54	Dead	702+/-	Y	N	Y	S	17
15	Gardner's syndrome; Gut AdenoPolyps	SB*,L,P,K,St,C	08/25/05	43	Dead	908+/-	Y	N	Y	L	16
16	FBP; Pseudo-obstruction neuropathic	SB,P	11/4/06	33	Alive	1007-/-	N	Y	N	L	14
17	Crohn's Disease	SB*	08/04/05	42	Alive	702+/-	Y	N	Y	E	20

The 17 subjects enrolled in this study had a wide variety of clinical recoveries, complications and outcomes. Tx, transplant; SB, small bowel; *, re-transplant of SB (patient 11 2nd SBT at day 128, patient 15, all time points are from 2nd SBT, patient 17 had a 2nd SBT beyond the time points shown here), L, liver; S, spleen; K, kidney; St, stomach; P, pancreas; C, colon; SGS, short gut syndrome; Rej, rejection; N, number of samples collected; PC, postclosure sample; Dec, decontamination of allograft; Ileo, Ileostomy type; S, Santouli ileostomy, mixing from colon is possible; L, loop ileostomy, mixing from colon not possible; E, end (terminal) ileostomy, no colon present. Nod2 mutations: WT, wild type; 702, Arg-702-Trp, a point mutation; 908, Gly-908-Arg, a point mutation; 1007, Leu-1007-frameshift-Cys, a frameshift mutation that results in protein truncation (*SI Text*).

Table S2. qPCR primers, their purported specificity and sensitivity, and the reference genome used to infer absolute quantity of the 16S rDNA gene

Primer	Taxonomy	Forward primer	Reverse primer	Tm	Amplicon(bp)	16S Position	Coverage	Reference	Standard curve	copies in 2 uL
UniFR	All bacteria	ACTCCTACGGGAGGCAGCAGT	ATTACCGCGGCTGCTGGC	65.5	180	334-514	68/71	Amann	E. coli	43000000
EntFR	Enterobacteriales	ATGGCTGTCGTCAGCTCGT	CCTACTTCTTTTGCAACCCACTC	60.5	177	1475-1652	59/34	Castillo	E. coli	43000000
LactFR	Lactobacillales	GGAAACRRRWGCTAATACCG	GAAGATTCCCTACTGCT	59	189	158-347	41*/83	This study	L. delbrueckii	20800000
ClostFR	Clostridiales	CGGTACCTGACTAAGAAGC	AGTTTYATTCTTGCGAACG	60	429	477-906	34^33^	Rintilla	R. productus	44000000
BdesFR	Bacteroidales	GGTGTCGGCTTAAGTGCCAT	CGGAYGTAAGGGCCGTGC	61	151	1038-1189	56/59	Rintilla	B.fragilis	81500000

The copy/μ L column was the number actually entered in the MJ Opticon3 software, Version 3.1 (BioRad). The formula used for calculation is given in *SI Text*. Using RDP ProbeMatch; *, Subgroup coverage of Lactobacillaceae is 78% and Enterococcaceae is 85%; ^, Subgroup coverage of Lachnospiraceae is 76%(F) and 65%(R).

Other Supporting Information Files

Dataset S1