

Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases

Daniel N. Frank*, Allison L. St. Amand*, Robert A. Feldman†, Edgar C. Boedeker‡, Noam Harpaz§, and Norman R. Pace*¶

*Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO 80309-0347; †Symbio Corporation, Menlo Park, CA 94025; ‡Department of Medicine, University of New Mexico, Albuquerque, NM 87131; and §Department of Pathology, Mount Sinai School of Medicine, New York, NY 10029

Contributed by Norman R. Pace, July 16, 2007 (sent for review June 7, 2007)

The two primary human inflammatory bowel diseases, Crohn's disease (CD) and ulcerative colitis (UC), are idiopathic relapsing disorders characterized by chronic inflammation of the intestinal tract. Although several lines of reasoning suggest that gastrointestinal (GI) microbes influence inflammatory bowel disease (IBD) pathogenesis, the types of microbes involved have not been adequately described. Here we report the results of a culture-independent rRNA sequence analysis of GI tissue samples obtained from CD and UC patients, as well as non-IBD controls. Specimens were obtained through surgery from a variety of intestinal sites and included both pathologically normal and abnormal states. Our results provide comprehensive molecular-based analysis of the microbiota of the human small intestine. Comparison of clone libraries reveals statistically significant differences between the microbiotas of CD and UC patients and those of non-IBD controls. Significantly, our results indicate that a subset of CD and UC samples contained abnormal GI microbiotas, characterized by depletion of commensal bacteria, notably members of the phyla *Firmicutes* and *Bacteroidetes*. Patient stratification by GI microbiota provides further evidence that CD represents a spectrum of disease states and suggests that treatment of some forms of IBD may be facilitated by redress of the detected microbiological imbalances.

Crohn's disease | culture-independent microbiology | ulcerative colitis | rRNA

The human inflammatory bowel diseases (IBD), ulcerative colitis (UC) and Crohn's disease (CD), are chronic gastrointestinal (GI) illnesses of uncertain etiology, with symptoms ranging from diarrhea and weight loss to ulceration, perforation, and complete obstruction of the GI tract (1–3). Both diseases are incurable and often require extensive medical and surgical management. Medical treatment of IBD with antibiotic, anti-inflammatory, and/or immunosuppressive drugs is generally supportive rather than curative (4). Clearly, novel approaches to disease management are needed and will require more thorough understanding of the factors that contribute to IBD.

IBD is manifest in the context of an immense and highly diverse population of commensal GI microbes. Studies of experimental animal models of IBD reveal that germ-free animals normally manifest few signs of inflammation (5–14); the full effects of experimental colitis typically are exhibited only upon exposure to natural microbial communities (9, 12, 13). A role for microbes in human IBD similarly is suggested by the observation that fecal flow exacerbates inflammation of involved tissues (15–18). Current models of human IBD posit that pathogenesis arises from, and is perpetuated by, interactions between host genetic and immune factors, GI microbes, and environmental triggers (8). The net effect is a dysregulated hypersensitive T_H1 -type response of the gut-associated lymphoid tissue to luminal microbes and/or their antigens (2, 8). Considerable controversy remains, however, as to whether the entire com-

mensal microbiota or individual pathogens [e.g., *Mycobacterium avium* spp. *paratuberculosis* (MAP)] are primarily responsible for induction of inflammation (19–24).

The commensal GI microbiota provides a multitude of beneficial services to the healthy host, including maintenance of immune homeostasis (25–27), modulation of GI development (28, 29), and enhanced metabolic capabilities (30). Perturbations of the normally stable GI microbiota might therefore be predicted to adversely affect the health of the host. Indeed, recent studies demonstrate that obesity in humans and *ob/ob* mice is associated with stereotypical imbalances in the normal gut microbiota (31–33). Likewise, previous studies of human IBD, using standard culture techniques (34–36) or molecular analysis (37–43), have noted alterations in the GI microbiota. However, most of these IBD studies have been limited in statistical power and precision of identification or have examined only the fecal microbiota, which differs substantially from that of the GI mucosa (44). Because of their immediate proximity to affected tissues, microbes associated with the gut wall likely are a more critical factor than are fecal microbes in promoting IBD pathogenesis (45, 46). These previous studies also have treated UC and CD as monolithic disorders, thus precluding the possibility that different subpopulations exist within IBD classifications (47, 48). To surmount these problems, we performed a comprehensive culture-independent phylogenetic analysis of microbes associated with a greatly expanded set of surgically resected tissue specimens. Our findings indicate significant abnormalities in the microbiotas of a subset of CD and UC individuals.

Results

Broad-Range PCR Analysis. One hundred ninety resected tissue samples were obtained in approximately equal numbers from CD, UC, and non-IBD control subjects (primarily individuals being treated for GI cancer). The clinical features of the specimen set are summarized in supporting information (SI) Table 2. To minimize the presence of luminal bacteria, tissues were rinsed in 0.15 M NaCl before extraction of total genomic DNA. Fifteen thousand one hundred and seventy-two small-subunit rRNA (SSU rRNA) genes (≈ 80 per sample) were cloned

Author contributions: D.N.F., E.C.B., N.H., and N.R.P. designed research; D.N.F., A.L.S.A., R.A.F., and N.H. performed research; D.N.F. contributed new reagents/analytic tools; D.N.F., E.C.B., N.H., and N.R.P. analyzed data; and D.N.F. and N.R.P. wrote the paper.

The authors declare no conflict of interest.

Abbreviations: IBD, inflammatory bowel disease; CD, Crohn's disease; UC, ulcerative colitis; SSU rRNA, small subunit rRNA; OTU, operational taxonomic unit; MAP, *Mycobacterium avium* spp. *Paratuberculosis*; PCA, principal components analysis; Q-PCR, quantitative PCR.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. EF695452–EF710623).

¶To whom correspondence should be addressed. E-mail: norman.pace@colorado.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0706625104/DC1.

© 2007 by The National Academy of Sciences of the USA

and sequenced from these samples after broad-range PCR with panbacterial primers. The kinds of microbes resident in the resected tissues were identified by molecular-phylogenetic sequence analysis. Few identical sequences were observed between different samples, and so similar sequences were clustered into operational taxonomic units (OTUs). OTUs were assembled at a range of sequence identity thresholds so that sample sets could be compared at multiple phylogenetic depths (SI Fig. 6). No significant differences in sampling coverage were evident between sample categories at any OTU clustering level (SI Table 3), thus permitting meaningful comparison of samples.

The overall phylogenetic distribution of the sequences identified in this study is shown in Fig. 1. Regardless of disease state or anatomical site of sampling, the majority of sequences were associated with only four phyla of the Bacteria (Figs. 1 and 2): *Firmicutes* (49% of clones), *Bacteroidetes* (23%), *Proteobacteria* (21%), and *Actinobacteria* (5%). Nearly half ($\approx 45\%$) belonged to just two subgroups: the order *Bacteroidales* and the family *Lachnospiraceae* (which comprises the *Clostridium* XIVa and IV groups within the order *Clostridiales*; ref. 49).

Compared with the colon, samples of the small intestine were substantially enriched in sequences of the *Bacillus* subgroup of *Firmicutes* (primarily *Streptococcaceae*). In the case of non-IBD control samples, *Streptococcaceae* sequences accounted for 23% and 5% of total clones isolated, respectively, from the small intestine and colon (Fig. 2 and SI Fig. 7). Actinobacterial sequences also were more prevalent in the small intestine (8% of clones in non-IBD samples compared with 2% from non-IBD colon). Members of the actinobacterial subgroups *Actinomycinaeae* and *Corynebacteriaceae* accounted for the majority of this enrichment. In contrast, small-intestine samples contained fewer sequences of *Bacteroidetes* and *Lachnospiraceae*, compared with colonic samples.

OTU distributions were significantly associated with disease state ($P = 0.02$) but not anatomy ($P = 0.39$) or gross pathology ($P = 0.25$). Relative to non-IBD control colon samples, UC and CD colonic samples contained fewer sequence types representative of *Bacteroidetes* and the *Lachnospiraceae* subgroup of *Firmicutes* and concomitantly more sequences representative of *Proteobacteria* and the *Bacillus* subgroup of *Firmicutes* (Fig. 2). Among the small-intestine samples, proteobacterial sequences were more abundant and *Bacillus* sequences less abundant in both UC and CD sample sets relative to non-IBD samples. *Lachnospiraceae* sequences of non-IBD samples were intermediate in abundance between UC and CD samples, whereas *Bacteroidetes* sequences were unaffected by disease status. Similar results were observed when clonal prevalences, rather than frequencies, were compared across sample sets (SI Fig. 8).

Exploratory Data Analysis. Because of the idiopathic nature of IBD, we conducted exploratory statistical analyses to determine whether any structure is manifest within the sequence data set, independently of diagnostic categories. Principal components analyses (PCA) of sample vs. OTU presence/absence matrices were performed for OTUs defined at several phylogenetic depths (97% OTU results are shown in Fig. 3). MANOVA tests of the 97% OTU PCA data indicated that variation along the first two principal coordinate axes was significantly influenced by disease status ($P < 0.001$) rather than anatomy ($P = 0.4$) or pathology ($P = 0.7$). Similar results also obtained for PCA of OTUs defined at the 99%, 95%, 90%, and 85% levels (SI Fig. 9 and SI Table 4).

Hierarchical clustering of samples (SI Fig. 10) on the basis of their first two principal component scores separated the samples into two primary clusters (outlined in Fig. 3). One cluster (labeled “Control subset”) includes most of the non-IBD control samples as well as $\approx 2/3$ of CD and $\approx 3/4$ of UC samples. The other nominal cluster is dominated by CD and UC samples and

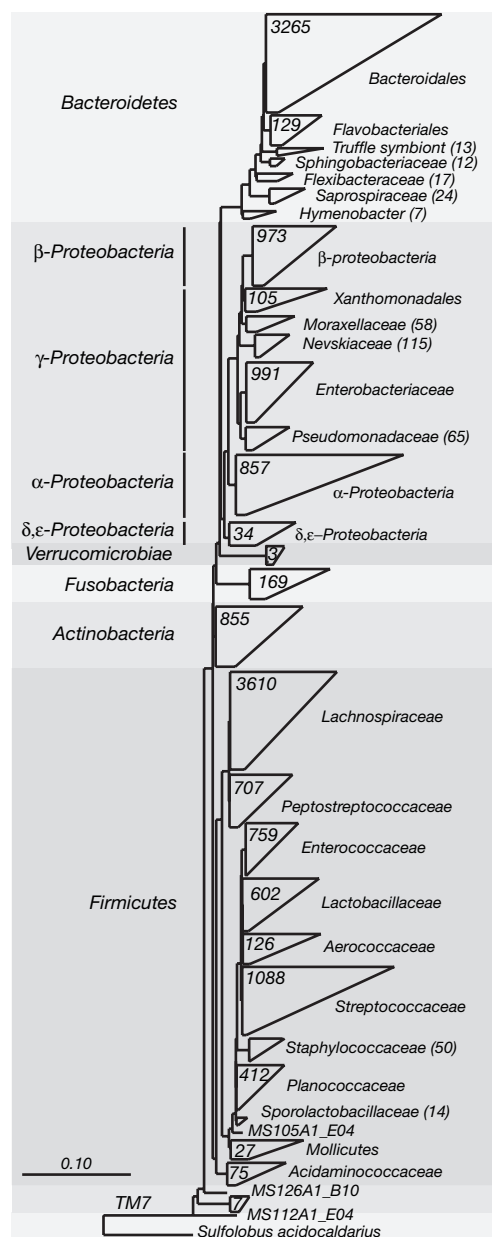


Fig. 1. Phylogenetic distribution of SSU rRNA sequences. The diagrammatic phylogenetic tree presents a summary of the rRNA sequences isolated in this study. The number of taxa in a clade is designated either on its wedge or parenthetically to its name. Phyla are named to the left of the tree and lower taxonomic levels to the right. Wedge widths represent the taxa with the longest (top bar) and shortest (bottom bar) distances within the clade. The scale bar represents base changes per site.

is thus designated the “IBD subset.” Hierarchical clustering of samples by raw OTU presence/absence produced similar results (SI Fig. 11), although more IBD samples clustered with the IBD subset. Furthermore, the UniFrac test (50), which assesses the significance of intercommunity differences in populations, strongly supports our hypothesis that IBD subset and Control subset specimens are composed of distinct microbial communities (SI Fig. 12).

The differences in rRNA clonal frequency that are evident from comparisons of disease state are even more pronounced when the data are parsed in terms of inclusion in, or exclusion from, the IBD subset (Fig. 2). Sequences representative of the

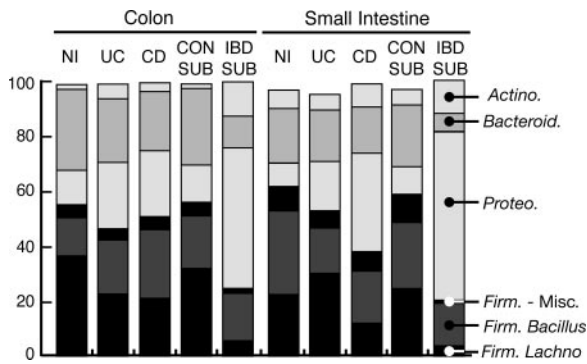


Fig. 2. Phylum-level comparison of disease state. Bars shading depicts percentages of cloned sequences in all samples of a particular category that belong to one of the predominant phylum of Bacteria identified in the sample set (*Firmicutes* are subdivided into *Bacillus*, *Lachnospiraceae*, and miscellaneous groups). Phyla of lower abundance were omitted, so bars do not sum to 100%. NI, non-IBD control samples; Con-Sub, Control subset; and IBD-Sub, IBD subset are explained in the text.

Bacteroidetes ($P < 0.001$) and *Lachnospiraceae* ($P < 0.001$) were greatly depleted, and those of the *Actinobacteria* ($P < 0.001$) and *Proteobacteria* ($P < 0.001$) were substantially more abundant in IBD-subset samples, relative to Control-subset samples. Summaries of the 97% OTU clusters that were the most depleted or enriched in prevalence in the IBD-subset samples are presented in Table 1 (more detailed lists are available in SI Tables 5 and 6). A diverse variety of *Actinobacteria* and *Proteobacteria* was increased in prevalence in the IBD subset. In contrast, most of the OTUs that were decreased in prevalence in the IBD subset belonged to either the *Lachnospiraceae* group or the phylum *Bacteroidetes*.

The disparities in microbiotas noted between the nominal IBD and Control subsets might simply be explained in terms of

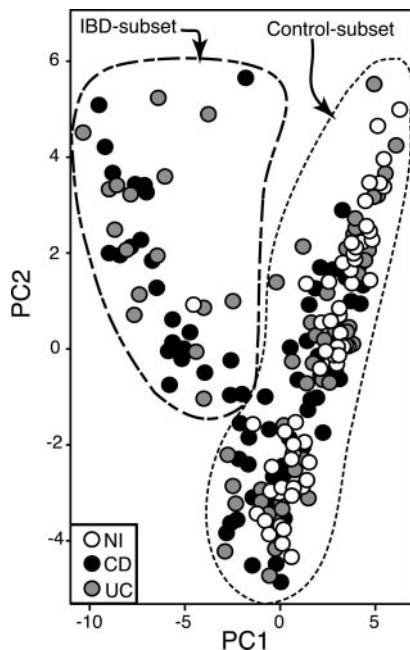


Fig. 3. PCA; 97% OTUs were encoded as presence/absence data for each sample, and the corresponding data matrix was subjected to PCA. Each circle, which is representative of a single sample and shaded according to disease status, is plotted along the first two principal component axes. Dashed lines denote nominal sample clusters, the Control subset and IBD subset.

different medical treatments to which patients were subjected. However, none of the treatments listed in patient records were significantly correlated with classification of samples into the IBD or Control subset, when analyzed individually or *in toto* (SI Table 7). In terms of reported disease complications, only the occurrence of abscesses in CD subjects was strongly predictive of IBD-subset inclusion: 8/24 IBD-subset CD samples and 0/44 Control-subset CD samples were positive for abscesses ($P < 0.001$; SI Table 7). CD subjects with a history of smoking were significantly more likely to be classified as Control subset than were reported nonsmokers ($P = 0.007$). Patient age at time of surgery also was correlated with the IBD/Control-subset split (SI Table 8); members of the IBD subset were, on average, 9.5 (CD; $P = 0.014$) or 6.3 (UC; $P = 0.08$) years younger than Control-subset subjects. Sex and previous appendectomy were not correlated with GI microbiota composition.

Quantification of Specific Microbial Groups. Shifts in OTU frequency in clone libraries could result from either blooms or diminished loads of particular microbial groups. To distinguish these possibilities, quantitative PCR (Q-PCR) analyses of total bacteria, *Bacteroidetes*, *Enterobacteriaceae* (a proxy for *Proteobacteria*), and *Lachnospiraceae* were performed for selected IBD- and Control-subset samples. As shown in Fig. 4, Q-PCR results revealed a 10-fold decrease in total bacterial load in the IBD subset ($P < 0.001$). Q-PCR also confirmed that members of the *Lachnospiraceae* ($P < 0.001$) and *Bacteroidetes* ($P < 0.001$) were diminished in quantity in the IBD-subset samples (statistical analyses were performed on data that were normalized to the Q-PCR measurements of total bacteria). The Q-PCR results indicated >300-fold (*Lachnospiraceae*) and 50-fold (*Bacteroidetes*) reductions in these microbial groups in the IBD subset, compared with the Control subset. In contrast, the measured levels of *Enterobacteriaceae* were not substantially different in IBD- and Control-subset samples ($P = 0.47$). The differences in IBD- and Control-subset clonal frequencies must therefore reflect diminution of *Lachnospiraceae* and *Bacteroidetes* microbial populations in patients' samples classified as IBD subset rather than increased absolute levels of proteobacteria.

Discussion

The chronic inflammation characteristic of UC and CD is suggestive of an inability on the part of patients to control basal level inflammation within gut-associated lymphoid tissues. The antigens responsible for immune system overstimulation are unknown but are likely to arise from the plethora of bacteria that reside within the human GI tract. In this study, we have used culture-independent phylogenetic analysis to characterize the GI microbiotas of IBD patients and non-IBD controls. The surgical samples analyzed provide access to the gut-wall-associated microbiotas of subjects, which may play more critical roles than do fecal microbes in IBD pathogenesis (45, 46). We acknowledge that this census reflects SSU rRNA gene copy number rather than true cell counts, and that the results are subject to biases inherent in PCR amplification and cloning. Furthermore, rRNA analysis will not detect functional changes in GI microbes, such as enhanced virulence, mucosal adherence, or invasion, that do not influence the relative proportions of species in the microbiota (45, 46, 51). Nonetheless, the results provide an internally consistent culture-independent assessment of the gut-wall-associated microbiota.

We report an in-depth molecular survey of microbiology within the human small intestine. Sequence analysis indicates that the wall of the distal small bowel is colonized by microbial populations that are not radically different from those of the large bowel. Both environments are dominated by bacteria of the phyla *Bacteroidetes* and *Firmicutes* but differ in the relative proportions in which subgroups of these phyla are present.

Table 1. Ninety-seven percent OTUs enriched or depleted in IBD subset

Top Blast hit*	Phylum [†]	ΔPrevalence, [‡] %
Depleted in IBD subset		
Bacterium mpn-isolate group 5	<i>Bacteroidetes; Bacteroidales</i>	-42.3
Bacterium mpn-isolate group 18	<i>Firmicutes; Lachnospiraceae</i>	-37.7
Butyrate-producing bacterium A2-A165	<i>Firmicutes; Lachnospiraceae</i>	-34.0
Butyrate-producing bacterium SR1/1	<i>Firmicutes; Lachnospiraceae</i>	-32.5
<i>Bacteroides thetaiotaomicron</i>	<i>Bacteroidetes; Bacteroidales</i>	-32.5
Butyrate-producing bacterium L2-L7	<i>Firmicutes; Lachnospiraceae</i>	-27.6
<i>Clostridium nexile</i>	<i>Firmicutes; Lachnospiraceae</i>	-26.2
Bacterium mpn-isolate group 19	<i>Firmicutes; Lachnospiraceae</i>	-26.2
Butyrate-producing bacterium SS2/1	<i>Firmicutes; Lachnospiraceae</i>	-24.1
<i>Alistipes</i> sp. WAL 8169	<i>Bacteroidetes; Bacteroidales</i>	-24.0
Enriched in IBD subset		
Drinking-water bacterium Y7	<i>Alphaproteobacteria</i>	76.0
Actinobacterium GWS-BW-H99	<i>Actinobacteria</i>	62.5
<i>Nocardioides</i> sp. NS/27	<i>Actinobacteria</i>	61.8
<i>Novosphingobium</i> sp. K39	<i>Alphaproteobacteria</i>	57.8
<i>Pseudomonas straminea</i>	<i>Betaproteobacteria</i>	53.7
Gamma proteobacterium DD103	<i>Gammaproteobacteria</i>	50.9
<i>Bacillus licheniformis</i>	<i>Firmicutes; Bacilli</i>	49.7
<i>Sphingomonas</i> sp. AO1	<i>Alphaproteobacteria</i>	49.6
<i>Actinomyces oxydans</i>	<i>Actinobacteria</i>	48.9
Acidimicrobiae Ellin7143	<i>Actinobacteria</i>	47.6

*Identified by Blast search of database culled of environmental clones.

[†]Determined by comparison to rRNA sequence database (49).

[‡]Difference in prevalence of an OTU between normal subset and IBD subset.

Members of the class *Bacillus* (predominantly the *Lactobacillales* clade) were significantly more abundant in the small intestine than in the colon. In contrast, a variety of members of the *Bacteroidetes* and the family *Lachnospiraceae* were less abundant in the small intestine. Moreover, the small-intestine samples, as a whole, contained less overall phylogenetic diversity of *Bacteroidetes* and *Firmicutes* than did the large intestine.

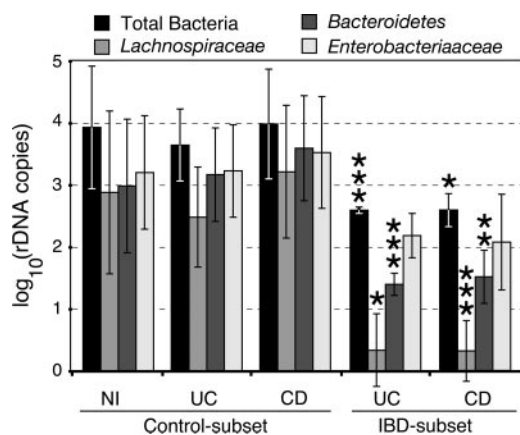


Fig. 4. Q-PCR analysis of predominant phylogenetic groups. Selected representatives of the Control- and IBD-subset clusters (see text for details) were assayed by Q-PCR for total bacteria, *Lachnospiraceae*, *Bacteroidetes*, and *Enterobacteriaceae*. Experiments were performed in triplicate. Data are presented as log₁₀-transformed SSU rDNA gene copy numbers, as interpolated from standard curves of plasmid controls. Error bars represent standard deviations of replicate experiments. NI, non-IBD controls. Control- and IBD-subset are explained in the text. The results of Student's *t* tests are indicated as *, *P* < 0.05; **, *P* < 0.01; and ***, *P* < 0.001. *t* tests compared the results of Control- and IBD-subset samples for a particular disease state and bacterial group (e.g., *Bacteroidetes*/CD/Control subset vs. *Bacteroidetes*/CD/IBD subset). Data for specific groups were normalized to total bacteria.

The role of MAP in CD pathogenesis continues to be the subject of contentious debate. Of the >15,000 SSU rRNA genes analyzed in this study, only 10 were from *Mycobacterium* spp., and all were more closely related to species other than MAP. Q-PCR assays of IBD and non-IBD genomic DNAs with MAP-specific (IS900) or *M. avium*-specific (rRNA) primers were negative (data not shown). Although these data cannot disprove the hypothesis that MAP causes CD, they establish upper limits on the abundance of this potential pathogen in the sample set.

The results of our bacterial survey (*i*) confirm previous reports of compositional shifts in IBD-associated GI microbial communities and (*ii*) fail to identify any individual species that is enriched in grossly abnormal IBD samples in a manner suggestive of an active etiologic agent. Recent studies of Manichanh *et al.* (41) and Gophna *et al.* (42) also have reported decreased loads of *Lachnospiraceae* in small sets (approximately six) of CD fecal samples or biopsies in comparison to UC and healthy controls. Our results demonstrate that a distinct subpopulation of IBD patients (the IBD-subset) harbor GI microbiotas that differ significantly from study controls and other CD and UC samples (Control-subset). Neither UC nor CD is characterized by a uniform, stereotypical microbiota. Because the phylogenetic distribution of Control-subset sequences is similar to that observed for healthy adults (32, 44), these samples likely are representative of healthy GI microbiotas. By this criterion, subjects classified in the IBD-subset harbor GI microbiotas that are significantly perturbed in composition relative to "normal" microbiotas.

Because uniform levels of *Enterobacteriaceae* were measured by Q-PCR in IBD- and Control-subset samples, it is unlikely that contamination of select samples resulted in skewed assessments of microbial diversity. Furthermore, reported patterns of recent (<2 years) antibiotic use, anatomical site of sampling, and gross pathology were not correlated with the microbiotas of IBD- or Control-subset samples. We therefore conclude that the observed differences in the microbiotas of IBD- and Control-subset

individuals are genuine and not the result of trivial experimental errors or variations in patient treatment.

The results demonstrate significant perturbations in the GI microbiotas of select IBD patients and thereby suggest that UC and CD are manifest in different forms that can be differentiated by microbial populations. *Bacteroidetes* and *Lachnospiraceae* were depleted 1.7 and 2.5 orders of magnitude, respectively, in specimens of the IBD subset. This is noteworthy, because these kinds of organisms promote GI health in myriad ways (30, 52). Several of the OTUs underrepresented in the IBD-subset are most closely related to cultured isolates of butyrate-producing bacteria (Table 1 and SI Table 5). Butyrate and other short-chain fatty acids, which are produced primarily by bacterial metabolism, are important sources of energy for colonic epithelial cells and may enhance epithelial barrier integrity and modulate the GI immune response (53–58). Loss of similar organisms has been correlated with antibiotic-associated diarrhea in a patient (59). We therefore propose a model in which dramatic reductions in the quantities of microbes that provide metabolic services to the host GI tract exacerbate some forms of IBD.

Our data do not address the causal relationships between the reported microbial community imbalances and IBD pathophysiology. Regardless of ultimate causation, the observed microbial imbalances, whether because of prolonged drug treatment or other factors, likely contribute to disease severity. That the hypothesized stratification of UC and CD patients on the basis of the observed microbial imbalances is clinically relevant is suggested by two observations. First, abnormal microbiotas were significantly correlated with the occurrence of abscesses in CD patients. The ability of the commensal microbiota to competitively exclude pathogens may have been disrupted in these cases, thus potentiating abscess formation. Second, patients with abnormal commensal microbiotas were significantly younger than those with normal microbiotas (8.1 years; $P = 0.002$) at the time of surgery. Loss of the normal GI microbiota might therefore be indicative of more-aggressive forms of IBD. Diet, long-term disease history, and human genetic factors (e.g., CARD15/NOD2 alleles) were not monitored in this study but constitute possible factors underlying the imbalances in selected commensal populations. Although antibiotic treatment *per se* was not correlated with the observed differences in microbial communities, patients' responses to treatment likely differed; perhaps individuals in the IBD subset were impaired in their capacities to reconstitute normal microbiotas after extensive medical treatment. Remediation of GI microbiotas in these cases, for example through application of specific probiotic regimens, might therefore ameliorate disease (60).

In addition to this study, at least two other large-scale culture-independent studies of the human GI tract have been reported (32, 44). Taken together, the >45,000 bacterial SSU rRNA genes surveyed reveal an unprecedented diversity of microbes associated with the human GI tract. Nearly all of these sequences (>98%) belong to only four bacterial phyla: *Firmicutes* (64%), *Bacteroidetes* (23%), *Proteobacteria* (8%), and *Actinobacteria* (3%). The lower taxonomic divisions, however, are quite diverse, and collector's curves (SI Fig. 13) do not plateau, indicating that much diversity remains uncharacterized. Nevertheless, we can place lower bounds on the richness of the human GI microbiome both by the S_{chao1} estimator (61) and by modeling bootstrap resampled collector's curves in a manner analogous to the use of Lineweaver–Burk plots in enzymology (Fig. 5). Both methods estimate that the human gut microbiome consists of at least 1,800 genera and $\approx 15,000$ –36,000 species of bacteria, depending on whether species are conservatively (97% OTUs) or liberally (99% OTUs) classified (SI Table 9). Extant sequences represent, at best, 50% of the predicted species-level diversity. Thus, despite the collective depth of these studies, a staggering level of microbial diversity remains to be characterized within the human gut environment.

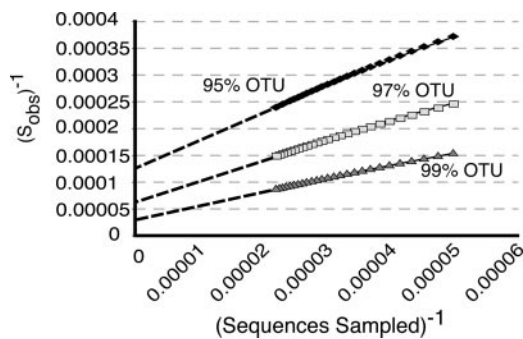


Fig. 5. Microbial diversity of the human intestine. OTU richness was estimated for a data set of sequences obtained in this study and those of Eckburg *et al.* (44) and Ley *et al.* (32). Sequences were clustered into OTUs and averaged collector's curves simulated by 1,000 random replicate samplings of the OTUs at sample sizes ranging from 1,000 to 45,000 sequences (total sequences, 45,172). The graph shows double reciprocal plots of mean S_{obs}^{-1} vs. sequences $^{-1}$ for 99%, 97%, and 95% OTUs. The Y intercepts of the graphs were extrapolated by linear regression of S_{obs}^{-1} and sequences $^{-1}$, from which expected OTU richness (i.e., S_{max}) was inferred. The dashed lines extending from each graph approximate the extrapolations.

Materials and Methods

Patients and Sample Collection. IBD was diagnosed on the basis of combined gross and microscopic features. After surgery at Mt. Sinai School of Medicine, 1.5 cm 2 of resected tissue was collected from each patient, placed in tubes containing ethanol/phenol/H $_2$ O, and shipped to Boulder, CO. The protocol was approved by the Institutional Review Boards of Mt. Sinai School of Medicine and the University of Colorado. The conditions that defined the negative control group are listed in *SI Text*.

Library Construction and Sequence Analysis. Molecular methods used standard protocols (62), elaborated in *SI Text*. DNA extraction and PCR steps included negative controls for detection of contamination by exogenous DNA. Multiple PCRs were performed for each sample, and products were pooled for further analysis. Species identifications were made by batch BLAST of rRNA databases and more detailed phylogenetic analyses of aligned sequences.

Q-PCR. Details of Q-PCR are described in *SI Text*. Plasmid quantification standards were prepared from representative clones of target organisms. Each Q-PCR experiment assayed all DNA templates under consideration and duplicate reactions of the plasmid dilution series and multiple negative-template controls. Quantification of template concentrations was made by linear extrapolation of data from standard curves. Q-PCR experiments were performed in triplicate for each primer set.

Statistical Comparisons of Communities. Statistical analyses, performed with the R software package (ver. 2.0.1), are described in greater detail in *SI Text*. For these analyses, sequence data were encoded as presence/absence of OTUs for each sample. Cluster analysis of samples used an agglomerative hierarchical clustering algorithm (63). Associations between medical treatments and membership in the IBD or Control subsets were examined by Fisher's exact test. Q-PCR data were normalized to human actin levels (measured for each sample) and analyzed by two-tailed Student's *t* test, without treating variances as equivalent. In the cases of specific bacterial groups, Q-PCR results were first normalized to total bacterial loads. The significance of environment clusters was assessed by 1,000

jack-knife resamplings by using the normalized weighted Uni-Frac test (50).

We thank Dr. D. Lang (National Institute of Allergy and Infectious Diseases (NIAID)/National Institute of Environmental Health Sciences)

for enthusiastic support and Dr. G. Carey, M. Amjadi, T. DeSantis, and M. Hamady for technical support. We thank Drs. L. Baumgartner and V. Visconti for valuable editorial comments. This project has been funded in whole or in part with Federal funds from the NIAID, including under contract no. N01-AI-30055.

- Farrell RJ, Peppercorn MA (2002) *Lancet* 359:331–340.
- Hendrickson BA, Gokhale R, Cho JH (2002) *Clin Microbiol Rev* 15:79–94.
- Shanahan F (2002) *Lancet* 359:62–69.
- Prantera C, Korelitz BI (1996) *Crohn's Disease* (Dekker, New York).
- Blumberg RS, Saubermann LJ, Strober W (1999) *Curr Opin Gastroenterol* 11:648–656.
- Pizarro TT, Arseneau KO, Cominelli F (2000) *Am J Physiol* 278:G665–G669.
- Rennick DM, Fort MM (2000) *Am J Physiol* 278:G829–G833.
- Sartor RB (2006) *Nat Clin Pract Gastroenterol Hepatol* 3:390–407.
- Dianda L, Hanby AM, Wright NA, Sebesteny A, Hayday AC, Owen MJ (1997) *Am J Pathol* 150:91–97.
- Matsumoto S, Okabe Y, Setoyama H, Takayama K, Ohtsuka J, Funahashi H, Imaoka A, Okada Y, Umesaki Y (1998) *Gut* 43:71–78.
- Onderdonk AB, Franklin ML, Cisneros RL (1981) *Infect Immun* 32:225–231.
- Rath HC, Herfarth HH, Ikeda JS, Grenther WB, Hamm TE, Balish E, Taugrog JD, Hammer RE, Wilson KH, Sartor RB (1996) *J Clin Invest* 98:945–953.
- Sellon RK, Tonkonogy S, Schultz M, Dieleman LA, Grenther W, Balish E, Rennick DM, Sartor RB (1998) *Infect Immun* 66:5224–5231.
- Taugrog JD, Richardson JA, Croft JT, Simmons WA, Zhou M, Fernandez-Sueiro JL, Balish E, Hammer RE (1994) *J Exp Med* 180:2359–2364.
- Harper PH, Lee EC, Kettlewell MG, Bennett MK, Jewell DP (1985) *Gut* 26:279–284.
- Oberhelman HA, Jr, Kohatsu S, Taylor KB, Kivel RM (1968) *Am J Surg* 115:231–240.
- Rutgeerts P, Goboos K, Peeters M, Hiele M, Penninckx F, Aerts R, Kerremans R, Vantrappen G (1991) *Lancet* 338:771–774.
- Winslet MC, Allan A, Poxon V, Youngs D, Keighley MR (1994) *Gut* 35:236–242.
- Prantera C, Scribano ML (1999) *Ital J Gastroenterol Hepatol* 31:244–246.
- Van Kruiningen HJ (1999) *Inflamm Bowel Dis* 5:183–191.
- El-Zaatari FA, Osato MS, Graham DY (2001) *Trends Mol Med* 7:247–252.
- Hermon-Taylor J (2001) *Gut* 49:755–756.
- Greenstein RJ (2003) *Lancet Infect Dis* 3:507–514.
- Sartor RB (2005) *Gut* 54:896–898.
- Rakoff-Nahoum S, Paglino J, Eslami-Varzaneh F, Edberg S, Medzhitov R (2004) *Cell* 118:229–241.
- Mazmanian SK, Liu CH, Tzianabos AO, Kasper DL (2005) *Cell* 122:107–118.
- Kelly D, Conway S, Aminov R (2005) *Trends Immunol* 26:326–333.
- Bry L, Falk PG, Midtvedt T, Gordon JI (1996) *Science* 273:1380–1383.
- Stappenbeck TS, Hooper LV, Gordon JI (2002) *Proc Natl Acad Sci USA* 99:15451–15455.
- Hooper LV, Midtvedt T, Gordon JI (2002) *Annu Rev Nutr* 22:283–307.
- Ley RE, Backhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI (2005) *Proc Natl Acad Sci USA* 102:11070–11075.
- Ley RE, Turnbaugh PJ, Klein S, Gordon JI (2006) *Nature* 444:1022–1023.
- Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI (2006) *Nature* 444:1027–1131.
- Giaffer MH, Holdsworth CD, Duerden BI (1991) *J Med Microbiol* 35:238–241.
- Krook A, Lindstrom B, Kjellander J, Jarnerot G, Bodin L (1981) *J Clin Pathol* 34:645–650.
- Merwe JPvd, Schroder AM, Wensick F, Hazenberg MP (1988) *Scand J Gastroenterol* 23:1125–1131.
- Seksik P, Rigottier-Gois L, Gramet G, Sutren M, Pochart P, Marteau P, Jian R, Dore J (2003) *Gut* 52:237–242.
- Mangin I, Bonnet R, Seksik P, Rigottier-Gois L, Sutren M, Bouhnik Y, Neut C, Collins MD, Colombel J-F, Marteau P, Dore J (2004) *FEMS Micro Ecol* 50:25–36.
- Prindiville T, Cantrell M, Wilson KH (2004) *Inflamm Bowel Dis* 10:824–833.
- Lepage P, Seksik P, Sutren M, de la Cochetiere MF, Jian R, Marteau P, Dore J (2005) *Inflamm Bowel Dis* 11:473–480.
- Manichanh C, Rigottier-Gois L, Bonnaud E, Gloux K, Pelletier E, Frangeul L, Nalin R, Jarrin C, Chardon P, Marteau P, et al. (2006) *Gut* 55:205–211.
- Gophna U, Sommerfeld K, Gophna S, Doolittle WF, Veldhuyzen van Zanten SJ (2006) *J Clin Microbiol* 44:4136–4141.
- Scanlan PD, Shanahan F, O'Mahony C, Marchesi JR (2006) *J Clin Microbiol* 44:3980–3988.
- Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA (2005) *Science* 308:1635–1638.
- Swidsinski A, Ladhoff A, Pernthaler A, Swidsinski S, Loening-Baucke V, Ortner M, Weber J, Hoffmann U, Schreiber S, Diemel M, Lochs H (2002) *Gastroenterology* 122:44–54.
- Swidsinski A, Weber J, Loening-Baucke V, Hale LP, Lochs H (2005) *J Clin Microbiol* 43:3380–3389.
- Greenstein RJ, Greenstein AJ (1995) *Mol Med Today* 1:343–348.
- Mishina D, Katsel P, Brown ST, Gilberts EC, Greenstein RJ (1996) *Proc Natl Acad Sci USA* 93:9816–9820.
- DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL (2006) *Appl Environ Microbiol* 72:5069–5072.
- Lozupone C, Knight R (2005) *Appl Environ Microbiol* 71:8228–8235.
- Darfeuille-Michaud A, Boudeau J, Bulois P, Neut C, Glasser AL, Barnich N, Bringer MA, Swidsinski A, Beaugerie L, Colombel JF (2004) *Gastroenterology* 127:412–421.
- Backhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI (2005) *Science* 307:1915–1920.
- Pryde SE, Duncan SH, Hold GL, Stewart CS, Flint HJ (2002) *FEMS Microbiol Lett* 217:133–139.
- Galvez J, Rodriguez-Cabezas ME, Zarzuelo A (2005) *Mol Nutr Food Res* 49:601–608.
- Scheppach W, Weiler F (2004) *Curr Opin Clin Nutr Metab Care* 7:563–567.
- Wong JM, de Souza R, Kendall CW, Emam A, Jenkins DJ (2006) *J Clin Gastroenterol* 40:235–243.
- Lara-Villoslada F, de Haro O, Camuesco D, Comalada M, Velasco J, Zarzuelo A, Xaus J, Galvez J (2006) *Eur J Nutr* 45:418–425.
- Segain JP, Raingeard de la Bletiere D, Bourreille A, Leray V, Gervois N, Rosales C, Ferrier L, Bonnet C, Blottiere HM, Galmiche JP (2000) *Gut* 47:397–403.
- Young VB, Schmidt TM (2004) *J Clin Microbiol* 42:1203–1206.
- Sartor RB (2004) *Gastroenterology* 126:1620–1633.
- Chao A (1984) *Scand J Stat* 11:265–270.
- Dalby AB, Frank DN, St Amand AL, Bendele AM, Pace NR (2006) *Appl Environ Microbiol* 72:6707–6715.
- Kaufman L, Rousseeuw PJ (2005) *Finding Groups in Data: An Introduction to Cluster Analysis* (Wiley, New York).