A Pyrosequencing Study in Twins Shows That Gastrointestinal Microbial Profiles Vary With Inflammatory Bowel Disease Phenotypes

BEN P. WILLING,[*] JOHAN DICKSVED,[*] JONAS HALFVARSON,[‡] ANDERS F. ANDERSSON,[§,‖] MARIANNA LUCIO,[¶] ZONGLI ZHENG,[#] GUNNAR JÄRNEROT,[‡] CURT TYSK,[‡,**] JANET K. JANSSON,[‡‡] and LARS ENGSTRAND[§]

[*]Department of Microbiology, Swedish University of Agricultural Sciences, Uppsala, Sweden; [‡]Division of Gastroenterology, Department of Medicine, [**]School of Health and Medical Sciences, Örebro University Hospital, Örebro, Sweden; [§]Department of Bacteriology, Swedish Institute for Infectious Disease Control, and Department of Microbiology, Tumor and Cell Biology, Karolinska Institute, Solna, Sweden; [‖]Science for Life Laboratory, Kunliga Tekniska hogskola Royal Institute of Technology, Stockholm, Sweden; [¶]Helmholtz-Zentrum Muenchen–German Research Center for Environmental Health, Institute for Ecological Chemistry, Neuherberg, Germany; [‡‡]Division of Earth Sciences, Lawrence Berkeley National Laboratory, Berkeley, California; and the [#]Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Solna, Sweden

Podcast interview: www.gastro.org/gastropodcast; see editorial on page 1816.

BACKGROUND & AIMS: The composition of the gastrointestinal microbiota is thought to have an important role in the etiology of inflammatory bowel diseases (IBDs) such as Crohn's disease (CD) and ulcerative colitis (UC). Interindividual variation and an inability to detect less abundant bacteria have made it difficult to correlate specific bacteria with disease. **METHODS:** We used 454 pyrotag sequencing to determine the compositions of microbial communities in feces samples collected from a cohort of 40 twin pairs who were concordant or discordant for CD or UC, and in mucosal samples from a subset of the cohort. The cohort primarily comprised patients who were in remission, but also some with active disease. **RESULTS:** The profiles of the microbial community differed with disease phenotypes; relative amounts of bacterial populations correlated with IBD phenotypes. The microbial compositions of individuals with CD differed from those of healthy individuals, but were similar between healthy individuals and individuals with UC. Profiles from individuals with CD that predominantly involved the ileum differed from those with CD that predominantly involved the colon; several bacterial populations increased or decreased with disease type. Changes specific to patients with ileal CD included the disappearance of core bacteria, such as *Faecalibacterium* and *Roseburia*, and increased amounts of *Enterobacteriaceae* and *Ruminococcus gnavus*. **CONCLUSIONS: Bacterial populations differ in abundance among individuals with different phenotypes of CD. Specific species of bacteria are associated with ileal CD; further studies should investigate their role in pathogenesis.**

Keywords: Gastrointestinal Microbiome; CCD; ICD; Pyrosequencing; Microbial Profiling.

Inflammatory bowel diseases (IBDs) including Crohn's disease (CD) and ulcerative colitis (UC), result from inappropriate activation of the gastrointestinal mucosal immune system by the intestinal microbiota. Although the exact role of the microbiota in etiology is not yet known, genetic susceptibility and environmental triggers have been shown to play an important role.[1] The current hypothesis is that the breakdown in the host–microbial mutualism is the consequence of an imbalance between protective and harmful bacteria (dysbiosis),[2] and there is growing interest in identification of microbes that kindle the inflammation. Some changes in the microbial community are shared in CD and UC including reduced diversity (particularly *Firmicutes*),[3] presence of noncommensals,[2,4] and increased concentrations of *Escherichia coli* including pathogenic strains.[5,6] However, some changes are specific, including reduced presence of the *Clostridium leptum* group, particularly *Faecalibacterium prausnitzii* in CD.[2,4,6–9] This reduction is particularly evident for individuals with CD localized in the ileum (ICD), compared with those with primarily colonic involvement (CCD).[3,6,10,11] A reduction of the *Clostridium coccoides* group, but no specific members, was reported for individuals with UC.[4,12]

Establishment of the correct diagnosis is important because treatment strategies for CD and UC differ. Recently, a gene expression assay for inflamed colonic biopsies was proposed.[13] However, many patients with IBD are in remission[14] and maintenance therapies differ depending on the initial diagnosis.[15] Thus, biomarkers are needed for patients both in remission and with active

Abbreviations used in this paper: CCD, Crohn disease with primarily colonic involvement; CI, confidence interval; ICCD, ileocolonic Crohn disease; ICD, Crohn disease localized in the ileum; OTU, operational taxonomic units; RDPII, Ribosomal Database Project.

0016-5085/$36.00
doi:10.1053/j.gastro.2010.08.049

disease for correct diagnosis of UC and CD (both ICD and CCD). There is also a need for better diagnostic tools for CD involving both the ileum and the colon, which is denoted ileocolonic Crohn's disease (ICCD).

Profiling the fecal microbiome using methods based on the 16S ribosomal RNA gene is less biased than cultivation-based approaches. In particular, pyrosequencing using parallel bar-coded sequence tags enables deep probing of multiple samples and provides high taxonomic resolution.[16,17] The power of pyrotag sequencing for exploration of the human microbiome has been shown in different body sites[16,18] and in response to antibiotic treatment.[19]

The aim of this study was to deeply explore the composition of the IBD gut microbiome using pyrotag sequencing and to identify specific bacterial signatures associated with IBD phenotypes. Because of the known interindividual variation in the human microbiome it is difficult to correlate specific bacteria with disease. Therefore, we studied a set of twin pairs to specifically focus on disease influence in genetically matched individuals. The cohort included twins who were concordant for disease and those who were discordant. A subset of the twins, assessed using low-resolution molecular fingerprinting techniques, previously were found to have significant differences in their gut microbiomes according to CD disease status.[6,20] Here, we applied pyrotag sequencing to a large twin cohort (40 twin pairs), including twins with UC, ICD, CCD, and ICCD, with the ultimate goal of defining bacterial signatures as potential diagnostic and monitoring targets.

Materials and Methods

Human Subjects

A set of 40 twin pairs (29 monozygotic, 11 dizygotic) obtained from a previously described Swedish population[21] were studied (Supplementary Table 1). Twin pairs with one twin previously hospitalized for IBD were identified by running the Swedish twin registry against the Swedish Hospital Discharge Register. After written consent from each twin, medical notes were scrutinized to verify the diagnosis of IBD[22] and to phenotype the disease according to the Montreal classification.[23] Twin pairs of the same sex, and if both twins in each pair had approved further contact and not undergone extensive IBD-related surgical resection (ie, colectomy), were invited to undergo colonoscopy. The twins were asked to send fecal samples 7–10 days before the colonoscopy and to fill in a questionnaire regarding environmental exposure, dietary habits, antibiotics, and drug use as previously described[20] (Supplementary Table 1). Antibiotic use beyond a year before sampling was not recorded. In UC, disease activity was classified using the Mayo score[24] in patients undergoing colonoscopy (n = 13) and using the Investigator Global Evaluation in patients who did not undergo colonoscopy (n = 3). In CD, the Harvey Bradshaw score[25] was calculated in patients undergoing colonoscopy (n = 24) and a modified Harvey Bradshaw score (without evaluation of a possible abdominal mass) was used in patients who did not undergo colonoscopy (n = 5). UC twins included 14 discordant pairs and 1 concordant pair. CD twins included 6 concordant pairs and 17 discordant pairs. Within the CD cohort, 15 individuals had ICD, 12 had CCD, and 2 had ICCD. Two healthy monozygotic twin pairs were included as controls and indicators of microbial similarities between healthy twins. Patient groups, as defined by disease, were similar in age (mean ± standard deviation [SD]) as follows: UC, 54.8 ± 12.4 (n = 16); CCD, 47.2 ± 7.6 (n = 12); healthy, 51.9 ± 11.5 (n = 35); ICD, 55.9 ± 16.1 (n = 15); and ICCD 49–67 (n = 2). Twins who were concordant for disease displayed phenotypic similarity. The use of human subjects for this study was approved by the Örebro County Ethical Committee (Dnr167/03).

Biopsy Collection

Biopsy specimens were analyzed from a subset of the twin cohort, which included a total of 9 twin pairs who were discordant or concordant for CD. The collection of biopsy specimens was described previously.[6]

DNA Extraction

DNA was extracted from 250 mg of feces using the MoBio Power Soil DNA Kit (Solana Beach, CA) according to the manufacturer's instructions. DNA from biopsies was isolated from the entire biopsy using the QIAamp DNA Mini Kit (Qiagen, Inc, Hilden, Germany) with the additional bead-beating steps on a FastPrep-24 (MP Biomedicals, Solon, OH) as previously described.[6]

454 FLX Sequencing

To investigate bacterial community composition in extracted DNA, V5 and V6 variable regions of the 16S ribosomal RNA gene were amplified by polymerase chain reaction (PCR) using forward primer (784f 5′AGGATTAGATACCCTGGTA3′) and reverse primer (1061r 5′CRRCACGAGCTGACGAC3′). The reverse primer was tagged with 1 of 4 labels at the 5' end along with the adaptor sequence (5′GCCTCCCTCGCGCCATCAG3′) to allow 4 samples to be included in a single 454 FLX sequencing lane as previously described.[16] Detailed PCR and purification procedures are described in the Supplementary Materials and Methods section. Pyrotag sequencing was performed on a 454 Life Sciences Genome Sequencer FLX machine (Roche, Broma, Sweden), at the Royal Institute of Technology (KTH, Stockholm, Sweden).

Taxonomic Analysis

Sequences were checked for quality and those that were less than 200 base pairs in length, contained incorrect

primer sequences, or contained more than 1 ambiguous base were discarded. Remaining sequences then were subjected to complete linkage clustering in the Ribosomal Database Project II (RDPII) by using a conservative 5% dissimilarity to define operational taxonomic units (OTUs) because of the short sequence length. The most abundant sequence from each OTU was selected as a representative sequence and was taxonomically classified by Basic Local Alignment Search Tool (BLAST) searching against a local BLAST database composed of 269,420 bacterial 16S ribosomal RNA gene sequences longer than 1,200 bp with good Pintail scores from RDP v. 10.7. The OTU inherited the taxonomy (down to genus level) of the best scoring RDP hit fulfilling the criteria of 95% or more identity over an alignment length of 180 base pairs or more.

Statistical Analysis

Statistical comparisons between disease phenotypes and the healthy group were made at the phylum, class, order, family, genus (full RDP taxonomy), and core OTUs (95% identity) levels. Because there were excessive zero counts for some variables, Vuong tests were performed to select standard- or zero-inflated negative binomial regression models.[26] Further tests of data overdispersion (shape parameter α) were used to select negative binomial or Poisson regression models. Logarithm of total number of reads for each patient was included in the models as offset for normalization. Clustered sandwich estimator of variance was used to adjust for correlation among twin pairs (Stata, 11.0; StataCorp LP, College Station, TX). For those variables with rare abundance (presented in only 1 or 2 patients and <10 reads of >2000 total reads), comparisons were not made for considerations of potential sequence misclassification and regression model convergence. The Benjamini and Hochberg[27] method was used to account for multiple comparisons, based on global P values of all variables compared, with a false-discovery rate of 5% (5 of 89 claimed differences). The comparisons of OTU numbers and core OTU numbers were performed using the same approach as described earlier. The comparison of diversity was made using a generalized linear regression model with clustered sandwich estimator of variance. The comparisons of similarities between individuals in twin pairs were analyzed by 1-way analysis of variance. Cluster analysis was performed using Bray–Curtis metrics.

We aimed to discriminate the microbial community profiles for different disease phenotypes. To this end we performed a partial least-squares discriminant analysis applying a Pareto scaling using SIMCA-P 12.0 (Umetrics, Umeå, Sweden). It was applied to a reduced data set of OTUs that were present in 5 or more of the samples. The partial least-squares discriminant analysis model was tested against the goodness-of-fit parameter and the predictive ability parameter, which is calculated by a 7-round internal cross-validation of the data. The different phenotypes represented the different classes and the OTUs were the predictors. Ignoring possible correlations between twin pairs, our approach was to split the data set into a training data set and a test data set. This in turn allowed testing of the performances of the model and the stability of the discrimination of different classes.

Results

A total of 248,320 16S sequences with acceptable quality were obtained with an average of 2178 reads (range, 541-3986 reads) per sample. By using 5% dissimilarity as an indicator of an OTU, a total of 6657 OTUs (6118 in fecal samples) were found in the 80 individuals with an average of 333 OTUs. A number of OTUs in fecal samples from healthy (95% confidence interval [CI], 317–410), CCD (95% CI, 293–433), and UC (95% CI, 328–459) were higher ($P < .001$) than in patients with ICD (95% CI, 110–259) (Figure 1*A*). The reduced number of OTUs was not a result of reduced read numbers because read numbers were similar (Figure 1*B*), but instead corresponded to reduced microbial diversity in ICD[28] (Figure 1*C*).

Healthy Fecal Core Microbiome

We defined 87 OTUs that were present in more than 50% of the healthy subjects as the healthy core microbiome, similar to a recent study[29] (Supplementary Table 2). These OTUs represented 44% of the reads from healthy individuals. Nearly all of the core OTUs (86 of 87) were present in healthy individuals in both concordant and discordant twin pairs, thus indicating that the healthy core microbiome was not skewed because of genetic predisposition to IBD. A majority of the core OTUs were represented by the 2 main phyla: *Firmicutes* (n = 57) and *Bacteroidetes* (n = 26); plus 2 *Actinobacteria: Collinsella aerofaciens* and *Bifidobacterium adolescentis*, and 2 *Proteobacteria*: *Bilophila*, a member of the *Desulfovibrionaceae* family, and an unclassified *Burkholderiales*. Although patients with UC and CCD shared the majority of the same core OTUs as healthy individuals, ICD subjects had dramatically fewer of the core OTUs ($P < .001$) (Figure 1*D*).

Discrimination Between Disease Phenotypes

We tested a multivariate statistical approach (partial least-squares discriminant analysis) as a means to distinguish disease phenotypes based on their bacterial compositions in fecal samples. The model derived from the training set led to an $R^2(Y)$ and $Q^2(Y)$ of 0.8 and 0.4, respectively. The model defined a clear divide among ICD, CCD, and healthy individuals. In contrast, UC and healthy samples could not be disentangled by the model as can be seen in the scatter plot in Figure 2. Moreover, the UC samples were the samples in the test set that were not predicted correctly. Our model also placed the samples from ICCD patients that

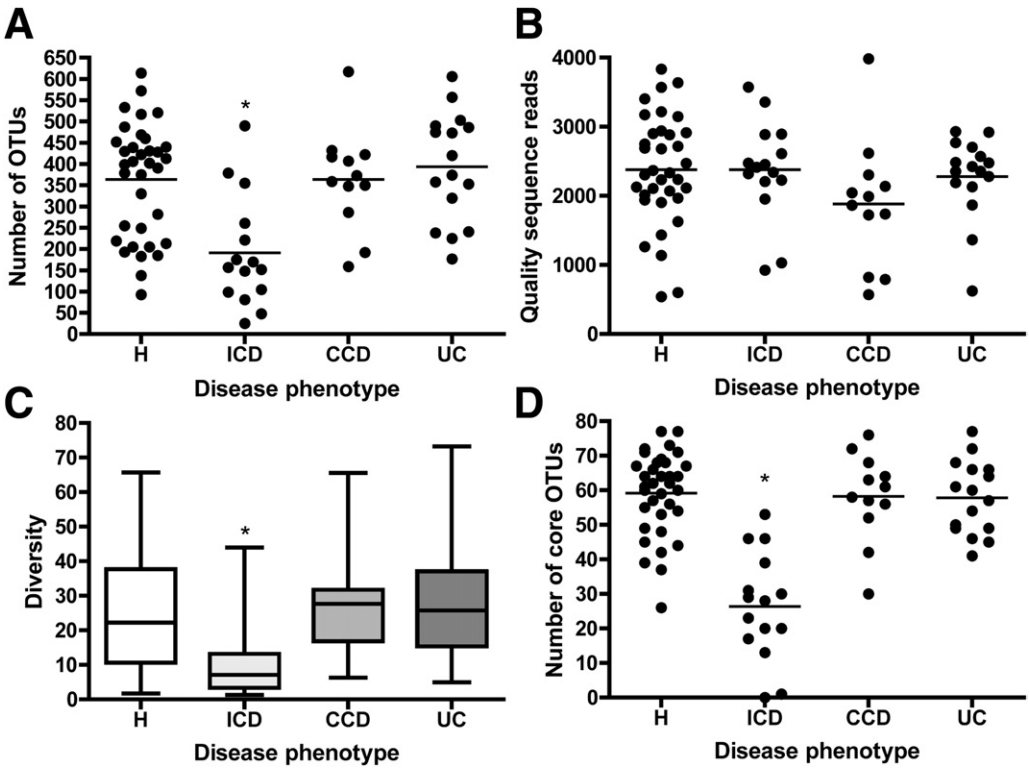


Figure 1. The affect of IBD disease phenotype on bacterial diversity. (*A*) Number of OTUs (95% identity) detected in fecal samples from patients with CCD, ICD, UC, and healthy subjects. (*B*) Number of reads per sample in each phenotype. (*C*) Simpson's diversity measure. (*D*) Number of healthy core OTUs detected in each individual. *Significantly different from all other phenotypes ($P < .05$).

could not be clinically classified as ICD or CCD, into the ICD group. As an outcome of the regression computation we derived the coefficients for ICD, CCD, and healthy samples. OTUs that contributed the most to discrimination between disease phenotypes (ie, had the strongest regression coefficients) are reported in Supplementary Table 3. When we ran the model with our sample cohort, a majority of healthy (97%), CCD (91.6%), and ICD (100%) samples were classified correctly, but none of the UC samples. Cross-validation of the model showed that prediction accuracy ranged between 71% and 98%. A small number of the patients (2 CCD and 4 ICD) had active disease; however, the model did not discriminate them from patients in remission.

Interphenotype Differences in Bacterial Clades

Abundances of different bacterial clades were compared between each disease with taxonomic classification from phylum to genus (Table 1). Although we originally planned to focus on differences between healthy and sick individuals within discordant twin pairs, we found that disease phenotype was a more important factor than genotype in discrimination of the data (as further discussed later). Differences between disease phenotypes could be seen at all taxonomic levels from phylum to OTU, and patterns were similar in fecal samples and biopsy specimens, but biopsy specimens were collected only from a subset of patients. It is noteworthy that relative abundances of some of the differentiating populations differed between feces and biopsy specimens (Figure 3).

As suspected from our partial least-squares discriminant analysis model (Figure 2), few differences were observed between UC and healthy at any taxonomic level from phylum to genus. There were, however, numerous differences between the fecal communities in CCD and ICD compared with healthy. We found that CCD and ICD showed opposite shifts in levels of several bacterial populations. At the phylum level CCD had more ($P < .02$) *Firmicutes* than healthy, whereas ICD tended to have fewer ($P < .06$). Conversely, ICD had more *Proteobacteria* than healthy subjects ($P < .01$), whereas this difference was not observed between CCD and healthy.

Four bacterial families were more abundant in CCD than healthy: 2 families of *Actinobacteria* (*Bifidobacteriaceae*, *Coriobacteriaceae*), 1 family of *Firmicutes* (*Ruminococcaceae*), and 1 family of *Tenericutes* (*Anaeroplasmataceae*) (Figure 4). ICD individuals had a shared absence of normal inhabitants of the gut microbiota (Figure 3). Of the 87 core OTUs, 24 were less abundant ($P < .05$) in ICD (Supplementary Table 2). This was particularly the case for members of the *Ruminococcaceae* family including *Faecalibacterium*, *Ruminococcaceae incertae sedis*, and unclassified *Ruminococcaceae* (Figure 3). *Ruminococcaceae* as a

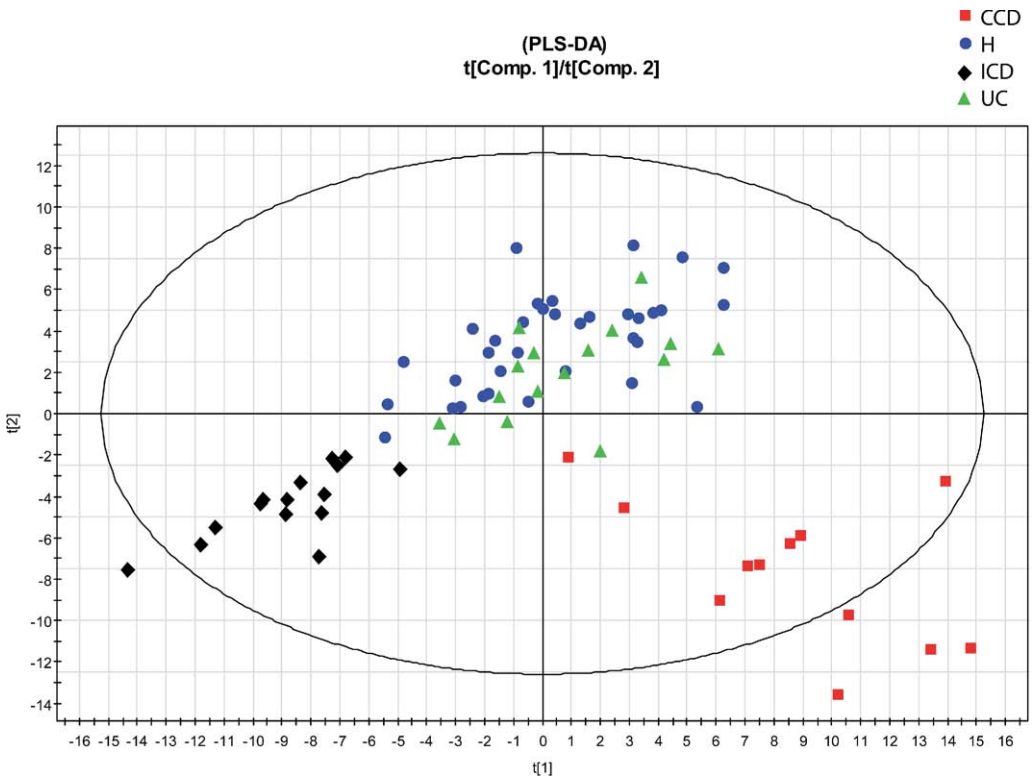


Figure 2. Discrimination of disease phenotypes based on microbial composition. Score scatter plot of partial least-squares discriminant analysis (PLS-DA) (predictive ability parameter [Q^2] [cum] = 0.4, goodness-of-fit parameter [R^2][Y] = 0.8) of microbial profiles classified as ICD, CCD, healthy, and UC patients.

group were underrepresented in ICD, to the point where not a single *Ruminococcaceae* sequence was detected in 3 ICD individuals. Other genera from diverse phyla showed the same pattern of reduced abundance in ICD patients including *Alistipes* (*Bacteroidetes*), *Collinsella* (*Actinobacteria*), and *Roseburia* (*Firmicutes*) (Figure 3), although not all were significant. *Lachnospiraceae*, the other main family of *Firmicutes* in the gut, did not show the same pattern as *Ruminococcaceae*; however, some members of this family including *Roseburia* were less abundant in ICD ($P < .01$). The increase in *Proteobacteria* in ICD was attributed mainly to OTUs representing *Enterobacteriaceae* (Figure 4). Although the nearest matches for many of these OTUs were unclassified *Enterobacteriaceae*, many were more than 99% identical to *E coli*.

Three of the core OTUs that were less abundant ($P < .05$) in ICD patients were also less abundant in CCD patients. The nearest matches for these OTUs were from the *Lachnospiraceae incertae sedis*, *Alistipes*, and unclassified *Ruminococcaceae* genera. Genera of the *Ruminococcaceae* family, including *Faecalibacterium*, *Ruminococcus*, and unclassified *Ruminococcaceae*, were more abundant in CCD than in healthy samples. Other genera that were more abundant in CCD included *Bifidobacterium* and *Collinsella*.

Although many OTUs were reduced dramatically in ICD, only 2 OTUs were remarkably higher in ICD compared with healthy and both were classified as *Lachnospiraceae*. The representative sequence of the most significant OTU was identical to *Ruminococcus gnavus* (ATCC 29149), which was also the most similar cultured bacterium for the other OTU (97% identity). These OTUs ranged from undetectable to as high as 2.86% of the population and were detected in 8 of 15 individuals with ICD with an average of 0.44%. Differences were even more pronounced in the biopsy samples where these 2 OTUs represented an average of 7.86% and 9.69% in the distal colon and ileum, respectively, and were detected in all patients sampled with ICD (Figure 3).

Consistency Between Biopsy Locations

Bray–Curtis cluster analysis showed that microbial profiles of biopsy specimens from the descending colon and ileum within an individual were highly similar to each other (Figure 5), and this similarity between locations was unaffected by disease. All biopsy specimens analyzed were from patients in remission; therefore, the effect of inflammation could not be tested in this sample set. The Bray–Curtis metrics of biopsy specimens from the same individual had the following 95% CIs: 59.0–71.0, 71.6–86.3, and 73.5–89.0 at the OTU, genus, and family levels, respectively (Figure 6).

Table 1. Bacterial Clades That Differed Between IBD Phenotypes and Healthy Subjects

Phylum	Class	Order	Family	Genus
Actinobacteria	**Ayctinobacteridae** ⇑	**Bifidobacteriales** ⇑	**Bifidobacteriaceae** ⇑	**Bifidobacterium** ⇑
	Coriobacteridae ⇑	**Coriobacteriales** ⇑	**Coriobacteriaceae** ⇑	**Collinsella** ⇑
Bacteroidetes	Bacteroidetes	Bacteroidales	Porphyromonadaceae	**un_Porphyromonadaceae** ↓↥
				Hallella ⇓
			Prevotellaceae	**Prevotella** ↓
			un_Bacteroidales ↓⇓	
Firmicutes ⇑	**Clostridia** ↓⇑	**Clostridiales** ↓⇑	Lachnospiraceae	**Roseburia** ↓
				un_Lachnospiraceae ↓
			Ruminococcaceae ↓⇑	**Ruminococcaceae Incertae Sedis** ↓
				Faecalibacterium ↓⇑
				Ruminococcus ⇑
				un_Ruminococcaceae ↓⇑
			IncertaeSedisXIII	**Anaerovorax** ↑
			Peptococcaceae ↓	**Peptococcus** ↓
			Un_Clostridiales ↓	
			Veillonellaceae	**Acidaminococcus** ↑
				Veillonella ↑⇓
		un_Clostridia ↓⇓		
	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	**Catenibacterium** ↓
	Bacilli ↑	**Lactobacillales** ↑	**Lactobacillaceae** ↑	**Lactobacillus** ↑
				Lactococcus ⇓
			Streptococcaceae ↓	**Streptococcus** ↓
Fusobacteria ↑⇓	**Fusobacteria** ↑⇓	**Fusobacteriales** ↑⇓	**Fusobacteriaceae** ↑⇓	**Fusobacterium** ↑⇓
Proteobacteria ↑	**Gamma-proteobacteria** ↑	Aermonadales	**Aeromonadaceae** ↑	**Aeromonas** ↑
		Enterobacteriales ↑	**Enterobacteriaceae** ↑	**Citrobacter** ↑
				Shigella ↑
				un_Enterobacteriaceae ↑
	Alpha-proteobacteria ⇓	**un_Alpha-proteobacteria** ⇓		
Tenericutes ↓⇑↓	**Mollicutes** ↓⇑↓	Anaeroplasmatales	**Anaeroplasmataceae** ↓⇑↓	**Asteroleplasma** ↓⇑↓

NOTE. IBD phenotypes included ICD, CCD, and UC. Bold entries indicate significant difference between groups ($P < .05$).
⇑⇓ Increase or decrease in CCD relative to healthy patients.
↑↓ Increase or decrease in ICD relative to healthy patients.
↓↑ Increase or decrease in UC relative to healthy patients.

Similarities in Twin Pairs

Biopsy specimens from concordant twin pairs grouped together, indicating some influence of genetics on mucosal-associated microbiota, but this grouping was not evident for discordant twin pairs. Also, for fecal samples, healthy twin pairs had very similar community compositions, with high similarities at the genus (69.6 and 77.6) and family (72.2 and 79.1) levels (Figure 6). In the majority (31 of 38) of comparisons of twin pairs with one or both individuals having disease, lower similarities were observed and similarities were comparable for concordant and discordant twins. Reduced similarity between twin pairs was most dramatic when at least one twin had ICD. Some diseased dizygotic twins were included in the cohort for comparison (5 UC, 1 CCD, 3 ICD, and 2 ICCD). However, there were no differences in their similarities to each other when compared with the monozygotic diseased twin pairs ($P = .57$).

Discussion

The results of this study indicate that some IBD diseases can be differentiated using the predictive model we created based on gut bacterial composition. In particular, several members of the gut microbiome differed in abundance in individuals with IBD and we identified several novel bacterial populations that are in higher or lower abundances in ICD and CCD compared with healthy. However, our model was not able to discriminate UC from healthy based on insufficient differences in their gut microbial compositions.

One of the goals of this study was to study twin pairs because they are matched genetically, and thus we could focus on the contribution of environmental factors, such as disease, on the microbiome.[30] Although we attempted to control for as many variables as possible we did not have information on antibiotic use beyond the most recent year, which may have affected microbial composition. In accordance with our findings, a previous study that monitored the gut microbiome of concordant obese and lean twins found that similarity between twins was independent of zygosity.[31] In our study, the effect of disease phenotype was greater than that of host genotype and because all of the dizygotic twins were discordant, we found no difference between monozygotic and dizygotic

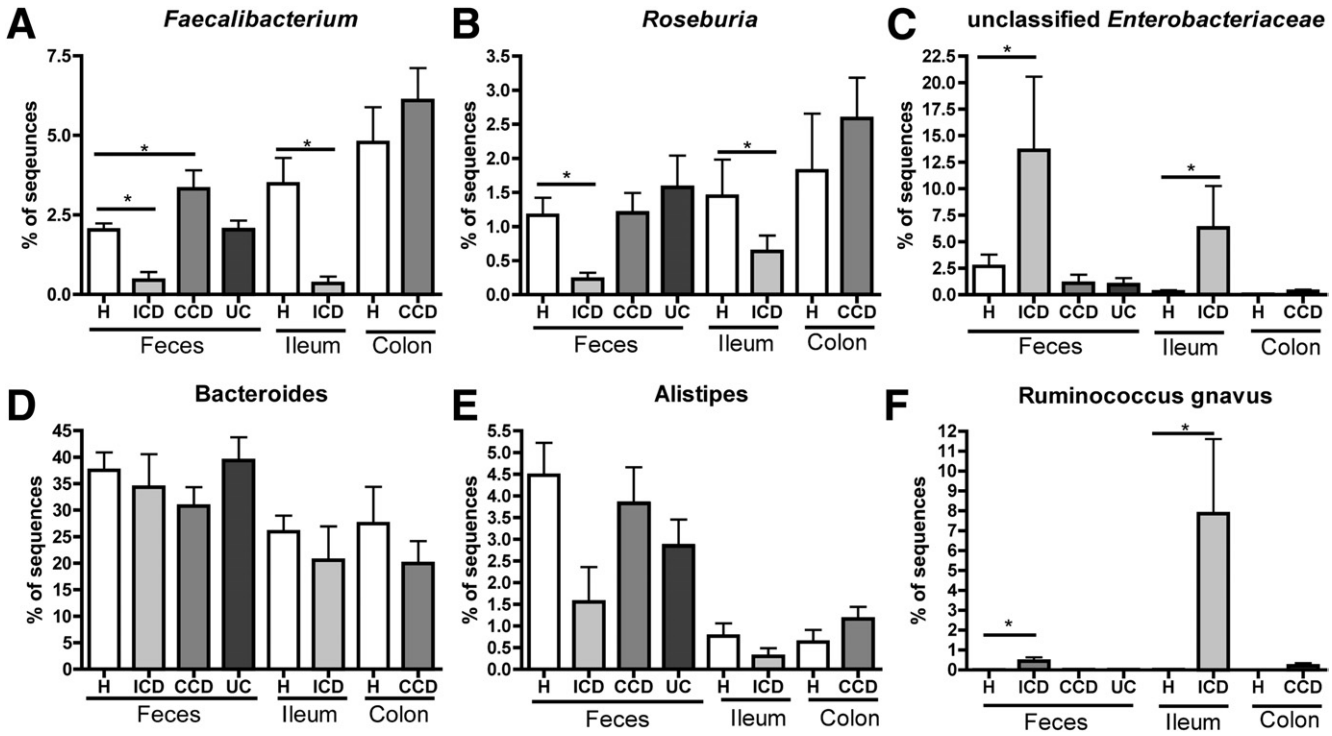


Figure 3. Changes in select bacterial genera in fecal samples from patients with (*A*) CCD, (*B*) ICD, (*C*) UC, and (*D*) healthy patients. *Significantly different from H ($P < .05$).

twin pairs. We did find that the gut microbiomes of healthy twin pairs were similar at the genus level but much less so at a finer level of phylogenetic resolution. However, even a similarity at the genus level strongly emphasizes that the host has an influence on the gut microbiome. This high similarity is especially noteworthy considering that the individuals we studied were older and had lived apart for years, or even decades. These data enforce the hypothesis that human beings are genetically programmed to maintain a certain gut microbiota, or that environmental exposure during early childhood has a strong and persistent impact on the adult microbiota. An even higher degree of similarity (OTU level) was found for mucosal and fecal samples from the same individual, as previously reported,[6] emphasizing the high degree of similarity in different regions of the bowel.

Similar to a recent report[29] we found that the healthy core microbiome was dominated by *Firmicutes*, particularly members of the *Lachnospiraceae* and *Ruminococcaceae* families. The single member of the *Coriobacteriaceae* family, *C aerofaciens*, previously described as a core member,[29] also was identified in the current study. We found notably more core OTUs from the *Bacteroidetes* phylum in the core microbiome compared with the previous study.[29] The increase corresponds to an overall increase in *Bacteroidetes* sequences detected in this study (54.7% of total sequences) as compared with the previous study (16.9%). These differences may be a result of differences in microbial compositions of individuals sampled or owing to differences in extraction methods, amplification bias, or our greater depth of analysis (average, 2178 vs 615). Conversely, we did not observe any members of the *Eubacteriaceae* family within the core microbiota, unlike the previous study.[29] There was also a 15-year average age difference between the 2 studies (52 vs 37 y). We also identified novel core *Proteobacteria*, including a *Bilophila* (*Desulfovibrionaceae* family), and an unclassified *Burkholderiales*.

Previous reports have indicated reduced abundance of *F prausnitzii* in CD patients;[6,32] however, here we show that *F prausnitzii* only represents one of many core members of the microbiota that are less abundant in ICD. *Roseburia*, which along with *Faecalibacterium* is a major butyrate producer in the gut,[33] represented 3 of the core OTUs that were less abundant in ICD. Although *F prausnitzii* has been shown to have therapeutic properties in IBD models,[11] our data indicate that replenishment of additional core members including *Roseburia*, *Alistipes*, *Collinsella*, and members of the *Ruminococcaceae* family may be warranted. The fact that many divergent genera from different phyla showed a similar pattern of reduced abundance in ICD questions what these bacteria share in common. A recent study of metabolites present in fecal water extracts from some of the same samples studied here also revealed a dramatic shift in metabolic profiles of patients with ICD compared with healthy patients.[34] Pathways involved in metabolism of amino acids, fatty acids, and bile acids were different in ICD, and some of

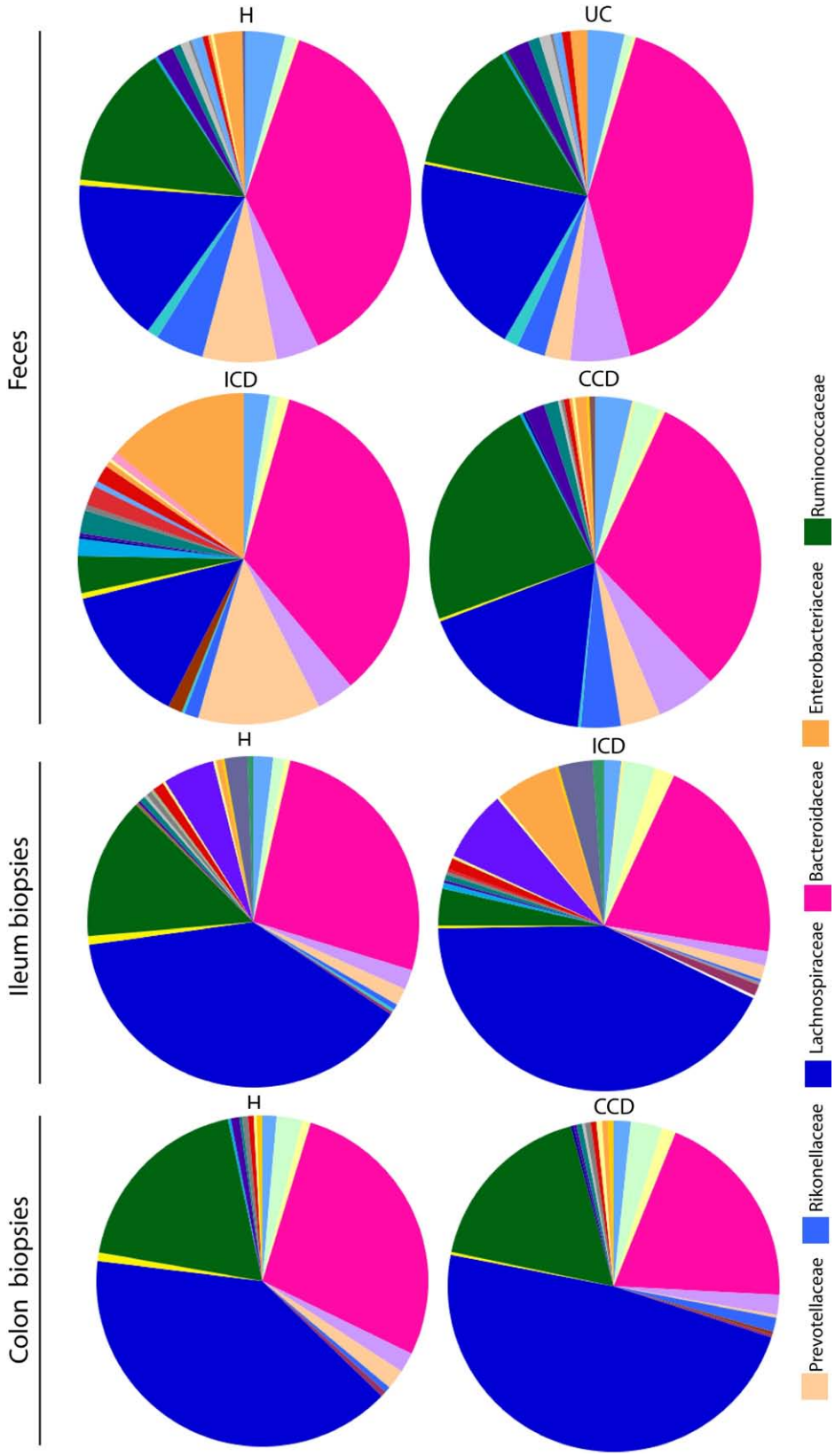
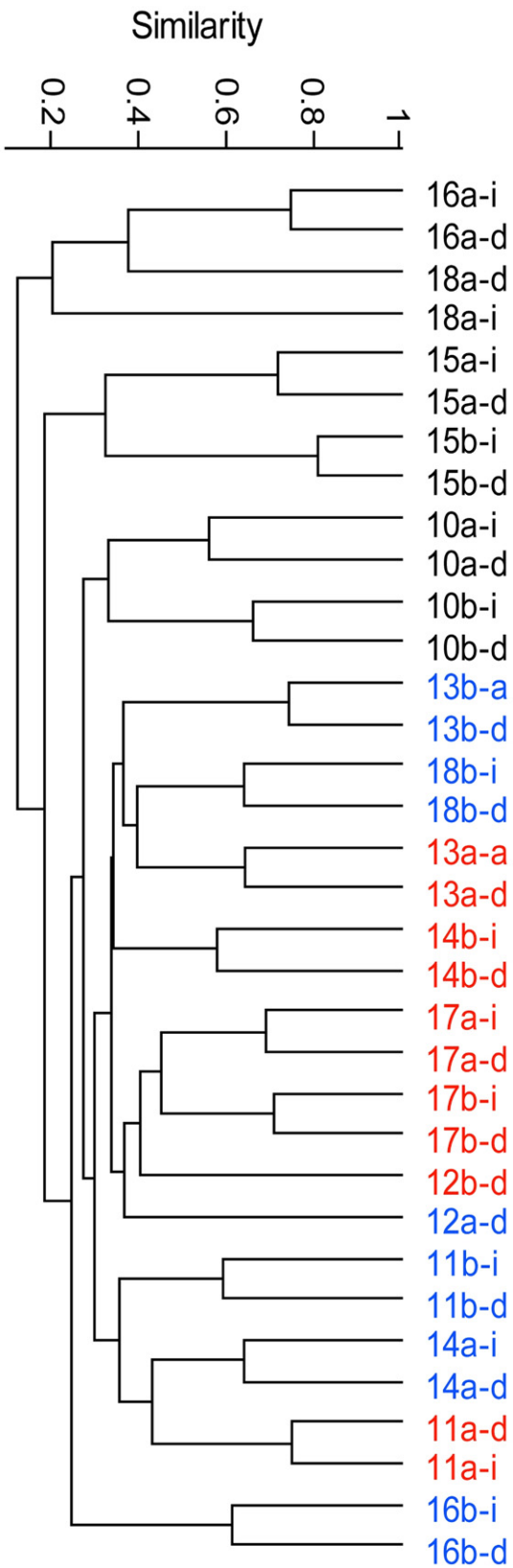


Figure 4. Average abundance of bacterial families from patients with CCD, ICD, UC, and healthy subjects in fecal samples and biopsy specimens from the ileum and descending colon.



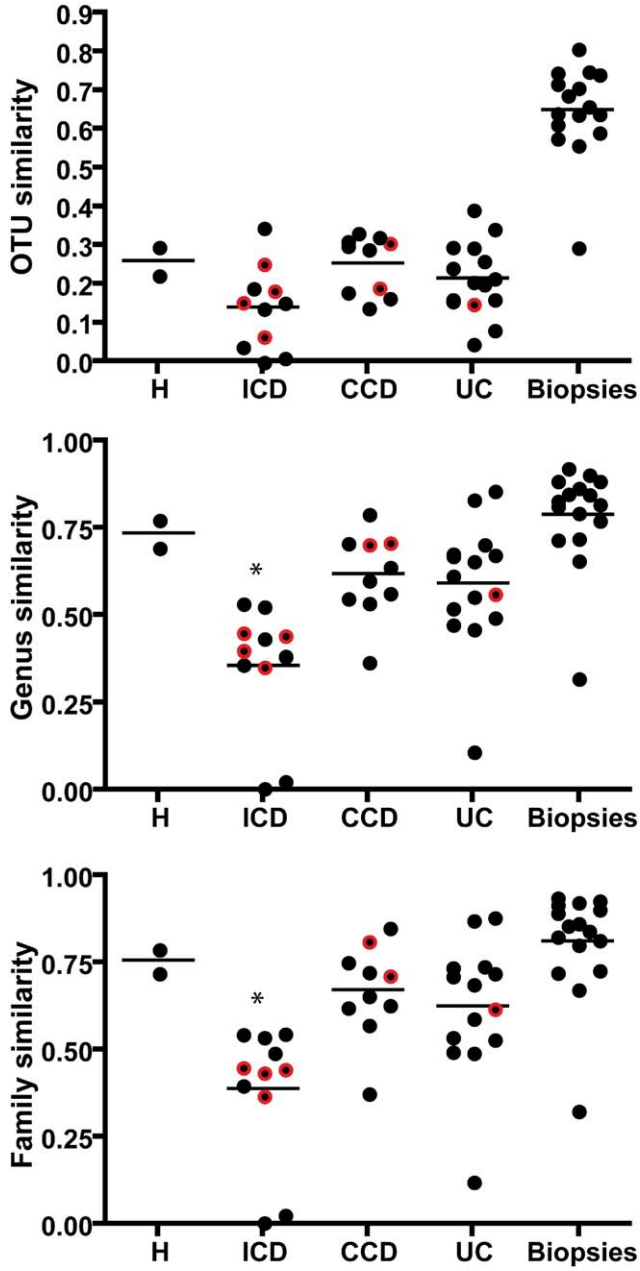


Figure 6. Similarity between microbial composition of concordant (*red*) and discordant (*black*) twin pairs using Bray–Curtis metrics at OTU (95% identity), genus and family classification levels in fecal samples from patients with CCD, ICD, UC, and healthy patients. Similarities between biopsy specimens within an individual collected from the ileum or ascending colon and descending colon also were compared. *Significantly different from all other phenotypes ($P < .05$).

Figure 5. Similarities between ileal and colonic mucosal microbiota. Cluster analysis using Bray–Curtis metrics of microbial profiles (OTU set at 5% dissimilarity) of biopsy samples from the ileum (i), ascending colon (a), and descending colon (d) of healthy subjects (*blue*), patients with ICD (*black*), or patients with CCD (*red*). Twin pairs share a common number.

these pathways correspond to the disappearance of specific members of the bacterial community, including *F prausnitzii*.[34]

The fecal microbiota could not be differentiated between patients with active disease and those in remission. Although some studies have shown that microbial diversity differs in inflamed and noninflamed biopsy tissues, no specific differences in bacterial composition were identified.[35,36] It also has been shown that the depletion of *F prausnitzii* in active CD is the result of the host immune response because high-dose cortisol treatment results in re-establishment of *F prausnitzii* abundance to normal levels.[32] It is yet unclear what immune mechanism is responsible for this shift or whether other bacteria also are suppressed by the host immune response and warrant further investigation.

Although ICD generally was distinguished by a reduction of bacterial populations, there were a few exceptions, including increased abundance of *Gammaproteobacteria*, primarily OTUs similar to *E coli*. In some ICD patients the representation of *Enterobacteriaceae* was as high as 86% of the community. Extreme levels of *Gammaproteobacteria* also have been observed in cases of inflammation with Salmonella,[37] which intriguingly also results in a Th1/Th17 response and deep mucosal damage similar to that seen in CD.[38] The increased abundance of several OTUs representative of *Enterobacteriaceae* in ICD supports the theory that these bacteria take advantage of inflammation to increase colonization levels.[39] Although there have been reports of *E coli* with pathogenic properties including adherence and invasion in Crohn patients,[40] the *E coli* in this study were not isolated and remain uncharacterized.

The only highly significant increase of a single OTU in individuals with ICD was related closely to *Ruminococcus gnavus*. Although this difference was observed in fecal samples, the difference was much larger in biopsy specimens, representing in some cases more than 20% of bacteria in both ileal and colonic biopsy specimens in ICD patients. There has been one other report of increased *R gnavus* levels in a subset of CD patients.[41] Intriguingly, an *R gnavus* strain has been reported previously to produce a lantibiotic with bacteriocidal activity against some clostridia and bifidobacteria species.[42] The increased abundance of this organism may therefore explain the reduced abundance of *Actinobacteria* and *Ruminococcaceae*. Further investigation of this interaction is warranted. Although the difference in *R gnavus* was more pronounced in biopsy samples, there were inadequate biopsy numbers of discordant disease location phenotypes to properly investigate differences in these regions.

One of the most striking observations was that ICD and CCD phenotypes showed microbial profiles that had opposite trends, particularly the case for the *Bifidobacteriaceae*, *Coriobacteriaceae*, and *Ruminococcaceae* families, questioning how these 2 disease phenotypes differ beyond localization. The striking difference between these 2 disease phenotypes may explain opposing reports of *Actinobacteria* abundance in CD.[2,3]

It was not possible to create a model to discriminate UC patients from healthy patients and only a few differences could be seen at the OTU level. Differences between UC and healthy patients previously have been observed in patients with active disease.[43,44] By contrast, the UC patients in this study were in remission, suggesting that the microbiota returns to a more normal state during remission.

In addition to improving our understanding of disease etiology, our aim was to find bacterial signatures as diagnostic and monitoring tools for IBD. There are a number of limitations to available diagnostic kits: accurate diagnosis is limited to active disease, some are invasive requiring blood or biopsy collection, and none stands alone and therefore is used in conjunction and as a supplement to endoscopy.[45] Therefore, it was encouraging that the predictive model that we generated successfully discriminated CD disease phenotypes based on fecal microbial composition. The ability to differentiate IBD patients who were in remission together with those having active disease indicates that profiling the microbiota may be a useful tool to aid in screening and diagnosis.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at doi: 10.1053/j.gastro.2010.08.049.

References

1. Sartor RB. Microbial influences in inflammatory bowel diseases. Gastroenterology 2008;134:577–594.
2. Frank DN, St Amand AL, Feldman RA, et al. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci U S A 2007;104:13780–13785.
3. Sokol H, Lay C, Seksik P, et al. Analysis of bacterial bowel communities of IBD patients: what has it revealed? Inflamm Bowel Dis 2008;14:858–867.
4. Sokol H, Seksik P, Rigottier-Gois L, et al. Specificities of the fecal microbiota in inflammatory bowel disease. Inflamm Bowel Dis 2006;12:106–111.
5. Darfeuille-Michaud A, Neut C, Lederman E, et al. Virulence factors of *Escherichia coli* strains isolated from ileal mucosa in Crohn's disease. Gastroenterology 1998;114:A958–A959.
6. Willing B, Halfvarson J, Dicksved J, et al. Twin studies reveal specific imbalances in the mucosa-associated microbiota of patients with ileal Crohn's disease. Inflamm Bowel Dis 2009;15:653–660.
7. Scanlan PD, Shanahan F, O'Mahony C, et al. Culture-independent analyses of temporal variation of the dominant fecal microbiota and targeted bacterial subgroups in Crohn's disease. J Clin Microbiol 2006;44:3980–3988.
8. Martinez-Medina M, Aldeguer X, Gonzalez-Huix F, et al. Abnormal microbiota composition in the ileocolonic mucosa of Crohn's disease patients as revealed by polymerase chain reaction-denaturing gradient gel electrophoresis. Inflamm Bowel Dis 2006;12:1136–1145.

9. Manichanh C, Rigottier-Gois L, Bonnaud E, et al. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. Gut 2006;55:205–211.
10. Baumgart M, Dogan B, Rishniw M, et al. Culture independent analysis of ileal mucosa reveals a selective increase in invasive *Escherichia coli* of novel phylogeny relative to depletion of *Clostridiales* in Crohn's disease involving the ileum. ISME J 2007;1:403–418.
11. Sokol H, Pigneur B, Watterlot L, et al. *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn's disease patients. Proc Natl Acad Sci U S A 2008;105:16731–16736.
12. Swidsinski A, Weber J, Loening-Baucke V, et al. Spatial organization and composition of the mucosal flora in patients with inflammatory bowel disease. J Clin Microbiol 2005;43:3380–3389.
13. Von Stein P, Lofberg R, Kuznetsov NV, et al. Multigene analysis can discriminate between ulcerative colitis, Crohn's disease, and irritable bowel syndrome. Gastroenterology 2008;134:1869–1881.
14. Solberg IC, Lygren I, Jahnsen J, et al. Clinical course during the first 10 years of ulcerative colitis: results from a population-based inception cohort. Scand J Gastroenterol 2009;44:431–440.
15. Rahier JF, Ben-Horin S, Chowers Y, et al. European evidence-based consensus on the prevention, diagnosis and management of opportunistic infections in inflammatory bowel disease. J Crohns Colitis 2009;3:47–91.
16. Andersson AF, Lindberg M, Jakobsson H, et al. Comparative analysis of human gut microbiota by barcoded pyrosequencing. PLoS One 2008;3:e2836.
17. Sogin ML, Morrison HG, Huber JA, et al. Microbial diversity in the deep sea and the underexplored "rare biosphere". Prod Natl Acad Sci U S A 2006;103:12115–12120.
18. Costello EK, Lauber CL, Hamady M, et al. Bacterial community variation in human body habitats across space and time. Science 2009;326:1694–1697.
19. Dethlefsen L, Huse S, Sogin ML, et al. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. PLoS Biol 2008;6:2383–2400.
20. Dicksved J, Halfvarson J, Rosenquist M, et al. Molecular analysis of the gut microbiota of identical twins with Crohn's disease. ISME J 2008;2:716–727.
21. Halfvarson J, Bodin L, Tysk C, et al. Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics. Gastroenterology 2003;124:1767–1773.
22. Lennardjones JE. Classification of inflammatory bowel-disease. Scand J Gastroenterol 1989;24:2–6.
23. Silverberg MS, Satsangi J, Ahmad T, et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a working party of the 2005 Montreal World Congress of Gastroenterology. Can J Gastroenterol 2005;19:5a–36a.
24. Schroeder KW, Tremaine WJ, Ilstrup DM. Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. A randomized study. N Engl J Med 1987;317:1625–1629.
25. Harvey RF, Bradshaw JM. A simple index of Crohn's-disease activity. Lancet 1980;1:514.
26. Lambert, D. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics 1992;34:1–14.
27. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc Ser B 1995;57:289–300.
28. Begon M, Harper J, Townsend C. Ecology: from individuals to ecosystems. 4th ed. Oxford, UK: Blackwell, 2006.
29. Tap J, Mondot S, Levenez F, et al. Towards the human intestinal microbiota phylogenetic core. Environ Microbiol 2009;11:2574–2584.
30. Zoetendal EG, Akkermans ADL, Akkermans-van Vliet WM, et al. The host genotype affects the bacterial community in the human gastrointestinal tract. Microbial Ecol Health Dis 2001;13:129–134.
31. Turnbaugh PJ, Hamady M, Yatsunenko T, et al. A core gut microbiome in obese and lean twins. Nature 2009;457:480–484.
32. Swidsinski A, Loening-Baucke V, Vaneechoutte M, et al. Active Crohn's disease and ulcerative colitis can be specifically diagnosed and monitored based on the biostructure of the fecal flora. Inflamm Bowel Dis 2008;14:147–161.
33. Louis P, Flint HJ. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. FEMS Microbiol Lett 2009;294:1–8.
34. Jansson J, Willing B, Lucio M, et al. Metabolomics reveals metabolic biomarkers of Crohn's disease. PLoS One 2009;4:e6386.
35. Seksik P, Lepage P, de la Cochetiere MF, et al. Search for localized dysbiosis in Crohn's disease ulcerations by temporal temperature gradient gel electrophoresis of 16S rRNA. J Clin Microbiol 2005;43:4654–4658.
36. Sepehri S, Kotlowski R, Bernstein CN, et al. Microbial diversity of inflamed and noninflamed gut biopsy tissues in inflammatory bowel disease. Inflamm Bowel Dis 2007;13:675–683.
37. Sekirov I, Tam NM, Jogova M, et al. Antibiotic-induced perturbations of the intestinal microbiota alter host susceptibility to enteric infection. Infect Immun 2008;76:4726–4736.
38. Grassl GA, Valdez Y, Bergstrom KSB, et al. Chronic enteric *Salmonella* infection in mice leads to severe and persistent intestinal fibrosis. Gastroenterology 2008;134:768–780.
39. Lupp C, Robertson ML, Wickham ME, et al. Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of *Enterobacteriaceae*. Cell Host Microbe 2007;2:119–129.
40. Darfeuille-Michaud A, Boudeau J, Bulois P, et al. High prevalence of adherent-invasive *Escherichia coli* associated with ileal mucosa in Crohn's disease. Gastroenterology 2004;127:412–421.
41. Prindiville T, Cantrell M, Wilson KH. Ribosomal DNA sequence analysis of mucosa-associated bacteria in Crohn's disease. Inflamm Bowel Dis 2004;10:824–833.
42. Dabard J, Bridonneau C, Phillipe C, et al. A new antibiotic produced by a *Ruminococcus gnavus* strain isolated from human feces. Appl Environ Microbiol 2001;67:4111–4118.
43. Mylonaki M, Rayment NB, Rampton DS, et al. Molecular characterization of rectal mucosa-associated bacterial flora in inflammatory bowel disease. Inflamm Bowel Dis 2005;11:481–487.
44. Bullock NR, Booth JC, Gibson GR. Comparative composition of bacteria in the human intestinal microflora during remission and active ulcerative colitis. Curr Issues Intest Microbiol 2004;5:59–64.
45. Li XH, Conklin L, Alex P. New serological biomarkers of inflammatory bowel disease. World J Gastroenterol 2008;14:5115–5124.

Received November 26, 2009. Accepted August 26, 2010.

Reprint requests

Address requests for reprints to: Lars Engstrand, MD, Department of Bacteriology, Swedish Institute for Infectious Disease Control, 17177 Solna, Sweden. e-mail: lars.engstrand@smi.se

Conflicts of interest

The authors disclose no conflicts.

Funding

This project was funded by the Söderbergs Foundation, the Ekhaga Foundation, the Örebro University Hospital Research Foundation, the Örebro County Research Foundation, the Bengt Ihre Foundation, the Uppsala BioX Micprof project, the Swedish University of Agricultural Sciences, and in part by the US Department of Energy under contract no. DE-AC02-05CH11231 with the Lawrence Berkeley National Laboratory.

Supplementary Materials and Methods

PCR Conditions and PCR Purification

Two microliters of DNA were added to each 25-μL PCR reaction containing 2.5 μL 10$\times$ PCR buffer (New England Biolabs, Frankfurt, Germany), 1 μL bovine serum albumin (10 mg mL^{-1}; Amersham Biosciences, Piscataway, NJ), 1 μL deoxynucleoside triphosphate (5 mmol/L), 0.25 μL Taq polymerase (5 U μL^{-1}; Amersham Biosciences), and 1 μL of each primer (10 μmol/L). PCR reactions were performed on a GeneAmp (Applied Biosystems, Foster City, CA) PCR system (5 min at 94°C, 30 cycles of 94°C for 45 s, 55°C for 40 s, and 72 for 1 min, and a final extension of 72°C for 7 min). Triplicate PCRs were pooled and 50 μL were run on 1% agarose gels at 80 V for 1.5 hours. PCR products of the appropriate size (approximately 340 base pairs) were gel purified (QIAquick Gel Extraction Kit; Qiagen, Inc, Hilden, Germany) and eluted in 50 μL of elution buffer. DNA quality was assessed on a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). DNA concentration was measured on a NanoDrop ND-1000 (NanoDrop Technologies, Wilmington, DE) and 25 ng of 4 samples, labeled with different tag sequences, were pooled and diluted in water for a total of 100 ng in 10 μL.