Highlighting New Phylogenetic Specificities of Crohn's Disease Microbiota

S. Mondot, PhD,* S. Kang, PhD,[†] J.P. Furet, MS,* D. Aguirre de Carcer, PhD,[†] C. McSweeney, PhD,[†] M. Morrison, PhD,[†] P. Marteau, PhD,[‡] J. Doré, PhD,* and M. Leclerc, PhD*

Background: Recent studies suggest that gastrointestinal (GI) microbes play a part in the pathogenesis of Crohn's disease (CD).

Methods: Fecal samples were collected from 16 healthy individuals and 16 CD patients (age- and sex-matched). The DNA extracted from these samples were subjected to two different methods of microbiome analysis. Specific bacterial groups were quantified by real-time polymerase chain reaction (PCR) methods using primers designed using a high-throughput in-house bioinformatics pipeline. The same DNA extracts were also used to produce fluorescently labeled cRNA amplicons to interrogate a custom-designed phylogenetic microarray for intestinal bacteria.

Results: Even though the intersubject variability was high, differences in the fecal microbiomes of healthy and CD patients were detected. *Faecalibacterium prausnitzii* and *Escherichia coli* were more represented in healthy and ileal CD patients, respectively. Additionally, probes specific for *Ruminococcus bromii*, *Oscillibacter valericigenes*, *Bifidobacterium bifidum*, and *Eubacterium rectale* produced stronger hybridization signals with the DNA samples from healthy subjects. Conversely, species overrepresented in CD patients were *E. coli*, *Enterococcus faecium*, and species from the Proteobacteria not normally found in the healthy human GI tract. Furthermore, we detected "healthy specific" molecular species or operational taxonomic units (OTUs) that are not closely related to any known species (*Faecalibacterium, Subdoligranulum*, and *Oscillospora* species), indicating that the phylogenetic dysbiosis is broader than at strain or species level.

Conclusions: These two techniques of microbiome analysis provided a statistically robust new picture of the dysbiosis in fecal microbiota from ileal CD patients. Specifically, we identified a set of six species discriminant for CD, which provides a preliminary diagnostic tool.

(*Inflamm Bowel Dis* 2011;17:185–192)

Crohn's disease (CD) is a multifactorial disease of the human digestive tract with a still unclear etiology. A genome-wide association study recently highlighted more than 30 distinct loci involved in CD pathogenesis.[1] Most of the human genes encountered in these human genomic loci were components of the autophagy pathway and innate immune system.[1] However, host genetic predisposition does not alone explain CD triggers, remission, and relapse. Environmental factors such as smoking,[2,3] appendectomy,[4] or endogenous microbiota are also involved in the onset of the disease. During the last decade, evidence implicating the host microbiota in CD has accumulated and it is now well established that changes in the gut microbiota are an important environmental component of inflammatory bowel diseases (IBDs).[5] Recent studies using rodent models have pinpointed the intricate dialog established between the host and the gut microbiota to maintain a mutual environment within the gut. For instance, Bouskra et al[6] demonstrated that the lymphoid tissue maturation and proliferation was driven by commensal gut bacteria through the NOD1 innate receptor. Additionally, the detection of enteric bacteria by Paneth cells through a MyD88-mediated activation protected host tissue against invasion by pathogenic and commensal bacteria.[7] Finally, Garrett et al[8] reported the existence of a communicable colitogenic microbiota shaped by T-bet cell deficiency in mice which was able to reinduce colitis in a genetically intact host. Taken together, these results indicate that the gastrointestinal tract is a strikingly complex ecosystem and its homeostasis requires finely balanced regulation and crosstalk between host cells and the resident microbiota. In humans, the gut microbiota is composed of trillions of microorganisms exceeding by a factor of 10 the total number of host cells,[9] with several thousands of different species detected.[10] A number of strategies have been used to examine the role of the intestinal microbiota in CD initiation and/or maintenance. First of all, the search for persistent pathogenic agents led to the detection of both *Mycobacterium avium paratuberculosis* (MAP)[11] and the *Escherichia coli*

TABLE 1. From Sequencing to Validation: Characteristics of the Subjects Included in the Study

	Laboratory 16S rDNA Molecular Inventories				Phylogenetic Specifities Validation	
	CPLII	ESL	CPL*	HSL*	CD Control Group	Healthy Control Group
n	2	8	6	6	16	16
Clinical status	CD	Healthy	CD	Healthy	CD	Healthy
Female	2	4	2	5	8	8
Mean age (SD)	37 (29–45)	78 (70–87)	30 (18–44)	32 (24–43)	31 (25–39)	30 (25–35)
Surgery		—		—		—
Disease location:						
Duodenum		—	1	—	0	—
Ileum		—	6	—	16	—
Colon		—	4	—	0	—
Anal		—	2	—	0	—
Treatment:						
Aminosalicylates		—	2	—	NA	—
Noninclusion criteria						
					Pregnancy	
					Colic resection	
					Antibiotherapy	
					Intestinal resection in the 5 later years	

SD, standard deviation.
*Samples from these subjects were also used for phylogenetic microarrays validation (Ref. 27).

AIEC pathovar[12] as potential infectious agents. However, these microorganisms are not always detected in CD patients,[5] suggesting other microbes might also play a role in CD. The non a priori approach of studying the gastrointestinal (GI) tract microbiota using DNA sequencing techniques has enabled the recovery of microbial diversity without the need for bacterial cultivation[13,14] and has provided some new links between gut microbiota composition and IBD.[15] Dysbiosis, defined as an imbalance of the gut microbial community,[16] has since been characterized by an alteration in the balance between the Bacteroidetes and Firmicutes phyla.[15,17] More precisely, the microbiota of CD patients is characterized by a decreased diversity within the *Clostridium leptum* group, with one species, *Faecalibacterium prausnitzii,* described as antiinflammatory.[18] In addition, increased numbers of the Proteobacteria and Actinobacteria phyla have also been observed in these molecular inventories.[10] However, there is a need to develop a finer resolution characterization of the gut microbiota in CD patients for a more precise measurement of health and disease.

The aim of this study was to provide new insights into CD-associated dysbiosis by combining molecular approaches through a "global strategy." First, a bioinformatics approach was developed and applied to investigate, without a priori, the existing microbial 16S rRNA gene (*rrs*) molecular inventories produced from earlier case–control studies of CD. The in silico predictions were then validated using real time quantitative polymerase chain reaction (RT Q-PCR) to quantify these molecular species in fecal DNA samples from a well-characterized cohort of age- and sex-matched ileal CD patients and healthy volunteers. The RT Q-PCR results were further validated and augmented using a custom-designed phylogenetic microarray for intestinal bacteria.

MATERIALS AND METHODS

Subjects and Samples

With laboratory molecular studies used to perform the in silico predictions and validation steps, a total of 54 subjects were included (Table 1) comprising 24 CD and 30 healthy individuals. The 16 CD patients used as the CD control group (CCG) were part of a previously published double-blind controlled trial.[19] These patients suffered from ileal CD requiring surgery ileal or ileocecal resection. The healthy control group (HCG) was composed of 16 individuals selected for both age and sex ratio concordance with the CCG. Fecal samples were handled and DNA extracted as previously described.[17] All DNA samples were checked by agarose gel electrophoresis (1% w/v) stained with ethidium bromide, and also quantified using a Nanodrop ND-1000 spectrophotometer (Labtech, Palaiseau, France).

Sequence Datasets

A total of 3302 sequences recovered from four different *rrs* gene libraries produced by our group were analyzed, two of which were metagenomic and previously
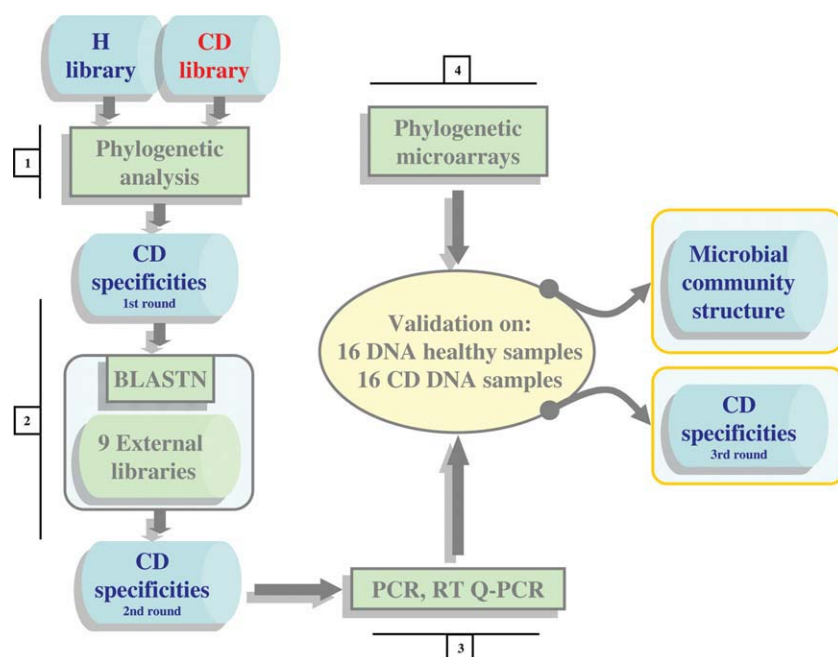
FIGURE 1. 1) CD and healthy 16S rRNA gene libraries from our laboratory datasets were analyzed by a phylogenetic approach. OTUs sequences retrieved from healthy or CD libraries or common to both were selected. 2) CD specificities sequences were then compared to nine external 16S rRNA gene libraries and a second round of OTU selection was applied. 3) Candidate OTUs were validated on 16 CD and 16 healthy fecal DNA samples using PCR and RT PCR. 4) In parallel, in the same samples, a phylogenetic microarray approach was applied to validate CD specificities and analyze the microbial community structure of CD microbiota.

published.[15] The other two libraries were obtained by PCR amplification of *rrs* genes using DNA from the Elderly Subject Library (ESL) and the CD Patient Library II (CPLII) as templates. The sequence data produced from these two sources are publicly available (Supporting Table 1). We also included 45,076 *rrs* gene sequences recovered from nine libraries produced by other investigators studying IBD, and deposited with the NCBI database prior to November 2007.

rrs Gene Sequence Analysis and Phylogenetic (OTU) Specificity Searches

A two-stage strategy was used for the in silico analysis of the *rrs* gene sequences (Fig. 1). First, the four libraries produced by our laboratory were analyzed to identify specific operational taxonomic units (OTUs) either associated with the "CD" or "healthy" datasets, or shared. Sequences were aligned using the MAFFT program[20] and distance matrices were computed with the DNADIST program. Chimeras and sequence anomalies were removed using the Mallard program.[21] An OTU was defined as sequences sharing at least 98% similarity. OTU detection and ecological estimates were determined using the DOTUR program.[22] OTUs were taxonomically assigned with Sequence Match application (RDP-II, release 9). The genetic diversity of CD patients and healthy subject libra-

ries were compared using SONS program, enabling the inference of CD specificities.[23] A CD specificity was defined based on OTU sequences preferentially retrieved in either healthy or CD libraries or common to both. Then CD specificities were refined by measuring, by BLASTN analysis, their prevalence in nine 16S rDNA sequences libraries from external datasets (Supporting Table 1). The resulting selected OTU sequences were finally used for RT Q-PCR primers design using the PRISE program[24] (Table 2). These OTUs were then searched for within the datasets downloaded from NCBI (Supporting Table 1). Conventional PCR and RT Q-PCR were then used to assess the detection rate of each candidate OTU among the 32 DNA samples (Table 1). Those OTUs either not detected in the DNA samples or with no correlation with subject status were withdrawn as uninformative. On the basis of this analysis, six informative OTUs were identified.

Primer Specificity and PCR-based Assays

Primer specificities were assessed using Probe-Check[25] and tested by RT Q-PCR on fosmid/plasmid from which the sequence originated, as well as cloned *rrs* gene amplicons produced from the cultured type strains. PCR was carried out using a PTC-100 thermocycler (MJ Research, Reno, NV) with HotStarTaq Master Mix (Qiagen, Courtaboeuf, France), following the conditions: 95°C

TABLE 2. Primer Sets Used in the Study

Target OTU	Primer	5′-3′ Sequence	Sources
OTUc03	hc03-F	CTCCTGTCGTTAGGGACGATA	This study
	hc03-R	TCTGCACTACTCAAGGCCAG	
OTUh06	ch06-F	GGTAAAGCTGAGGACTCTGGTGG	This study
	ch06-R	GTGGTTTGGGGATTCGCTC	
OTUe10	he10-F	GGTCTTGACATCCAACTAACGAAGT	This study
	he10-R	TTTTGTCAACGGCAGTCCTAT	
OTUa08	ha08-F	TCTTTTGTCAGGGACGAGTAGAA	This study
	ha08-R	GCCTCTCCGATACTCAAGAAGAGT	
OTUa05	B. vulg 01	CCCGGGCTTAAATTGCAGA	Furet et al (26)
	B. vulg 02	GCGGGACCTGTTAGTAACTGACAA	
OTUe11	E. coli F	CATGCCGCGTGTATGAAGAA	Huijsdens et al (34)
	E. coli R	CGGGTAACGTCAATGAGCAAA	

for 10 minutes and 25 cycles of amplification (95°C for 15 sec, 60°C for 1 min, 72°C for 30 sec) followed by a final elongation step at 72°C for 10 minutes. The RT Q-PCR was carried out as previously described[26] using an ABI Prism 7700 (Applied Biosystems, Courtaboeuf, France) using SYBR Green Master Mix (Applied Biosystems). The results for the RT Q-PCR primer specificity tests for uncultured *Subdoligranulum sp.* (OTUc03), *Parabacteroides distasonis*, (OTUh06), *Ruminococcus bromii* (OTUe10), *Oscillibacter valericigenes* (OTUa08), *Bacteroides vulgatus* (OTUa05), *E. coli* (OTUe11) are shown in Supporting Table 2.

Bacterial Strains, Fosmids/Plasmids-bearing *E. coli* DH10B, Culture Conditions, and DNA Extraction

The anaerobic and aerobic bacterial strains used were: *P. distasonis* DSM20701T, *Fusobacterium nucleatum* DSM19508T, *Fusobacterium naviforme* DSM20699, *Prevotella copri* 18205T, *Prevotella ruminicola* ATCC19189T, *O. valericigenes* DSM18026T, *Ruminococcus bromii* 51896T, *B. vulgatus* ATCC8482, *E. coli* UEPSD S123 and DH10B. Detailed culture media and growth conditions are provided in the supplemental information data (Supporting Text).

Microarray

The 32 DNA samples were also used to prepare fluorescently labeled cRNA fragments of the *rrs* gene that were hybridized to a custom-designed phylogenetic microarray. A previous description and validation of the microarray are reported in Kang et al.[27]

Statistical Analysis

The chi-square test was used to test selected OTU independence toward CD status. All statistics were computed using R software and both ade4[28] and MADE4[29] packages. A principal coordinate analysis (PCA) was performed on the full distance matrix to map the overall genetic diversity retrieved in the laboratory dataset and point out CD specificities. A strip chart plot was performed on RT Q-PCR data to map CD and healthy microbiotas according to CD specificities. Medians were computed for each group using OTU Cycle threshold (Ct) values, and Wilcoxon's test was applied to RT Q-PCR data to assess significant differences between CD and healthy groups. Similarity between microbiotas according to their microarray profiles was assessed by a correspondence analysis (CoA). A "between-class" analysis was applied on the whole CoA dataset. A heatplot map of the first coordinate between class analysis was projected to highlight discriminating probes between healthy versus CD groups.

RESULTS

CD Microbiota Imbalances Revealed

The in silico analyses identified a CD specificity, defined as an OTU preferentially retrieved in either healthy or CD libraries, or common to both. These OTUs are associated with uncultured *Subdoligranulum sp.* (OTUc03), *P. distasonis* (OTUh06), *R. bromii* (OTUe10), *O. valericigenes* (OTUa08), *B. vulgatus* (OTUa05), and *E. coli* (OTUe11; Supporting Table 2; Supporting Fig. 1). Except for OTUc03, the OTU sequences were 99% similar to the 16S *rrs* gene sequence of the closest strain.

The OTUs found to be most depleted in CD patients (Fig. 2) by both in silico and the RT Q-PCR analyses belonged principally to the *C. leptum* group and to a lesser extent within the Bacteroidetes phylum. Although the "unknown" OTUc03 was detected in most of the fecal samples (68.75% and 100% of CD and HCG, respectively) the
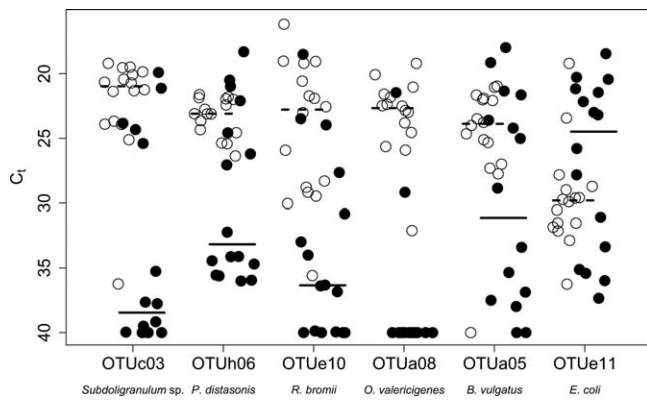
FIGURE 2. Strip chart diagram plot of the detected Ct obtained by RT Q-PCR targeting CD specificities in 32 DNA samples. Full and open circles correspond to CD ($n = 16$) and healthy samples ($n = 16$), respectively. Dashed and solid lines are median values computed for healthy and CD Ct data, respectively.

RT Q-PCR analyses showed bacteria assigned to this OTU were present in significantly greater abundance in HCG than in CCG patients ($\Delta Ct = 17.48$; $P < 0.001$) (Fig. 2). The *R. bromii* OTUe10 was also detected in most of the samples (75% of CCG and 100% of HCG samples) but RT Q-PCR suggested that these bacteria were more abundant in the HCG than CCG group ($\Delta Ct = 13,56$; $P < 0.001$). However, the most striking difference was observed for *O. valericigenes* (OTUa08), which was detected in only 12.5% of the CCG patients but in almost all (93.75%) of the HCG patients. Additionally, its abundance was significantly higher in healthy than in CCG ($\Delta Ct = 17.33$, $P < 0.001$) (Fig. 2).

Based on the in silico predictions and RT Q-PCR data, *P. distasonis* (OTUh06) was detected in all 32 samples, but the abundance of bacteria assigned to this OTU was significantly higher in the healthy than in the CCG ($\Delta Ct = 10.07$; $P < 0.05$). The *B. vulgatus* OTU (a05) was detected in most of the fecal samples (81.25% of CCG and in 93.75% of HCG groups), but the RT Q-PCR suggested there were no significant differences in the abundance of these bacteria between the two groups ($\Delta Ct = 7.24$, $P > 0.05$). Finally, the *E. coli* OTU (e11) was detected in all 32 samples and there were no significant differences between the two groups in terms of its abundance ($\Delta Ct = 5.32$, $P > 0.05$).

Collectively, these results indicated that some commensal gut bacterial inhabitants of healthy microbiota were depleted from CD samples. Moreover, as previously shown, *E. coli* was found to be overrepresented in some CD microbiota.

More Individual Variability in Microbiota from Patients with Active CD than in Healthy Individuals

The RT Q-PCR analyses suggested that in addition to the differences outlined above, the CD patients showed a
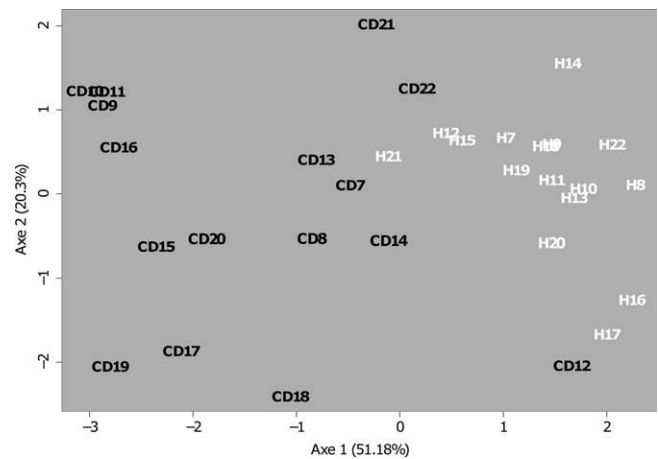

FIGURE 3. PCA was computed using the RT Q-PCR Ct values obtained for the six candidate OTUs. The two first components (principal components 1 and 2) were plotted and represented 71.48% of whole inertia. Individual microbiota, CD (black, $n = 16$) and healthy (white, $n = 16$) were scattered according to their RT Q-PCR profiles (Fig. 2).

greater degree of variance in the abundance of these OTUs, compared to healthy individuals (Fig. 3), suggestive of greater intraindividual variations among CD patients. The chi-square test of the RT Q-PCR median distributions validated the observed topology, and indicated that all selected OTUs were not independent of the clinical status of the individuals. According to both PCA axes 1 and 2, which accounted for 71.48% of the variance, the fecal microbiota in CD patients appeared more scattered than those in healthy subjects. Interestingly, two fecal samples from healthy subjects (H17 and H21) were dispersed and closer to the CD datasets, while the fecal samples from two CD
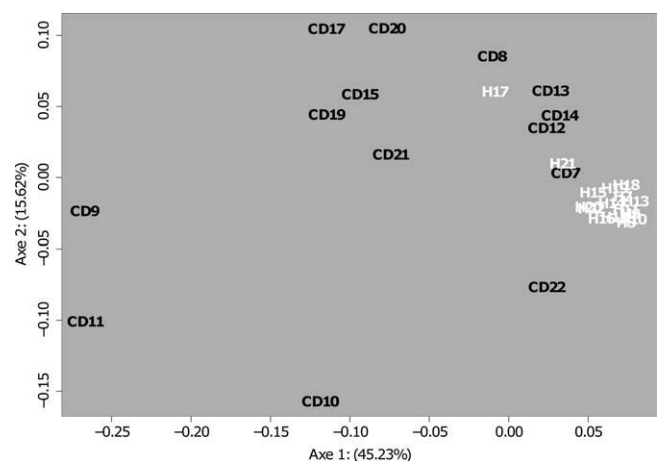

FIGURE 4. CoA was computed on the whole phylogenetic microarray dataset and both CoA axes 1 and 2 were plotted which corresponded to 60.85% of whole inertia. CD (black, $n = 14$) and healthy (white, $n = 16$) microbiota were mapped according to their phylogenetic profiles.

patients (CD12 and CD22) appeared most similar to the profiles obtained from healthy individuals. Such observations suggest that patients considered otherwise "healthy" or "diseased" might actually be developing a microbiota that is "at risk" or "remedial" for CD, respectively. Finally, based on the PCA axis 1, the sensibility and specificity for detecting CD status were 93.75% and 87.5%, respectively.

Microarrays Confirmed the Discrimination Between CD and Healthy Microbiota

The same 32 fecal samples were used to produce fluorescently labeled cRNA probes and hybridized to a phylogenetic microarray specific for gut bacteria.[27] Correspondence analysis of the microarray data showed a tightly clustering schema for the profiles produced from most of the samples from healthy microbiota when PCA axes 1 and 2, accounting for 60.85% of the variance, were considered (Fig. 4). Conversely, the samples from most of the CD patients demonstrated a large distribution spectrum, which was in accordance with RT Q-PCR data. Similarly, it was noticeable that samples H17 and H21 from HCG were again closest to the CD samples and CD12 and CD22 samples were more closely associated with the healthy samples. Furthermore, the co-inertia analysis supported the co-structure produced from both the RT Q-PCR and microarray datasets (data not shown). Taken together, the healthy individuals and the patients suffering from an ileal CD were strongly discriminated according to their GI tract microbiota composition. However, whatever the method used the dysbiosis was not absolute.

Further Insights into CD Dysbiosis as Revealed by Microarray Analysis

Although the in silico and RT Q-PCR data provided insight into the fluctuations and abundance of six bacterial OTUs in healthy and CD patients, the microarray analyses provided additional and new insights. The discriminating probes, set as 5-fold different from the median between HCG and CCG (Fig. 5) revealed that *F. prausnitzii, R. bromii, Eubacterium rectale, Ruminococcus albus, Bifidobacterium bifidum, P. distasonis*, and *Alistipes putredinis* were more abundant in healthy than in CD microbiota. Very consistently, the OTUs predicted by the in silico analysis (and validated by RT Q-PCR), such as *R. bromii* and *P. distasonis* (Fig. 2), were also in greater abundance. Although *O. valericigenes* was overrepresented in the HCG it was not similarly identified by the microarray, probably due to the stringent threshold we set for microarray statistical analysis. *F. prausnitzii* was not shown as statistically discriminated by RT Q-PCR because the species diversity was larger than an OTU (data not shown), as confirmed by the five probes highlighted in the microarray. Nevertheless,

F. prausnitzii was detected as a healthy specific species. Furthermore, several OTUs that were identified by in silico predictions, but not statistically supported by PCR data (data not shown) were confirmed by microarray as overrepresented in healthy microbiota such as *B. bifidum*.

Overrepresented Species in CD Microbiota

Both the RT Q-PCR and microarray analyses showed that *E. coli* was overrepresented in the CD samples. Furthermore, other members of the Proteobacteria phylum, including *Proteus vulgaris* and *Enterobacter cowanii*, were also present in greater abundance in CD patients. In relation to Gram-positive bacteria, *Enterococcus faecium* species appeared to be overrepresented in a subset of patients and some patients were found to have a greater abundance of *Streptococcus* spp.

DISCUSSION

This study confirms recent data but also provides new insights into the phylogenetic specificities of CD-associated microbiota by combining bioinformatics predictions with biological validation, via a combination of RT Q-PCR and microarray approaches.

First, combination of RT Q-PCR targeting six human gut commensal bacteria in healthy and diseased individuals provided a preliminary diagnostic tool with both relevant sensitivity (93.75%) and specificity (87.5%) for discriminating subjects according to their health status. Then, species shown in this study as being "healthy specific" were also part of the recently described human intestinal phylogenetic core.[30] These species were also identified by the in silico analysis of data arising from non-European studies, and this shows that the dominant and prevalent bacterial species associated with healthy microbiota are likely universal. As such, these species should be targeted to monitor "normal" human GI tract health. Among the healthy microbiota, *O. valericigenes* was shown to be a very good phylogenetic marker of clinical status in the cohort examined here. Recently, *Bifidobacterium longum* and "*Oscillospira* sp.," a genus close to *O. valericigenes,* were shown to grow cooperatively; the exopolysaccharides produced by *B. longum* were associated with the increased abundance of "*Oscillospira* sp." in the human GI tract.[31] Our study also confirmed that *F. prausnitzii,* recently shown to produce antiinflammatory factors, was present in lower abundance in CD.[18,32] Furthermore another member of the *C. leptum* group, *R. bromii*, was reduced in subjects with inflammation.

The bacterial genomes for some members of this "healthy microbial core" are now being sequenced via the Human Microbiome Project (http://www.hmpdacc.org/) and the International Human Microbiome Consortium (http://www.human-microbiome.org/), and this will certainly help to better understand their functional role in the human GI
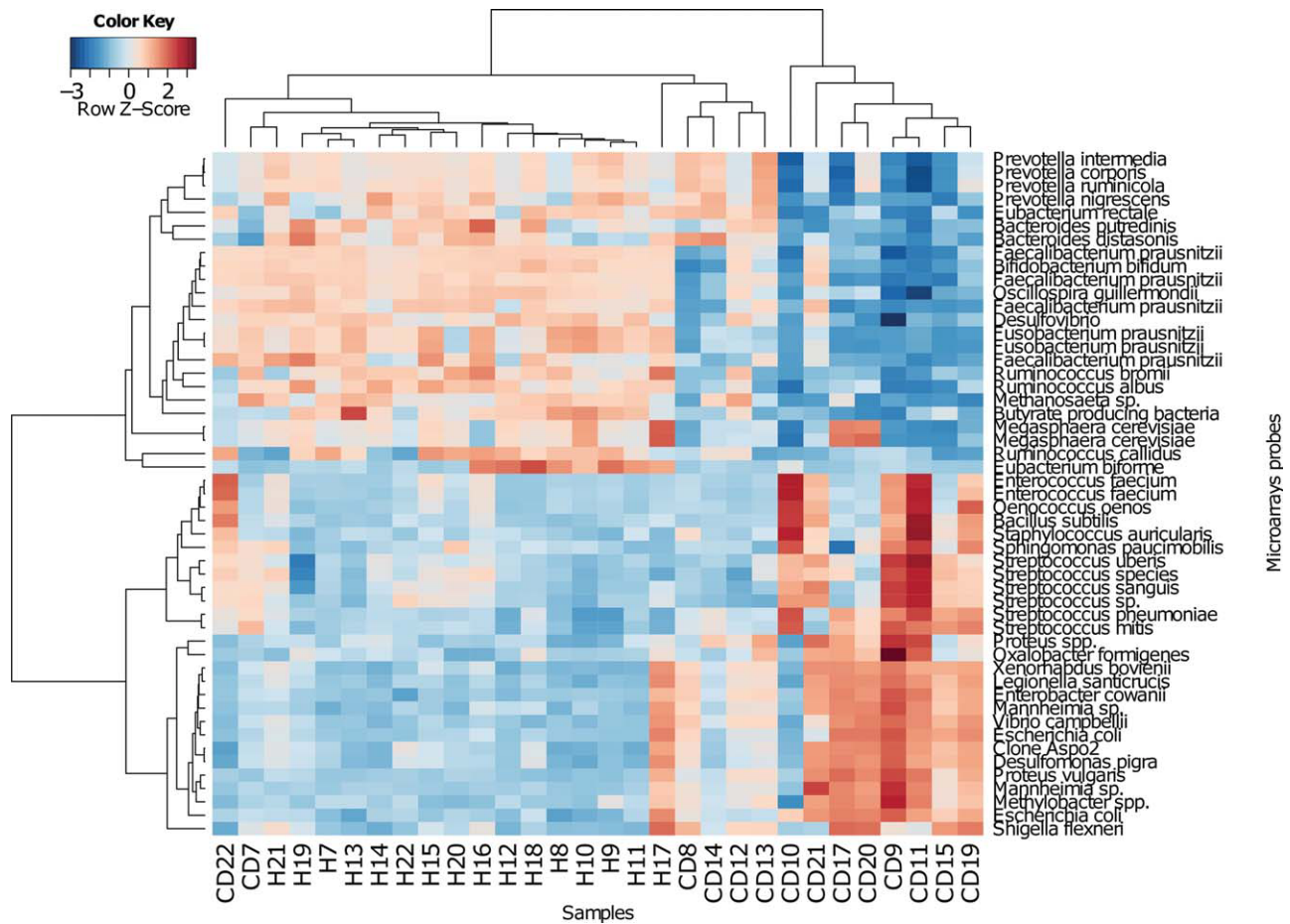
FIGURE 5. A between-class analysis was computed on the CoA results from the whole dataset. A heat plot map was applied on the discriminated probes retrieved from the first coordinate of the between-class analysis. Red indicates high intensity probes and corresponds to an overrepresentation of species in samples. Conversely, blue indicates an underrepresentation.

tract. Such information will be critical to augment our understanding of the roles of these bacteria in health and disease. Despite these advances, many of the "core" microbes, their interactions and processes, and their role in health (or disease initiation and progression) remain cryptic. For instance, "healthy" bacterial structural components such as pili, flagella, and fimbriae, as well as metabolites and other gene products that might interact with human immune receptors (e.g., Toll-like receptors, Nod-like receptors), require better characterization. Dysbiosis in CD is probably not due to a single species, and our present study demonstrates the importance of a non a priori method approach, since it provided new species linked to active CD. For instance, OTUc03, which seems of interest, represents a species far from any sequence in the database and should therefore be targeted by single-cell genomics.

Some of the healthy species were missing, but, conversely, other species were more prevalent in CD patients microbiota. *E. coli*, as previously demonstrated, was overrepresented in CD patients.[12] In addition, several species

from the Proteobacteria phylum were linked to the CD status. Most of these species (*Streptococcus* spp., *Enterococcus* spp., *Staphylococcus* spp., *Legionella* spp., *Vibrio* spp., and *Proteus* spp.) are known to be facultative anaerobes, and it seems thus that the microbiota during active ileal CD is adapted to a high redox potential.

The variability among CD patients appeared higher than among healthy individuals, as reported in the literature.[32] Such variability among CD patients might explain why OTUs specific to CD status have been difficult to detect. A hypothesis is that individuals who are predisposed to CD are less able to regulate the microbial makeup of their gut, leading to an unstable microbial population. Moreover, the CD microbial profiles highlighted by our data are consistent with the deleterious schema believed to be caused by changes in antimicrobial peptide production which result in the killing of *B. vulgatus, E. coli,* and *E. faecalis*.[33]

This study focused on patients suffering from an active ileal form of CD, which is of particular interest

because of its proximity to the Peyer patches and immune cells. Because CD is a disease with different phenotypes in terms of pathology and progression, host genetics and lifestyle differences might explain the discrepancies between results arising from different cross-sectional studies. Several ongoing projects using high-throughput metagenomics approaches (such as the EU-funded project METAHIT, http://www.metahit.eu/) will allow deeper analyses that will provide additional confirmation of our in silico predictions. These metagenomic datasets from CD patients will help identify which bacterial genes are under- or overrepresented during inflammation. Clinically, dysbiosis can be studied further by a large screening of individuals by a fast quantification of the six OTUs we detected here as discriminatory of CD status.

ACKNOWLEDGMENTS

REFERENCES

1. Barrett JC, Hansoul S, Nicolae DL, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet*. 2008;40:955–962.
2. Cosnes J, Carbonnel F, Carrat F, et al. Effects of current and former cigarette smoking on the clinical course of Crohn's disease. *Aliment Pharmacol Ther*. 1999;13:1403–1411.
3. Seksik P, Nion-Larmurier I, Sokol H, et al. Effects of light smoking consumption on the clinical course of Crohn's disease. *Inflamm Bowel Dis*. 2009;15:734–741.
4. Loftus EV Jr. Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences. *Gastroenterology*. 2004;126:1504–1517.
5. Sartor RB. Microbial influences in inflammatory bowel diseases. *Gastroenterology*. 2008;134:577–594.
6. Bouskra D, Brezillon C, Berard M, et al. Lymphoid tissue genesis induced by commensals through NOD1 regulates intestinal homeostasis. *Nature*. 2008;456:507–510.
7. Vaishnava S, Behrendt CL, Ismail AS, et al. Paneth cells directly sense gut commensals and maintain homeostasis at the intestinal host-microbial interface. *Proc Natl Acad Sci U S A*. 2008;105:20858–20863.
8. Garrett WS, Lord GM, Punit S, et al. Communicable ulcerative colitis induced by T-bet deficiency in the innate immune system. *Cell*. 2007;131:33–45.
9. Savage DC. Microbial ecology of the gastrointestinal tract. *Annu Rev Microbiol*. 1977;31:107–133.
10. 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.
11. Bull TJ, McMinn EJ, Sidi-Boumedine K, et al. Detection and verification of Mycobacterium avium subsp. paratuberculosis in fresh ileocolonic mucosal biopsy specimens from individuals with and without Crohn's disease. *J Clin Microbiol*. 2003;41:2915–2923.
12. 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.
13. Suau A, Bonnet R, Sutren M, et al. Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. *Appl Environ Microbiol*. 1999;65:4799–4807.
14. Eckburg PB, Bik EM, Bernstein CN, et al. Diversity of the human intestinal microbial flora. *Science (New York)*. 2005;308:1635–1638.
15. 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.
16. Tamboli CP, Neut C, Desreumaux P, et al. Dysbiosis in inflammatory bowel disease. *Gut*. 2004;53:1–4.
17. Seksik P, Rigottier-Gois L, Gramet G, et al. Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. *Gut*. 2003;52:237–242.
18. Sokol H, Pigneur B, Watterlot L, et al. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A*. 2008;105:16731–16736.
19. Marteau P, Lemann M, Seksik P, et al. Ineffectiveness of Lactobacillus johnsonii LA1 for prophylaxis of postoperative recurrence in Crohn's disease: a randomised, double blind, placebo controlled GETAID trial. *Gut*. 2006;55:842–847.
20. Katoh K, Kuma K, Toh H, et al. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*. 2005;33:511–518.
21. Ashelford KE, Chuzhanova NA, Fry JC, et al. New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. *Appl Environ Microbiol*. 2006;72:5734–5741.
22. Schloss PD, Handelsman J. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol*. 2005;71:1501–1506.
23. Schloss PD, Handelsman J. Introducing SONS, a tool for operational taxonomic unit-based comparisons of microbial community memberships and structures. *Appl Environ Microbiol*. 2006;72:6773–6779.
24. Fu Q, Ruegger P, Bent E, et al. PRISE (PRImer SElector): software for designing sequence-selective PCR primers. *J Microbiol Methods*. 2008;72:263–267.
25. Loy A, Arnold R, Tischler P, et al. probeCheck—a central resource for evaluating oligonucleotide probe coverage and specificity. *Environ Microbiol*. 2008;10:2894–2898.
26. Furet JP, Firmesse O, Gourmelon M, et al. Comparative assessment of human and farm animal faecal microbiota using real-time quantitative PCR. *FEMS Microbiol Ecol*. 2009;68:351–362.
27. Kang S, Denman SE, Morrison M, et al. Dysbiosis of faecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray. *Inflamm Bowel Dis*. 2010 (in press).
28. Chessel D, Dufour A, Thioulouze J. The ade4 package. I. One-table methods. *R News*. 2004;4:5–10.
29. Culhane AC, Thioulouse J, Perriere G, et al. MADE4: an R package for multivariate analysis of gene expression data. *Bioinformatics (Oxford)*. 2005;21:2789–2790.
30. Tap J, Mondot S, Levenez F, et al. Towards the human intestinal microbiota phylogenetic core. *Environ Microbiol*. 2009;11:2574–2584.
31. Salazar N, Gueimonde M, Hernandez-Barranco AM, et al. Exopolysaccharides produced by intestinal Bifidobacterium strains act as fermentable substrates for human intestinal bacteria. *Appl Environ Microbiol*. 2008;74:4737–4745.
32. 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.
33. Nuding S, Fellermann K, Wehkamp J, et al. Reduced mucosal antimicrobial activity in Crohn's disease of the colon. *Gut*. 2007;56:1240–1247.