Twin studies reveal specific imbalances in the mucosa-associated microbiota of patients with ileal Crohn’s disease

Short title: Mucosal microbiota of Crohn’s twins

Ben Willing,*, †, PhD, Jonas Halfvarson, ‡†, MD PhD, Johan Dicksved,*, PhD, Magnus Rosenquist,*, PhD, Gunnar Järnerot, ‡, MD PhD, Lars Engstrand, ‡§, MD PhD, Curt Tysk‡¶, MD PhD, and Janet K Jansson***††, PhD

* Swedish University of Agricultural Sciences, Department of Microbiology, Uppsala, Sweden
‡ Division of Gastroenterology, Department of Medicine, Örebro University Hospital, Örebro, Sweden.
§ Department of Oncology, Radiology and Clinical Immunology, Uppsala University Hospital, Uppsala, Sweden
§§ Swedish Institute for Infectious Disease Control, Department of Bacteriology, Solna, Sweden
¶ School of Health and Medical Sciences, Örebro University, Örebro, Sweden.
** Lawrence Berkeley National Laboratory, Earth Sciences Division, Berkeley, CA

† Authors contributed equally to this work

†† Corresponding author: Janet K. Jansson, Lawrence Berkeley National Laboratory, Earth Sciences Division, 1 Cyclotron Road, MS 70A-3317D, Berkeley, CA 94720 USA, Phone number: +1 (510) 486-7487, Email: jrjansson@lbl.gov
Sources of support: This project was funded by the Örebro University Hospital Research Foundation, Örebro County Research Foundation, Bengt Ihre’s Foundation, the Uppsala BioX Micprof project, the Swedish University of Agricultural sciences and, in part, by the U.S. Department of Energy Contract DE-AC02-05CH11231 with Lawrence Berkeley National Laboratory.

Figures: 2
Tables: 2
ABSTRACT AND KEY WORDS

Large inter-individual variation in the composition of the intestinal microbiota between unrelated individuals has made it challenging to identify specific aspects of dysbiosis that lead to Crohn’s disease. To reduce variations in exposure during establishment of the gut flora and influence of genotype, we studied the mucosa-associated microbiota of monozygotic twin pairs that were discordant (n=6) or concordant (n=4) for Crohn’s disease. DNA was extracted from biopsies collected from 5 locations between the ileum and rectum. Bacterial 16S ribosomal RNA genes were amplified and community composition assessed by terminal-restriction fragment length polymorphism, cloning and sequencing and quantitative real-time PCR. The microbial compositions at all biopsy locations for each individual were similar, regardless of disease state, but there were differences between individuals. In particular, individuals with predominantly ileal Crohn’s had a dramatically lower abundance \((P<0.001)\) of *Faecalibacterium prausnitzii* and increased abundance \((P<0.03)\) of *Escherichia coli* compared to healthy co-twins and those with Crohn’s localized in the colon. This dysbiosis was significantly correlated to the disease phenotype rather than genotype. The reduced abundance of *F. prausnitzii* and increased abundance of *E. coli* are indicative of an ileal Crohn’s disease phenotype, distinct from colonic Crohn’s disease and the relative abundances of these specific bacterial populations are promising biomarker candidates for differential diagnosis of Crohn’s and eventually customized treatment.

Key Words: Monozygotic discordant twins; ileal Crohn’s disease; *Faecalibacterium prausnitzii*; *Escherichia coli*; mucosa-associated microbiota

INTRODUCTION
Defining the etiology of Crohn’s disease (CD) has proven difficult due to the complicated interplay of environmental and genetic factors leading to disease development. Increasing evidence suggests that CD is ultimately the result of a breakdown in the détente relationship between commensal bacteria in the intestine and the host immune system. A disease concordance rate of approximately 50% in monozygotic twins identifies genetics as a major contributing factor towards CD development. However, an equal discordance rate in monozygotic twins points to the importance of environmental factors, including the composition of the intestinal microbiota.

The current hypothesis is that sufficient antigenic stimulation to cause disease is provided by an imbalance of beneficial and detrimental commensal organisms or ‘dysbiosis’. Recent reports indicate increased abundances of Bacteroides and Enterobacteriaceae, particularly Escherichia coli, and a reduction in microbial diversity in CD. However, no single bacterial species has been convincingly indicated as a biomarker of CD to date. Since host genotype plays a substantial role in shaping microbial population, as evidenced by similarities in monozygotic twins, it is unclear whether microbial imbalance is associated with host genotype or disease. This can be clarified by studying monozygotic twin pairs that are discordant (one is healthy and one is sick) for disease.

The mucosal microbiota differs from that of the digesta and is of particular interest because the mucosal surface is the site of microbial recognition by the host that if improperly regulated could lead to the inflammatory responses typical of CD. Therefore, we studied biopsy samples collected at colonoscopy from previously studied monozygotic twin pairs that were either discordant or concordant for CD. Our hypothesis was that studying discordant monozygotic twins in depth
using a combination of molecular approaches would facilitate the identification of specific bacteria that are correlated with CD incidence rather than host genotype, including those not yet cultivated.

MATERIALS AND METHODS

Human subjects

A total of 10 monozygotic twin pairs obtained from a previously described Swedish twin population were studied: 2 pairs discordant for predominantly ileal Crohn’s (ICD), 4 pairs discordant for predominantly colonic Crohn’s (CCD), 2 pairs concordant for ICD, and 2 pairs concordant for CCD. Table 1 summarizes relevant clinical data for patients and responses to a questionnaire regarding the usage of antibiotics, non-steroidal anti-inflammatory drugs within the preceding 12 months, gastroenteritis within the last 3 months and specific dietary habits that have been previously reported. Subjects were born between 1936 and 1986 with patient groups, as defined by disease status, being similar in age (mean±SD) CCD 49.0±18.5 (n=8), healthy (HC) 52.8±17.2 (n=6), ICD 50.8±4.5 (n=6)). According to Harvey Bradshaw score, all individuals were in remission with the exception of two (10b and 15a), who previously had undergone ileocecal resection. However, none of these had postoperative recurrence, endoscopic recurrence score <2 at colonoscopy, strongly suggesting that their Harvey Bradshaw score reflected a co-existing functional disorder or possibly bile acid malabsorption and not active CD. The use of human subjects for this study was approved by the Örebro County Ethical Committee (Dnr167/03).

Biopsy collection

Patients received standard bowel cleansing with the polyethylene glycol preparation, Laxabon® (BioPhausia, Stockholm, SE), the evening before colonoscopy. Biopsies
were taken from distal ileum, ascending, transverse and descending colon and rectum, if technically possible. Two biopsies were collected from each location in each subject, using disposable biopsy forceps (Olympus, Center Valley, PA). Biopsies were immediately placed in 0.5 ml freezing buffer (5 mM K$_2$HPO$_4$, 1.3 mM KH$_2$PO$_4$, 2 mM Na$_3$C$_6$H$_5$O$_7$, 1mM MgSO$_4$·7H$_2$O and 17.2% glycerol), in CryoPlus tubes (Sarstedt AG & Co, Germany), frozen on dry ice, and stored at -70°C.

**DNA extraction**

Biopsies were thawed on ice and centrifuged at 15000 x g for 5 min. Freezing buffer was removed by aspiration and DNA was isolated from the entire biopsy using the QIAamp® DNA Mini Kit (Qiagen Inc., Hilden, Germany). Extraction was performed according to the manufacturer’s instructions, with the addition of two 45 s bead-beating steps at level 5 on a FastPrep®-24 (MP Biomedicals, Solon, OH) at the beginning of the protocol.

**Terminal-restriction fragment length polymorphism (T-RFLP)**

To obtain a profile of mucosa-associated bacteria, the total extracted DNA from each biopsy was analyzed by T-RFLP using a protocol previously developed for DNA extracted from faeces.²² Briefly, bacterial 16S rRNA genes were specifically amplified by PCR using broad-range bacterial primers Bact-8F, ²⁵ 5’-end labeled with 6-carboxyfluorescein, and 926r²⁶ (Table 2). Resulting PCR products were digested with HaeIII (New England BioLabs, Ipswich, MA) and separated on an ABI 3730 capillary sequencer (Applied Biosystems, Foster City, CA). Electrophoreograms were processed using Peak Scanner v1.0 (Applied Biosystems) and relative peak areas of terminal restriction fragments (TRFs) corresponding to sizes between 50 and 500 bp were calculated by dividing individual peak area by total peak area within this size.
constraint. Only peaks above a threshold of 0.5% were included in further analyses. A minimum threshold of $10^4$ 16S rRNA gene copies, as measured by qPCR (see below), was set to avoid bias due to low total abundance.

**Cloning and sequencing**

16S rRNA genes were cloned from selected DNA samples (11a, 11b, 13a and 13b) to aid in the identification of bacteria corresponding to TRF sizes of interest based on the analysis of T-RFLP data. Extracted DNA samples were amplified using primers Bact-8F and 926r.\textsuperscript{25,26} PCR products were gel purified (Qiagen), cloned into TOPO TA pCR 4.0 vector (Invitrogen, Carlsbad, CA) and transformed into \textit{E. coli} TOP 10 chemically competent cells (Invitrogen). Transformants were screened by PCR amplification with M13 primers (Invitrogen) and appropriate PCR product sizes were determined by agarose gel electrophoresis. The resulting PCR products were then analyzed by T-RFLP as described above and products that resulted in TRF lengths corresponding to peaks of interest were sequenced. Sequences obtained were subjected to phylogenetic assignment using RDP’s Naïve Bayesian rRNA Classifier Version 2.0. Unique sequences were deposited to Genbank at NCBI under accession numbers EU668957 to EU668961.

**Quantitative real-time PCR (qPCR)**

All qPCR assays were performed in duplicate 25 µl reactions in 96-well plates using an iQ\textsuperscript{TM}5 Real-time Detection System (BioRad, Hercules, CA). Total 16S rRNA gene copies for bacteria were determined for each sample with the general bacterial primers (Table 2) using iQ\textsuperscript{TM} supermix (Biorad). Human cell numbers were determined for biopsies from the ileum and descending colon using self designed primers specific for the $\beta$-globin gene (Table 2), that were designed using MacVector 9.5.4 (Accelrys Software, Inc., San Diego CA). 16S rRNA gene copies for the
specific bacterium, *Faecalibacterium prausnitzii*, were also determined using specific primers (Table 2). *E. coli* was quantified using self-designed primers that amplify *uidA* (Table 2), encoding β-glucuronidase, which is specific to *E. coli* and *Shigella*. Specificity to *E. coli* was confirmed by negative PCR for *Shigella* specific *ipaH*.^{29} qPCR assays for β-globin, *uidA* and *F. prausnitzii* were performed using SYBR®GreenER™ qPCR Supermix for iCycler® (Invitrogen). Resulting melting curves were visually inspected to ensure specificity of product detection. For quantification of target DNA copy number, standard curves were generated for each gene of interest using gel extracted PCR products quantified using a Qubit™ Fluorometer (Invitrogen).

**Statistical analysis**

To assess the diversity of the microbial community, defined by evenness and richness of TRFs, Simpson’s index of diversity^{30} was applied to the T-RFLP data. Similarity indices were calculated within individuals, comparing each location to the consensus of all locations, using Manhattan metrics^{31}. Differences between discordant twins were calculated using Manhattan metrics and significance tested by ANOVA (SPSS, Chicago, IL). qPCR data and abundance data for individual TRFs were analyzed as a one-way ANOVA using the general linear model (SPSS). Group means were separated by REGWF with significance of $P<0.05$. Correlation coefficients were calculated using the Pearson Correlation Procedure (SPSS).

**RESULTS**

**Mucosa-associated community profiles**

Of the 94 biopsies collected, 8 (3 CCD, 2 ICD, 3 HC) were not included in T-RFLP analysis because of low bacterial numbers in those samples (< $10^4$ 16S rRNA gene copies). However, a minimum of 3 biopsies was included from each subject for
further analyses. A total of 111 TRFs, representing different ribotypes, were found in the 20 individuals studied. The average number of TRFs±SE, per biopsy was 28.2±2.94, 27.4±1.67 and 23.6±5.99 for CCD, HC and ICD respectively.

**Microbial Density**

Bacterial numbers were not affected by disease phenotype ($P=0.376$) or sample location ($P=0.295$). Mean 16S rRNA gene copy numbers per copy β-globin±SE were 35.5±15.6 and 25.2±15.3 for CCD, 27.4±15.7 and 11.4±5.7 for HC, 17.0±10.3 and 10.6±5.3 for ICD, for the ileum and distal colon respectively.

**Diversity**

Microbial diversity (mean±SE), measured as a combination of TRF richness and evenness, was lower ($P<0.004$) in individuals with ICD (0.798±0.05) than those with CCD (0.850±0.010) or HC (0.877±0.018). This result may be confounded by surgery, however the microbial diversity in CCD individuals who had undergone surgery was not different ($P=0.67$) from CCD individuals who had not undergone surgery.

**Consistency between locations**

Cluster analysis of T-RFLP profiles, using Bray Curtis metrics, showed that the microbial compositions in samples collected from an individual grouped together, regardless of biopsy location (Figure 1). Similarity scores for biopsies, using the Manhattan index, as compared to the mean of all biopsies within an individual were generally higher than 80% with an average of 88.5%. No difference in the microbial community structures at the different sampled locations within individuals was observed in the different sampled groups ($P=0.25$); with similarity scores (mean±SE): HC (86.0±5.3), ICD (88.8±3.3) and CCD (90.3±1.7) groups.

**Similarity between twins**
The average microbial community structures from all biopsy locations for each individual were compared between matched twin pairs. The discordant twin pairs, with ICD, had lower similarity scores (36-53%), measured using abundance data (Manhattan index), compared to the discordant twins with CCD (52-66%). This was also reflected in lower average similarities between concordant twins pairs with ICD (48-62%) compared to concordant twins with CCD (56-67%). These data show a trend ($P<0.07$) for individuals with ICD to be less similar to their twin than CCD individuals, however concordant twins were not more similar ($P=0.29$) than discordant twins.

**Phenotypic differences in mucosa-associated microbiota**

*Faecalibacterium prausnitzii*

Analysis of T-RFLP data identified TRF223 to be abundant in all HC and CCD individuals, but scarce or absent in subjects with ICD. This TRF corresponded to *Faecalibacterium prausnitzii* (98.2-99.5% similarity) based on screening of 16S rRNA gene clone sequences obtained from the same individuals. It was noteworthy that all sequences were more similar to uncultured clones deposited in databases than to the cultured *F. prausnitzii* type strain.

We found that *F. prausnitzii*, measured by qPCR as a percent of total 16S rRNA genes (Figure 2A), was dramatically lower in individuals with ICD ($P<0.001$) than those with CCD or HC, with respective values of (mean±SE) 0.4±0.89 (ICD), 11.3±3.36 (CCD) and 8.7±2.49 (HC). Discordant ICD twins (pairs 16 and 18) segregated to disease phenotype rather than twin pair for *F. prausnitzii* abundance.

*Escherichia coli*

We designed novel primers targeting the *uidA* gene for specific quantitation of *Escherichia coli* by real-time PCR (Figure 2B). *E. coli* was significantly more
abundant \((P < 0.03)\) in biopsies at all locations from ICD patients than in HC or those with CCD. In 8 of 9 positive diseased individuals, whether CCD or ICD, when \(E. coli\) was present at detectable levels it was found at all locations \((1.3-4.5 \log uidA\) per \(10^6\) 16S rRNA genes). \(E. coli\) was also detected in some healthy individuals \((11b, 12a,\) and \(14a)\) however, only in a maximum of 2 locations, and at a level of 1.2-2.5 \(\log uidA\) per \(10^6\) 16S rRNA genes. However, in one pair of ICD concordant twins \((pair 10)\) \(E. coli\) was only detected in one of the individuals, but at all locations. Similar to the situation we observed for \(F. prausnitzii\), the discordant twins with one individual having CD localized in the ileum, segregated according to disease phenotype rather than twin pair with respect to \(E. coli\) abundance.

**DISCUSSION**

It is known that the host genotype partly determines the microbial community composition in the human gut.\(^{15-17}\) Therefore, our hypothesis was that by studying the mucosa-associated microbiota of identical twins it would be easier to untangle the respective contributions of host genetics and intestinal bacteria towards CD etiology. We found that discordant ICD twins segregated according to their disease phenotype rather than twin pair with respect to the abundance of some specific bacterial populations. These results suggest that the bacterial composition in the intestine is associated with disease rather than host genotype.

We used two molecular approaches to quantify the abundance of specific bacterial populations in the biopsy samples. Both T-RFLP and qPCR approaches provided significant indications that specific bacterial species were more or less abundant in individuals with ICD compared to CCD or healthy individuals. In particular, TRF223, corresponding to \(F. prausnitzii\), was abundant in all HC and CCD individuals, but scarce or absent in subjects with ICD. We previously detected the
same TRF in faecal samples collected from 21 out of 22 healthy individuals; including some of the same individuals included in the present study and in 89 out of 90 healthy children in a separate study, but at that time we did not have the sequence data to match this TRF to a particular species. The abundant and consistent presence of *F. prausnitzii* in healthy individuals suggests that it is an important member of the mucosal microbiota. *F. prausnitzii* produces butyrate, formate and D-lactate as a product of fermentation, providing the major energy source for colonic epithelial cells and having effects on epithelial barrier integrity and immune modulation. The functional activity of *F. prausnitzii* in affecting host physiology was recently demonstrated by Li et al., where they showed that its population variation was associated with eight urinary metabolites of diverse structure including butyrate isoforms. The loss of *F. prausnitzii*, and thus reduced butyrate production may trigger gut inflammation in genetically susceptible individuals. The absence of this organism may also leave a gap that can be filled by other organisms resulting in a less stable population.

Using qPCR the abundance of *F. prausnitzii* was confirmed to be significantly lower in all biopsies collected from ICD patients, compared to healthy controls and those with disease location in the colon. Recently *Faecalibacteria* and *Subdoligranula* were found to represent a combined total of 0.4%, 15% and 26% of bacterial 16S rRNA gene clones from ileal biopsies in ICD, healthy controls and CCD subjects respectively. Martinez-Medina et al. also observed *F. prausnitzii* more frequently in healthy (87%) compared to CD (53%) subjects, however they did not separate CD phenotypes in their analysis or compare abundance data. Vasquez et al. detected *F. prausnitzii* by FISH in only 5 of 22 CD patients that had undergone ileal resection, and at low levels (1.8±0.5%) in positive individuals. Additional reports of reduced
abundance and diversity of Firmicutes, particularly from the Clostridium leptum group of which F. prausnitzii is a member, coincide with our findings, however, F. prausnitzii was not specifically mentioned in those studies.

Swidsinski et al. recently reported that low faecal levels of F. prausnitzii, along with faecal leukocyte counts, differentiated active CD from UC. Although they did not indicate disease location in CD patients, they found that treatment with high-dose cortisol therapy or infliximab restored F. prausnitzii levels indicating that this bacterium is being suppressed by the host immune system. Intriguingly, all enrolled twins were in endoscopic remission, yet differences in F. prausnitzii were observed in ICD.

Another bacterial population that differentiated according to disease phenotype was E. coli. Increased abundance of E. coli in biopsies from CD patients has previously been reported. However, our study and one other identify a particular increase in E. coli levels in subjects with ICD, compared to HC. We found a consistent presence of E. coli in CD patients, spanning from the ileum to the rectum. Previously, an increased E. coli abundance in rectum biopsies was found in CD patients. Although the E. coli were not isolated and typed in this experiment, E. coli isolated from CD biopsies in previous studies have shown pathogen-like behaviour in vitro, including survival within macrophages, and may play a role in the inflammatory process. Reduced production of α-defensins and reduced antimicrobial activity against E. coli has been observed in ICD but not CCD as compared to healthy mucosa. This phenomenon would directly explain increased E. coli abundance in ICD observed in this and another study. Intriguingly, butyrate induces the expression of some anti-microbial peptides, thus a reduced abundance of
butyrate producing \textit{F. prausnitzii} may ultimately result in reduced secretion of antimicrobial peptides and proteins allowing \textit{E. coli} to proliferate.

We were unable to identify any TRFs that clearly differentiated CCD individuals from HC, although there were a few TRFs that were more common in CCD than HC and ICD. However, an increasing number of reports have indicated differences between ICD and CCD phenotypes, including beneficial effects of antibiotic therapy in CCD, but not ICD.\textsuperscript{44,45} The reduced effectiveness of antibiotic therapy in ICD has not been explained, however may be a result of the differences in microbial composition in these two disease phenotypes, and the intracellular localization of \textit{E. coli} within macrophages.\textsuperscript{46} When we previously studied faecal samples from these subjects, we observed that community profiles of twins with CCD were similar to those of healthy individuals, while those with ileal involvement separated from the others.\textsuperscript{22}

The consistency of the bacterial composition along the intestinal tract in each individual, ranging from the distal ileum to the rectum, suggests that dysbiosis associated with CD is not localized to the area of disease. This consistency between locations regardless of disease phenotype confirms previous reports in healthy and IBD individuals.\textsuperscript{19,37,47}

Reduced bacterial diversity in CD individuals compared to their healthy co-twins agrees with numerous other reports in non twin populations,\textsuperscript{12-14,48,49} but this is the first finding that specifically points to a lower diversity for a specific disease phenotype (i.e. ICD). Reports of increased bacterial numbers associated with biopsies from CD patients as compared to healthy controls have been inconsistent.\textsuperscript{47,50,51} We did not observe any differences in bacterial numbers associated with disease phenotype, however, this may be due to the nature of the control patients, as
individuals related to IBD patients may also have disrupted mucosal surfaces although they do not suffer from disease.⁵²,⁵³

We acknowledge that these results are subject to biases inherent in PCR amplification and that we may have missed differences in some members of the microbial community, however, biases were consistent in all groups. In this investigation we studied patients who were in remission and not diseased, although the gut microbiota is altered in a diseased compared to a quiescent state.⁴⁸ Confounding factors including surgery and medication in ICD may also give rise to some skepticism, however similar surgeries were performed in individuals with CCD (12b and 13a), which had normal levels of *F. prausnitzii* and an absence of *E. coli*. Frank et al.⁵¹ also found that appendectomy was not correlated with the subsequent microbial composition in colonic biopsies of IBD patients.

In conclusion we identified specific bacterial species that are significantly increased or decreased in abundance in individuals with Crohn’s disease in the ileum, compared to individuals with CD in the colon and to healthy individuals. Specifically, we propose that lower amounts of *F. prausnitzii* and higher amounts of *E. coli* are indicative of ileal CD and can be used to distinguish the ileal disease phenotype from colonic CD and from healthy individuals. The data from discordant twins indicate that the host genotype is not solely responsible for the observed differences in the gut microbiota, but instead support the hypothesis that ICD and CCD are different subgroups of the disease. Therefore, we propose that the relative abundances of *F. prausnitzii* and *E. coli* can serve as indicators to distinguish Crohn’s disease phenotype, and may eventually support customized treatment.

**Acknowledgements**
This project was funded by the Örebro University Hospital Research Foundation, Örebro County Research Foundation, Bengt Ihre’s Foundation, the Uppsala BioX Micprof project, the Swedish University of Agricultural sciences and, in part, by the U.S. Department of Energy Contract DE-AC02-05CH11231 with Lawrence Berkeley National Laboratory. We thank Kerstin Eriksson for her assistance with collection of biological material.
REFERENCES


Table 1 Clinical data for patients and responses to a questionnaire regarding the usage of antibiotics, non-steroidal anti-inflammatory drugs within the preceding 12 months and gastroenteritis within the last 3 months (same number for twin sets, followed by a small letter “a” or “b” for each individual in a pair).

<table>
<thead>
<tr>
<th>ID</th>
<th>Status</th>
<th>Birth year</th>
<th>Sex</th>
<th>Nod2 status</th>
<th>Gastroenteritis</th>
<th>Disease duration</th>
<th>Antibiotics</th>
<th>NSAID</th>
<th>Surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td>9a</td>
<td>CCD</td>
<td>1947</td>
<td>M</td>
<td>wt</td>
<td>No</td>
<td>19</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>9b</td>
<td>CCD</td>
<td>1947</td>
<td>M</td>
<td>wt</td>
<td>No</td>
<td>20</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>10a</td>
<td>ICD</td>
<td>1962</td>
<td>F</td>
<td>wt</td>
<td>Yes</td>
<td>22</td>
<td>No</td>
<td>Yes</td>
<td>ileal res + righ hemi</td>
</tr>
<tr>
<td>10b</td>
<td>ICD</td>
<td>1962</td>
<td>F</td>
<td>wt</td>
<td>Yes</td>
<td>21</td>
<td>No</td>
<td>Yes</td>
<td>ileocecal res</td>
</tr>
<tr>
<td>11a</td>
<td>CCD</td>
<td>1986</td>
<td>F</td>
<td>snp 8</td>
<td>No</td>
<td>4</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>11b</td>
<td>H</td>
<td>1986</td>
<td>F</td>
<td>snp 8</td>
<td>No</td>
<td>N/A</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>12a</td>
<td>CCD</td>
<td>1953</td>
<td>M</td>
<td>wt</td>
<td>No</td>
<td>33</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>12b</td>
<td>CCD</td>
<td>1953</td>
<td>M</td>
<td>wt</td>
<td>No</td>
<td>33</td>
<td>No</td>
<td>Yes</td>
<td>ileocecal res</td>
</tr>
<tr>
<td>13a</td>
<td>CCD</td>
<td>1943</td>
<td>F</td>
<td>wt</td>
<td>No</td>
<td>18</td>
<td>No</td>
<td>Yes</td>
<td>segm colonic res</td>
</tr>
<tr>
<td>13b</td>
<td>H</td>
<td>1943</td>
<td>F</td>
<td>wt</td>
<td>No</td>
<td>N/A</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>14a</td>
<td>H</td>
<td>1936</td>
<td>M</td>
<td>wt</td>
<td>No</td>
<td>11</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>14b</td>
<td>CCD</td>
<td>1936</td>
<td>M</td>
<td>wt</td>
<td>No</td>
<td>11</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>15a</td>
<td>ICD</td>
<td>1953</td>
<td>M</td>
<td>snp 8</td>
<td>No</td>
<td>31</td>
<td>No</td>
<td>No</td>
<td>ileal res</td>
</tr>
<tr>
<td>15b</td>
<td>ICD</td>
<td>1953</td>
<td>M</td>
<td>snp 8</td>
<td>No</td>
<td>31</td>
<td>No</td>
<td>No</td>
<td>ileocecal res</td>
</tr>
<tr>
<td>16a</td>
<td>ICD</td>
<td>1954</td>
<td>F</td>
<td>wt</td>
<td>No</td>
<td>33</td>
<td>No</td>
<td>Yes</td>
<td>Ileo res + right hemi</td>
</tr>
<tr>
<td>16b</td>
<td>H</td>
<td>1954</td>
<td>F</td>
<td>wt</td>
<td>No</td>
<td>N/A</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>17a</td>
<td>CCD</td>
<td>1976</td>
<td>M</td>
<td>wt</td>
<td>No</td>
<td>11</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>17b</td>
<td>CCD</td>
<td>1976</td>
<td>M</td>
<td>wt</td>
<td>No</td>
<td>11</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>18a</td>
<td>ICD</td>
<td>1953</td>
<td>M</td>
<td>wt</td>
<td>No</td>
<td>34</td>
<td>No</td>
<td>No</td>
<td>ileal res + right hemi</td>
</tr>
<tr>
<td>18b</td>
<td>H</td>
<td>1953</td>
<td>M</td>
<td>wt</td>
<td>No</td>
<td>N/A</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Abbreviations: ileal res, ileal resection; right hemi, right sided hemicolecotomy; ileocecal res, ileocecal resection; segm colonic res, segmental colonic resection; nd, no data; wt, wildtype; CCD, colonic Crohn’s disease; ICD, ileal Crohn’s disease; H, healthy

Table 2 Oligonucleotides for PCR, qPCR and terminal restriction fragment length polymorphism (T-RFLP)

<table>
<thead>
<tr>
<th>Target (purpose)</th>
<th>Sequence (5’-3’)</th>
<th>Annealing temp (°C)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacterial 16S rRNA gene (qPCR)</td>
<td>F: TCCTACGGGAGGCACGCTT</td>
<td>60</td>
<td>this study</td>
</tr>
<tr>
<td></td>
<td>R: GACTTACGGGTATCTACGTT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P: GGTATACGGGGGTCTGCGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F. prausnitzii (qPCR)</td>
<td>F: GATGGCGCGCGTCGATT</td>
<td>58</td>
<td>this study</td>
</tr>
<tr>
<td></td>
<td>R: TCCTTACCGTGCTCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>uidA (qPCR)</td>
<td>F: AGCTTGGTATGCTG</td>
<td>58</td>
<td>this study</td>
</tr>
<tr>
<td></td>
<td>R: CTGCTGCTGCTGTTGTCCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ipaH (PCR)</td>
<td>F: TGAGAAACTTCAGTCCCCTGCGG</td>
<td>60</td>
<td>this study</td>
</tr>
<tr>
<td></td>
<td>R: TTCTGATGCTGCTGAGCCAGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>β-globin (qPCR)</td>
<td>F: CACACTTCATCCCGATGCCC</td>
<td>58</td>
<td>this study</td>
</tr>
<tr>
<td></td>
<td>R: GAAGGCCAAAGGACACGTT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16S rRNA gene (T-RFLP/cloning)</td>
<td>F: AGAGTTGATCTGCTGCTGCAA</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R: CCGTCAATTCTCTTTGTT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURE LEGENDS

**Figure 1.** Similarity tree using Bray Curtis metrics of bacterial 16S rRNA gene terminal restriction fragment profiles from biopsies collected from the ileum (I), ascending colon (A), transverse colon (T), descending colon (D) and rectum (R) of individuals with predominantly ileal Crohn’s disease (magenta bars), predominantly colonic Crohn’s disease (green bars) and healthy individuals (blue bars). Same number for twin sets, followed by a small letter “a” or “b” for each individual in a pair.

**Figure 2.** Quantitative real-time PCR for *F. prausnitizii* 16S rRNA gene (A) and *E. coli uidA* (B) in mucosal DNA from the ileum, ascending colon, transverse colon, descending colon and rectum (displayed from left to right) from individuals with predominantly colonic Crohn’s disease (CCD), healthy controls (HC) and with predominantly ileal Crohn’s disease (ICD). Same number for twin sets, followed by a small letter “a” or “b” for each individual in a pair. (\* = CCD individuals with some ileal involvement. \** = ICD individuals with some colonic involvement)
Figure 2