Appl. Environ. Microbiol. doi:10.1128/AEM.02006-15

Mucosa-associated *Faecalibacterium prausnitzii* phylotype richness is reduced in inflammatory bowel disease patients.

AUTHORS

MSc. Mireia Lopez-Siles [1], PhD. Margarita Martinez-Medina [1], Prof. Carles Abellà [1], MD. David Busquets [2], MD. Miriam Sabat-Mir [3], PhD. Sylvia H. Duncan [4], MD. PhD, Xavier Aldeguer [2], Prof. Harry J. Flint [4], Prof. L. Jesús Garcia-Gil [1#]

AUTHORS' AFFILIATIONS

1. Laboratory of Molecular Microbiology, Biology Department, Universitat de Girona, Girona, Spain.

2. Department of Gastroenterology, Hospital Dr. Josep Trueta, Girona, Spain

3. Department of Gastroenterology, Hospital Santa Caterina, Salt, Girona, Spain

4. Microbial Ecology Group, Rowett Institute of Nutrition and Health, University of Aberdeen, Aberdeen, United Kingdom

RUNNING TITLE:

Phylotype richness of *Faecalibacterium prausnitzii*

[#] **For correspondence:**

Prof. L. Jesús Garcia-Gil,

Laboratory of Molecular Microbiology

Departament de Biologia

Universitat de Girona

Carrer de Maria Aurèlia Capmany, 69, E-17071, Girona, Spain.

Tel: +34 972 418175.

Fax: +34 972 418150.

e-mail: jesus.garcia@udg.edu

ABSTRACT

Faecalibacterium prausnitzii depletion in intestinal diseases has been extensively reported, but little is known about intra-species variability. This work aims to determine if subjects with gastrointestinal disease host different mucosa-associated *F. prausnitzii* populations from healthy individuals. A new species-specific polymerase chain reaction-denaturing gradient gel electrophoresis (PCR-DGGE) targeting the 16SrRNA gene was developed to fingerprint *F. prausnitzii* populations in biopsies from 31 healthy controls (H), 36 Crohn's disease (CD), 23 ulcerative colitis (UC), 6 irritable bowel syndrome (IBS), and 22 colorectal cancer (CRC) patients. The richness of *F. prausnitzii* subtypes was lower in inflammatory bowel disease (IBD) patients than in H subjects. The most prevalent operational taxonomic units (OTU) consisted of four phylotypes (OTU99), which were shared by all groups of patients. Their distribution and the presence of some disease-specific *F. prausnitzii* phylotypes allowed us to differentiate IBD and CRC population from that in H. At the OTU97 level, two phylogroups accounted for of 98% the sequences. Phylogroup I was found in 87% of H but in under 50% of IBD patients (P=0.003). In contrast, phylogroup II was detected in >75% of IBD patients and only in 52% of H subjects (P=0.005). This study reveals that despite the main members of *F. prausnitzii* population are present in both H and individuals with gut diseases, richness is reduced in the latter, and an altered phylotype distribution exists between diseases. This approach may serve as a basis for addressing the suitability of their quantification as putative biomarkers of disease, and depicting the importance of these subtype losses in disease pathogenesis.

KEYWORDS

Faecalibacterium prausnitzii, DGGE, richness, inflammatory bowel disease, Crohn's disease, ulcerative colitis, colorectal cancer, irritable bowel syndrome

INTRODUCTION

Metagenomic studies have shown that the human gut microbiota is constituted by a relatively limited number of dominating bacterial phyla. While in healthy adults, Bacteroidetes and Firmicutes are the most abundant phyla, Proteobacteria, Verrucomicrobia, Actinobacteria and Fusobacteria are relatively scarce (1-3). The Firmicute *Faecalibacterium prausnitzii* (Ruminococcaceae) is one of the three most abundant species, representing approximately 6-8% of the gut microbial community in healthy subjects, although it can reach up to 20% in some individuals (1, 4-11). In contrast, depletion of *F. prausnitzii* has been reported to occur in several pathological disorders (for review see (12) and references therein) such as Crohn's disease (CD)(12-19), ulcerative colitis (UC) (11, 14, 15, 17, 20-26), irritable bowel syndrome (IBS) of alternating type (27), colorectal cancer (CRC) (28, 29), and diabetes (30-32).

Many studies have shown the potential role of *F. prausnitzii* in promoting gut health through the secretion of anti-inflammatory compounds such as butyrate (16, 33-36), and in reducing the severity of induced-colitis in mice (16, 37). Despite being a relatively abundant bacteria, capable of regulating gut homeostasis (38, 39) and interacting in several host pathways (40), few studies have paid attention to the distribution of phylotypes within faecalibacteria populations in the human gut. Random Amplified Polymorphic DNA-polymerase chain reaction (RAPD-PCR) fingerprinting of 18 isolates from fecal samples of ten healthy subjects showed that an individual can have up to four different *F. prausnitzii* strains and that these are grouped by individual (35). Besides, 16S rRNA gene analysis of these isolates indicated that each strain has its unique sequence, but they group into two *F. prausnitzii* phylogroups that have a 97% 16S rRNA gene sequence similarity. These two phylogroups coexist in healthy individuals (35) and comprise approximately 97% of *F. prausnitzii* 16S rRNA

sequences found in feces (10, 41). However, it still remains to be elucidated how many different *F. prausnitzii* are hosted by patients with gut diseases, and it is still unknown if the *F. prausnitzii* population of patients suffering intestinal disorders differs from that found in healthy subjects.

This work describes the *F. prausnitzii* populations present in inflammatory bowel disease (IBD), CRC and IBS patients by using a species-specific polymerase chain reaction (PCR) followed by denaturing gradient gel electrophoresis (DGGE) and then sequencing bands. The main objective therefore was to determine if phyloytpe profiles correlated with certain intestinal disorders. We have also investigated whether or not certain phylotypes are associated with patients' clinical characteristics in order to reveal potential biomarkers for diagnosis support and/or prognosis.

MATERIALS AND METHODS

Patients, clinical data and sampling.

A Spanish cohort consisting of 118 volunteers (36 CD, 23 UC, 6 IBS, 22 CRC, and 31 control (H) subjects) was included (Table 1). Subjects were recruited by the Gastroenterology Services of the Hospital Universitari Dr. Josep Trueta (Girona, Spain) and the Hospital Santa Caterina (Salt, Spain). Patients were gender- and age-matched, except CRC patients who were significantly older than all the other groups ($P<0.001$), and H subjects who were older than those with IBD ($P\leq0.013$). IBD patients were diagnosed according to standard clinical, pathological and endoscopic criteria and categorized as stated in the Montreal classification (42). Rome III criteria was used to diagnose IBS patients (available at http://www.romecriteria.org/criteria/). CRC diagnosis was established by colonoscopy and biopsy, The control group (H) consisted of subjects with normal colonoscopy who underwent this procedure for reasons such as rectorrhagia (N=9), colorectal cancer familial history (N=11), and abdominal pain (N=11). None of the subjects received antimicrobial treatment for at least two months before colonoscopy.

Prior to colonoscopy, patients were subjected to cleansing of the gastrointestinal tract using Casenglicol® following manufacturer's guidelines. During routine colonoscopy, a biopsy sample from transverse colon was collected for each subject, following standard procedures. When it was not technically possible, rectal samples were taken instead, because the mucosa-associated community profile is rather stable along the gut (15, 43). All biopsies were immediately placed in sterile tubes without any buffer and stored at -80 ºC following completion of the whole endoscopic procedure and upon analysis.

Ethical consideration.

This work was approved by the Ethics Committee of Clinical Research of the Hospital Universitari Dr. Josep Trueta (Girona, Spain) and the Institut d'Assistència Sanitària of Girona (Salt, Spain) on 24th February 2009 and 21st April 2009, respectively. Informed consent from the subjects was obtained before enrollment.

Sample treatment and DNA extraction.

Prior to DNA extraction, biopsies were subjected to two mild ultrasound wash cycles to discard transient and loosely attached bacteria as previously reported (15). DNA was extracted using the NucleoSpin® Tissue Kit (Macherey-Nagel GmbH &Co., Germany). The support protocol for Gram positive bacteria and the RNAse treatment step were carried out. Genomic DNA was stored at -80 ºC until use. DNA concentration and purity of the extracts were determined with a NanoDrop ND-100 spectrophotometer (NanoDrop Technologies, USA).

Primers design to study *F. prausnitzii* population.

A conventional PCR assay consisting of a species-specific primer set targeting the 16S rRNA gene was designed. 16S rRNA gene sequences from *F. prausnitzii* and from other Ruminococcaceae (Table S1) were recovered from GenBank and aligned using Clustal W (44). Specific primers targeting DNA regions exclusive for *F. prausnitzii* were manually designed and further checked using the softwares Primer Express® version 3.0 software (Applied Biosystems, Foster City, CA, USA) and NetPrimer® software (available at http://premierbiosoft.com/netprimer, PREMIER Biosoft International, California, USA) to check for primer-dimer structures, hairpins and possible cross dimer interactions between oligonucleotides.

The final primer set designed in this study was: Fpra 427F 5'-TAAACTCCTGTTGTTGAGGAAGAT-3' and Fpra 1127R 5'-

138 TTTGTCAACGGCAGTCYKG-3', flanking a ~700-bp fragment that includes variable
139 regions V3 to V6.

140 Oligonucleotides specificity was tested *in silico* by comparing against the
141 Ribosomal Database Project II (45) and GenBank database through Seqmatch and
142 BLAST (46) tools, respectively. Additionally, an *in vitro* inclusivity/exclusivity test was
143 performed (see details in Supplemental Materials, Table S2). Primer set coverage was
144 evaluated using the SILVA TestPrime 1.0 (available at http://www.arb-
145 silva.de/search/testprime/). A sensitivity test to determine the detection limit of the
146 reaction (i.e. the lowest concentration at which 95% of the positive samples are detected
147 (47)) was performed (Supplemental Materials).

148 **PCR amplification and DGGE fingerprinting.**

149 For *F. prausnitzii* population profiling, the 16S rRNA gene was partially
150 amplified from extracted genomic DNA using the new primer set designed in this study.
151 To obtain suitable PCR products for DGGE analysis, a GC-rich sequence was attached
152 to the 5' end of the forward primer: 5'-
153 CGCCCGCCGCGCCCCGCGCCCGGCCCGCCGCCCCCGCCCC-3' according to
154 Muyzer et al. (48). All primers were obtained from Metabion (Steinkirchen, Germany).

155 PCR reaction was optimized testing different concentrations of $MgCl_2$ (0.25 to
156 15 mM), deoxyribonucleoside triphosphates (dNTPs; 0.1 to 3 mM), and primers (50 to
157 900 nM of each primer). The optimized reaction was used in all samples and was
158 composed of: 1× of buffer (II) 10× (Applied Biosystems, Foster City, California),
159 2.5 mM of $MgCl_2$ (25 mM; Applied Biosystems), 0.2 mM of dNTPs (10 mM; Applied
160 Biosystems), 0.2 µM of GC-Fpra 427F primer, 0.2 µM of Fpra 1127R (10 pmol/mL
161 each), 0.05 U/µL of AmpliTaq DNA polymerase (5 U/mL, Applied Biosystems), and 1
162 µL of genomic DNA as a template in a total volume of 50 µL.

All PCR were performed in a GeneAmp® PCR System 2700 thermocycler (Applied Biosystems, Perkin-Elmer®, CA, USA). PCR conditions were optimized by testing different annealing temperatures (52ºC to 70ºC). The optimized cycling program was used, and consisted of: 10 minutes at 95ºC for initial denaturation and DNA polymerase activation, followed by 35 cycles of 30 seconds at 95ºC (denaturation), 30 seconds at 65ºC (annealing), and 1 minute at 72ºC (extension), with a final extension of 10 minutes at 72ºC. Products were visualized under UV light after gel electrophoresis on 1.5% (wt/vol) agarose gels in 1× TAE buffer (Tris-acetate-EDTA, pH 8.0) stained with ethidium bromide (0.5 µg/mL).

DGGE was carried out with an Ingeny-phorU2 system (Ingeny®, Goes, The Netherlands) in 6% (w/v) acrylamide gels, with a vertical denaturing gradient ranging from 30% to 70% urea/formamide. Electrophoreses were run in 0.5× TAE buffer at 60ºC and at a constant voltage of 120 V for a minimum of 16 hours. Gels were then stained with 1× SYBR Gold® (Molecular Probes Europe, Invitrogen™) for 45 minutes in the dark, visualized under UV light and photographed.

Sequencing, sequence editing and analyses.

All detectable bands were excised from the gel and DNA extracted as previously reported (15). DNA was then reamplified by PCR as described above, using the forward primer Fpra 427F without the GC clamp. Positive PCR products were cleaned and sequenced in both directions (forward and reverse) by Macrogen Inc (Seoul, Korea). Prior to analysis, sequences chromatograms were manually inspected and none presented double peaks. Quality of sequences was also checked with Sequence Scaner Software2 (Applied Biosystems, Foster City, CA, USA). Only those with high trace score (which is the average basecall quality value of bases in the post-trim sequence) were considered for further analyses. Sequences were assembled to obtain high-quality

consensus sequences which were further manually refined using BioEdit Sequence Alignment Editor version 7.0.9.0 (49). The presence of chimeras was subsequently checked using DECIPHER (50).

To validate the accuracy of sequences, two biopsies from the same patient were analysed as previously described, and identical consensus sequences were obtained from the bands recovered (data not shown).

Consensus sequences were compared against NCBI database by using the BLASTN search tool (46) in order to determine the closest previously reported sequence and isolate. An alignment of the consensus sequences was performed using Clustal W (44), with manual curation if needed. For further analysis sequences of 581 nucleotides in length (positions 525-1106 in *Escherichia coli* numbering, V3-V6 regions) were used, which corresponded to those positions recovered for all sequences.

A Neighbor-Joining [NJ, (51)] distance matrix using the Jukes-Cantor (JC) correction was calculated in Mothur [http://www.mothur.org, (52)], which was then used to assign sequences to Operational Taxonomic Units (OTUs) using the farthest neighbor method at a cutoff of 0.03 and 0.01 (i.e., 97% and 99% minimum similarity for any pair of sequences to belong to the same phylogroup or phylotype, respectively). Representative sequences for each OTU (named thereafter OTU97 and OTU99, respectively) were identified and used for further analyses of distribution by group of patients. We will refer to the OTUs defined at >99% sequence identity as “phylotypes” and those defined at >97% as “phylogroups”. Unique sequences (100% similarity, named thereafter OTU100) were also considered to compare sequence-based population composition and to study the richness by groups of patients.

Nucleotide sequence accession numbers

The 16S rDNA sequences obtained in this study were deposited in the GenBank/EMBL/DDBJ database under the accession numbers KP005458 through KP005741.

Statistics

Clustering of all samples was performed according to the presence/absence of unique sequences with a hierarchical analysis by the inter-groups joining method based on the Dice coefficient. This same analysis was performed for group of patients clustering according to OTU99 and OTU97 composition.

The non-parametric Kruskal–Wallis and the Mann–Whitney U tests were used to test differences in richness in variables with more than two categories and for pairwise comparisons, respectively according to subject diagnostics, IBD subtype, activity status (active CD and UC patients when CDAI > 150 (53) and a Mayo score >3, respectively), smoking habit, intestinal resection, medication and tumour state for CRC patients.

Pearson's X^2 test was carried out to compare the prevalence of OTUs defined at a 99% and at 97% of similarity, taking into account the same clinical variables in order to determine specific OTUs of each condition.

All statistical analyses were conducted via SPSS 15.0 (SPSS Inc., Chicago, IL). Significance levels were established for P values ≤ 0.05.

RESULTS

Features of the novel PCR-DGGE assay to study *F. prausnitzii* population.

In this study, a novel oligonucleotide set was designed to specifically amplify *F. prausnitzii* (Supplemental Materials). The detection limit of the reaction was 2623 target genes, which indicates that if we had many replicate samples with 2623 copies of 16S rRNA gene of *F. prausnitzii*, no more than 5% failed reactions should occur. The banding pattern obtained by DGGE represents the major constituents of the analyzed community (54). Species that contribute <1% of the total population would not be readily detected by this molecular approach (48). This is expected to have a minor effect in the present study however because *Faecalibacterium prausnitzii* represents at least a 6% of the mucosa-associated microbial community in healthy subjects (1, 4-11).

The *in silico* analysis of the oligonucleotide set chosen showed that primer Fpra427F was specific for *F. prausnitzii* and targeted all the isolates, whereas the Fpra1127R primer was genus-specific. Coverage of the Fpra 427F-Fpra1127R primer set was of 70.6% of the *Faecalibacterium* sp. sequences in the SILVA dataset. The remaining 29.4% should be regarded as sequences of this genus but not *F. prausnitzii* (probably other species that have not yet been described exist within this genus).

Inclusivity-exclusivity tests were conducted *in vitro* by testing DNA from nine *F. prausnitzii* isolates and from 71 additional representative bacterial species (Table S2, Supplemental Materials) confirmed that the PCR was totally specific. All the *F. prausnitzii* isolates were successfully detected thus producing a single PCR product, and there was no cross-reaction with any of the non-target microorganisms (Supplemental Materials).

Sequence analysis and *F. prausnitzii* population composition

Applied and Environmental Microbiology

From the samples from the 118 volunteers engaged in the study, a total of 284 partial *F. prausnitzii* 16S rRNA gene sequences (66 from H, 17 from IBS, 48 from UC, 97 from CD and 56 from CRC) were obtained from PCR-DGGE gels. This represented an 88.75% of the observed bands, whereas the remaining (i.e. 11.25% of the observed bands) were not recovered due to methodological problems (i.e. not proper band excision from gel, incapability to re-amplify the bands for further sequencing) and thus were not included in the analysis. Among the retrieved sequences, 135 unique sequences were found, that could be grouped into 40 OTUs based on a >99% sequence similarity criteria and 5 OTUs based on >97% of sequence similarity criteria (Tables S3, S4 and S5). Correspondence between OTUs at different cutoffs of similarity is shown in Table S6 (i.e. which OTU100 are included within a given OTU97 and OTU99).

F. prausnitzii population similarity and richness (OTU100 analysis)

Sequences were grouped in 135 unique sequences (i.e. OTU100; Table S3), and analyzed in order to compare sequence-based population composition and to estimate *F. prausnitzii* richness between groups of patients.

Clustering analysis of OTU100 hosted by each subject revealed that *F. prausnitzii* populations were rather individual-specific as few patients featured identical population composition. Furthermore, 24.6% of the subjects had *F. prausnitzii* population composition very different from that found in any other patient studied, thus featuring their own branching point in the cluster analysis (Figure 1). Noticeably, these were mainly H subjects (54.8%), whereas only 16.7% of IBS, 21.7% of UC, 4.5% of CRC, and 22.2% of CD patients had particular *F. prausnitzii* population profiles composed by a set of sequences not found in any other patient (P=0.001). Within CD patients, over 54% of C-CD featured unique *F. prausnitzii* populations, but this was

only found in approximately 10% of patients with ileal disease location (either I-CD or IC-CD) (P=0.025).

The remaining 75.4% of the subjects grouped into five main clusters. Interestingly, no H subjects were grouped in cluster III, which included over 30% of IBD and CRC patients (P=0.005). This clustering was not explained by any other patient variable tested (age, gender, smoking habit, disease activity index, age at the onset of disease, intestinal resection, and medication).

As concerns to *F. prausnitzii* richness, all volunteers had populations ranging between one and five unique sequences of this species. The average number of unique sequences of *F. prausnitzii* per subject was lower in IBD patients (1.7±0.8 in UC and 1.6±0.8 in CD) than in H, IBS and CRC subjects (2.1±1.1 in H, 2.2±1.0 in IBS and 2.5±1.2 in CRC) (P=0.064). Clinical data of patients did not explain differences in the number of unique sequences found.

Percentage of subjects with three or more *F. prausnitzii* unique sequences was higher in H, IBS and CRC groups in comparison with IBD patients (P=0.027) (Figure 2). These data suggested that IBD patients featured less *F. prausnitzii* microdiversity whereas CRC and IBS patients presented a higher number of *F. prausnitzii* unique subtypes in the gut, being similar to numbers found in H subjects.

F. prausnitzii population composition at the phylotype level (OTU99 analysis).

For this analysis all the samples from patients with the same diagnostics were analyzed together in order to compare the *F. prausnitzii* populations hosted by H subjects and patients with intestinal disorders. We will refer to the 40 OTUs defined at >99% sequence identity as “phylotypes”. All volunteers had populations ranging between one and four phylotypes of this species, but no significant differences in the average number of OTU99 per patient between groups were observed (P=0.558). In H

subjects 20 different phylotypes were recovered, however in IBS, UC, CD and CRC patients values were lower (7, 9, 16 and 14 respectively) (Figure 3). The highest complexity of the community found in H subjects was confirmed from estimates of the Shannon diversity index (H'_H=2.24), whereas the lowest diversity index was calculated in CD patients (H'_{CD}=1.90) in spite of having more subjects with this disorder in the cohort analyzed. Besides, the other gut disorders also presented lower diversity index values than H subjects (H'_{IBS}=1.49, H'_{UC}=1.58, H'_{CRC}=1.83).

Cluster analysis of phylotypes by patient group revealed that those with IBD and CRC host different *F. prausnitzii* populations when compared to H subjects (Figure 3). The most prevalent phylotype (OTU99_1) accounted for approximately 20% of the sequences in H and IBS subjects, and about 40% in IBD and CRC patients (P=0.002) (Figure 3, Table S4). The next three most prevalent OTUs (OTU99_2, OTU99_3 and OTU99_4), were also detected in all the patient groups (Figure 3, Table S4), representing between 8% and 50% of the sequences depending on the patients group, but without reaching statistically significant differences in their prevalence between groups of patients. Twelve OTUs (representing 54.8% of the sequences recovered from H) were exclusive of the H group. On the other hand, 22.2% of sequences from CD patients (seven OTUs) were not shared with patients with other intestinal disorders. Similarly, four UC and four CRC specific OTUs were also found, accounting for the 17.4% and the 18.2% of sequences from each group of patients respectively. Finally, OTU99_17 accounted for 16.7% of IBS sequences and was only found in this group of patients.

Altogether these observations suggest that the most prevalent *F. prausnitzii* phylotypes (OTU99_1 throught OTU99_4) are present in both H and diseased patients, but that there existed rare phylotypes not found in all the patients (from OTU99_5 to

Applied and Environmental Microbiology

OTU99_40), which could have emerged or disappeared in certain gut conditions, since they are disease-specific or exclusively found in healthy subjects.

F. prausnitzii phylogroup distribution by group of patients (OTU97 analysis)

Of the total 284 sequences recovered, 279 sequences grouped together in two main OTU97 which corresponded to the previously defined phylogroups I and II (35) (Figure 4, Table S5). Thus, we will refer to the OTUs defined at >97% sequence identity as "phylogroups. Phylogroup II (OTU97_1) consisted of 56.7% of the total sequences recovered (corresponding to 161 sequences) whereas phylogroup I (OTU97_2) was slightly less prevalent (41.55%, 118 sequences). The remaining five sequences, representing less than 2% of sequences recovered, were grouped into three different OTUs: a tripleton with two CD and one CRC sequences (OTU97_3), a singleton from an IBS patient (OTU97_4), and a singleton from an H subject (OTU97_5). Noticeably, OTU97_5 presented 95% of similarity to the nearest sequence present in databases (Table S5).

All subjects had populations ranging between one and three phylogroups of *F. prausnitzii*. Significant differences in the average number of OTU97 per patient between groups were not found (P= 0.285). However, phylogroups I and II prevalences differed between groups of patients (Figure 4). More than 52% of the sequences from H and IBS subjects belonged to phylogroup I, whereas most of the sequences from UC, CD and CRC patients belonged to phylogroup II (62.5%, 66% and 62.5% respectively) (P=0.001), suggesting that phylogroup I is more frequently compromised in the latter. Cluster analysis of phylotgroups by patient group also revealed that those with IBD and CRC host different *F. prausnitzii* populations when compared to H subjects at this cut-off level (Figure 4).

Co-occurrence analysis of both phylogroups (Figure 5) showed that only in 26.1% of UC and 22.2% of CD presented both phylogroups simultaneously while more than 38% of H, IBS and CRC subjects had sequences from both phylogroups, although differences did not reach statistical significance (P=0.270). It is of note that no I-CD patients had simultaneously both phylogroups (P=0.060).

These results suggest that patients with intestinal disorders feature an altered prevalence of phylogroups, mostly characterized by the presence of mono-phylogroup populations in some IBD patients, especially those with I-CD.

Putative indicator sequences for differential diagnosis and/or disease prognostics.

We observed differences in phylogroup and phylotype prevalence between groups of patients. Therefore, we further explored which OTUs could be considered as potential biomarkers for disease diagnosis. The prevalence of four OTU99 was statistically different between the groups of patients (P≤0.027) (Table 2). Particularly, OTU99_1 was found to be more prevalent in IBD and CRC patients than in H subjects (P=0.020), OTU99_10 was found in 16.7% of IBS patients and 9.1% CRC patients, but not in H nor in IBD subjects (P=0.027), and OTU99_11 and OTU99_17 were exclusively found in a 16.7% of IBS patients (P=0.001).

Several OTU99 were found to be associated with clinical characteristics of the diseased status. Within IBD patients, OTU99_24 was exclusively found in a 33% of UC inactive patients (P=0.015) whereas OTU99_8 was especially absent in active CD patients (P=0.008). Remarkably, OTU99_8 and OTU99_4 were found in all CRC patients featuring the most severe tumor state (T4) (P<0.001 and P=0.033 respectively). Moreover, OTU99_7 was not found in CRC patients with tumour states T3 and T4 but

it was carried by all CRC patients with T1, a 12.5% of T2, and a 5.6% of CD patients (P=0.013).

Referring to phylogroups (OTU97), the 87.1% of H subjects and the 83.3% of IBS patients had phylogroup I, whereas this value was reduced to 63.6% in CRC, 47.8% in UC and 43.2% in CD patients (P=0.003) (Table 2). In contrast, a higher prevalence of phylogroup II was observed in IBD and CRC patients (78.3% in UC, 83.8% in CD and 90.9% in CRC) in comparison to H and IBS subjects (51.6% and 66.7% respectively; P=0.005). No further differences in relation to patients' clinical characteristics were observed at the phylogroups level.

Applied and Environmental Microbiology

DISCUSSION

In this study we describe for the first time the richness and the distribution of phylotypes and phylogroups of *F. prausnitzii*. Our data reveal that *F. prausnitzii* populations are rather individual-specific, especially in H subjects, in line with previous evidences from *F. prausnitzii* isolates which grouped by individual according to their 16S rRNA gene sequence and their RAPD-PCR fingerprint (35). Despite the fact that we have observed that *F. prausnitzii* populations can include up to five different sequences per individual, we found that in over 87% of IBD patients their populations consisted of just one or two different *F. prausnitzii* sequences. In contrast, at least 30% of H, IBS and CRC subjects hosted populations with more than three different sequences. Multiple 16S rRNA gene copies, with slight sequence variation can be hosted by a bacterial species, which can affect bacterial community analyses (55). According to data on *rrn*DB (56) *F. prausnitzii* S3L/3 has a single copy of the 16S rRNA gene which makes our results more meaningful. To date only one genome of this species has been fully sequenced and annotated, therefore the possibility that future studies reveal that several copies of 16S rRNA gene are hosted by other members of this species cannot be ruled out. This will reduce the number of unique phylotypes per patient observed. In our study, no differences in the average number of unique sequences per patient have been observed between active and inactive IBD patients, IBD or CRC subtypes, intestinal resection, medication or smoking habit, suggesting that richness remains reduced over time even if there is endoscopic and clinical remission, and regardless of the treatment used.

BLAST analysis revealed that the sequences recovered were of high similarity to previously characterised strains of this species. Only two of the sequences featured ≤95% similarity with previously found *F. prausnitzii*. This finding suggests that novel

and rare phylotypes of *F. prausnitzii* are yet to be retrieved by cultivation techniques. Sequences for both phylogroups I and II,(35) were detected in all groups of patients; but in some patients, especially IBD, we detected only one of the two main phylogroups. In general all patients suffering from an intestinal disorder exhibited a reduction in phylotype richness which was not recovered during periods of remission or in mild states of the disease, suggesting that alterations in this population struggle to normalise with the current patient's treatments. New therapies to recover all the diversity of *F. prausnitzii* in these patients should be considered.

Differences in phylotype and phylogroup prevalence between diagnostics allowed us to discriminate patients suffering intestinal disease, especially those with IBD and CRC, from H subjects. This finding is in agreement with previous studies reporting that phylotypes related to isolate M21/2 (phylogroup I) and/or isoalte A2-165 (phylogroup II) are depleted in CD patients when compared to H subjects (15, 22). As these results might indicate differences in abundance of these phylogroups, it would be of interest to conduct quantitative analyses in order to explore their usefulness as biomarkers for intestinal disorders diagnostics and/or prognostics. Additionally, further investigation addressing the role of disease-specific sequences and the absence of H-associated phylotypes in gut disease patients could shed light on how *F. prausnitzii* can contribute to/prevent gut diseases pathogenesis.

Our data have revealed that H subjects and those with intestinal disorders host distinctive *F. prausnitzii* populations. This discrimination can be mainly attributable to differences in the prevalence of the common *F. prausnitzii* phylotypes, but also to the presence of rare OTUs found specifically in each group of patients. These differences in presence/absence of a specific phylotype can be explained by different capabilities to adapt to a diseased gut environment or mucosa status. For instance, it has been reported

that CD patients often have acidic stools, with elevated bile salt concentrations (57, 58) and also other factors such as oxidative stress or thiol availability are likely to lead to a significantly altered microbiota (59, 60). Supporting this hypothesis, *F. prausnitzii* representatives are sensitive to changes in the gut physico-chemical conditions that may occur in disease status, such as pH reduction or bile salt content (35), and are extremely sensitive to oxygen although they can persist in oxygen environments due to a flavin-thiol dependent extracellular electron shuttle (61). These observations should be taken into account for future treatment strategies aiming to restore *F. prausnitzii* population in patients suffering intestinal disorders. Novel treatments like prebiotics could be a strategy in order to boost the remaining *F. prausnitzii* populations in the gut disorders. The treatment strategies based on *F. prausnitzii* as probiotics using strains that are more tolerant to withstand gut conditions in intestinal disturbances could also be a suitable approach.

The observed imbalance in the prevalence of phylogroups between subjects provides an additional view to understand the role of this species in IBD since differences in their prevalence between healthy and IBD subgroups must have biological relevance, eg. reflecting differential responses to the host environment. Currently there is no phenotypic trait that consistently distinguishes *F. prausnitzii* members from one or other phylogroup (35), but the existence of differences to be present in members of different phylogroups would not be surprising, for instance due to horitzontal gene transfer. Phenotypic differences have been observed between isolates from different phylogroups (which also have their characteristic 16S rRNA sequence) with respect to their capability to use carbohydrates from diet and/or host-derived, as well as their tolerance to bile salts and pH (35). It can be hypothesised that differences in phylogroup composition reflect variations in sensitivity to such

environmental factors or to interactions with the host and it has been demonstrated that *F. prausnitzii* ATCC2768 (phylogroup I) and *F. prausnitzii* A2-165 (phylogroup II) are linked with the modulation of different urinary metabolites related to different host pathways (40). Our study does not allow deciphering the biological relevance of the observed changes in population composition, but points out that the *F. prausnitzii* population hosted by different groups of subjects is different, and further studies based on isolation and characterisation of *F. prausnitzii* from subjects with these disorders and controls would be interesting to address this question.

CONCLUSION

Although members of the *F. prausnitzii* population are shared between healthy subjects and those with gut diseases, there is a loss of richness in IBD patients and a different distribution of specific phylotypes. The imbalance in phylogroups (OTU97_1 and OTU97_2), and abundance of specific phylotypes can be used as biomarkers to distinguish some intestinal diseases as IBD or CRC.

ACKNOWLEDGEMENTS

This work was partially funded by the Spanish Ministry of Education and Science through project SAF2010-15896. Mireia Lopez-Siles was recipient of an FI grant from the Generalitat de Catalunya (2010FI_B2 00135), which receives support from the European Union Commissionate. Prof. Harry J. Flint and Dr. Sylvia H. Duncan acknowledge support from the Scottish Government Food, Land and People programme. We thank PhD. Teresa Mas-de-Xaxars for assistance in CRC samples recruitment, Ms Natàlia Adell from the Serveis Tècnics de Recerca for statistical assistance, Mr. Pau Boher for assistance with heat maps analysis, and Dr. Marc Llirós and Ms. Carla Camprubí-Font for critically revising the manuscript. We appreciate the

generosity of the patients who freely gave their time and samples to make this study possible, and the theatre staff of all centers for their dedication and careful sample collection.

REFERENCES

1. **Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA.** 2005. Diversity of the human intestinal microbial flora. Science **308:**1635-1638.
2. **Sekirov I, Russell SL, Antunes LC, Finlay BB.** 2010. Gut microbiota in health and disease. Physiol Rev **90:**859-904.
3. **Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Jian M, Zhou Y, Li Y, Zhang X, Qin N, Yang H, Wang J, Brunak S, Dore J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, Bork P, Ehrlich SD.** 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature **464:**59-65.
4. **Nava GM, Stappenbeck TS.** 2011. Diversity of the autochthonous colonic microbiota. Gut Microbes **2:**99-104.
5. **Baumgart M, Dogan B, Rishniw M, Weitzman G, Bosworth B, Yantiss R, Orsi RH, Wiedmann M, McDonough P, Kim SG, Berg D, Schukken Y, Scherl E, Simpson KW.** 2007. 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 **1:**403-418.
6. **Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S, Dore J, Antolin M, Artiguenave F, Blottiere HM, Almeida M, Brechot C, Cara C, Chervaux C, Cultrone A, Delorme C, Denariaz G, Dervyn R, et al.** 2011. Enterotypes of the human gut microbiome. Nature **473:**174-180.
7. **Hold GL, Schwiertz A, Aminov RI, Blaut M, Flint HJ.** 2003. Oligonucleotide probes that detect quantitatively significant groups of butyrate-producing bacteria in human feces. Appl Environ Microbiol **69:**4320-4324.
8. **Schwiertz A, Jacobi M, Frick JS, Richter M, Rusch K, Kohler H.** 2010. Microbiota in pediatric inflammatory bowel disease. J Pediatr **157:**240-244
9. **Suau A, Rochet V, Sghir A, Gramet G, Brewaeys S, Sutren M, Rigottier-Gois L, Doré J.** 2001. *Fusobacterium prausnitzii* and related species represent a dominant group within the human fecal flora. Systematic and Applied Microbiology **24:**139-145.
10. **Walker AW, Ince J, Duncan SH, Webster LM, Holtrop G, Ze X, Brown D, Stares MD, Scott P, Bergerat A, Louis P, McIntosh F, Johnstone AM, Lobley GE, Parkhill J, Flint HJ.** 2011. Dominant and diet-responsive groups of bacteria within the human colonic microbiota. ISME J**:**220-230.
11. **Swidsinski A, Loening-Baucke V, Lochs H, Hale LP.** 2005. Spatial organization of bacterial flora in normal and inflamed intestine: a fluorescence in situ hybridization study in mice. World J Gastroenterol **11:**1131-1140.

12. **Miquel S, Martin R, Rossi O, Bermudez-Humaran L, Chatel J, Sokol H, Thomas M, Wells J, Langella P.** 2013. *Faecalibacterium prausnitzii* and human intestinal health. Curr Opin Microbiol **16:**255-261.
13. **Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR.** 2007. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci U S A **104:**13780-13785.
14. **Lopez-Siles M, Martinez-Medina M, Busquets D, Sabat-Mir M, Duncan SH, Flint HJ, Aldeguer X, Garcia-Gil LJ.** 2014. Mucosa-associated *Faecalibacterium prausnitzii* and *Escherichia coli* co-abundance can distinguish Irritable Bowel Syndrome and Inflammatory Bowel Disease phenotypes. Int J Med Microbiol **304:**464-475.
15. **Martinez-Medina M, Aldeguer X, Gonzalez-Huix F, Acero D, Garcia-Gil LJ.** 2006. 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 **12:**1136-1145.
16. **Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermudez-Humaran LG, Gratadoux JJ, Blugeon S, Bridonneau C, Furet JP, Corthier G, Grangette C, Vasquez N, Pochart P, Trugnan G, Thomas G, Blottiere HM, Dore J, Marteau P, Seksik P, Langella P.** 2008. *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. Proc Natl Acad Sci U S A **105:**16731-16736.
17. **Sokol H, Seksik P, Furet JP, Firmesse O, Nion-Larmurier I, Beaugerie L, Cosnes J, Corthier G, Marteau P, Dore J.** 2009. Low counts of *Faecalibacterium prausnitzii* in colitis microbiota. Inflamm Bowel Dis **15:**1183-1189.
18. **Swidsinski A, Loening-Baucke V, Vaneechoutte M, Doerffel Y.** 2008. Active Crohn's disease and ulcerative colitis can be specifically diagnosed and monitored based on the biostructure of the fecal flora. Inflamm Bowel Dis **14:**147-161.
19. **Willing B, Halfvarson J, Dicksved J, Rosenquist M, Jarnerot G, Engstrand L, Tysk C, Jansson JK.** 2009. Twin studies reveal specific imbalances in the mucosa-associated microbiota of patients with ileal Crohn's disease. Inflamm Bowel Dis **15:**653-660.
20. **Furet JP, Kong LC, Tap J, Poitou C, Basdevant A, Bouillot JL, Mariat D, Corthier G, Dore J, Henegar C, Rizkalla S, Clement K.** 2010. Differential adaptation of human gut microbiota to bariatric surgery-induced weight loss: links with metabolic and low-grade inflammation markers. Diabetes **59:**3049-3057.
21. **Hansen R, Russell RK, Reiff C, Louis P, McIntosh F, Berry SH, Mukhopadhya I, Bisset WM, Barclay AR, Bishop J, Flynn DM, McGrogan P, Loganathan S, Mahdi G, Flint HJ, El-Omar EM, Hold GL.** 2012. Microbiota of de-novo pediatric IBD: increased *Faecalibacterium prausnitzii* and reduced bacterial diversity in Crohn's but not in ulcerative colitis. Am J Gastroenterol **107:**1913-1922.
22. **Jia W, Whitehead RN, Griffiths L, Dawson C, Waring RH, Ramsden DB, Hunter JO, Cole JA.** 2010. Is the abundance of *Faecalibacterium prausnitzii* relevant to Crohn's disease? FEMS Microbiol Lett **310:**138-144.
23. **Kabeerdoss J, Sankaran V, Pugazhendhi S, Ramakrishna BS.** 2013. *Clostridium leptum* group bacteria abundance and diversity in the fecal

microbiota of patients with inflammatory bowel disease: a case-control study in India. BMC Gastroenterol **13:**20.

24. **Machiels K, Joossens M, Sabino J, De Preter V, Arijs I, Eeckhaut V, Ballet V, Claes K, Van Immerseel F, Verbeke K, Ferrante M, Verhaegen J, Rutgeerts P, Vermeire S.** 2013. A decrease of the butyrate-producing species *Roseburia hominis* and *Faecalibacterium prausnitzii* defines dysbiosis in patients with ulcerative colitis. Gut.

25. **McLaughlin SD, Clark SK, Tekkis PP, Nicholls RJ, Ciclitira PJ.** 2010. The bacterial pathogenesis and treatment of pouchitis. Therap Adv Gastroenterol **3:**335-348.

26. **Vermeiren J, Van den Abbeele P, Laukens D, Vigsnaes LK, De Vos M, Boon N, Van de Wiele T.** 2012. Decreased colonization of fecal *Clostridium coccoides/Eubacterium rectale* species from ulcerative colitis patients in an in vitro dynamic gut model with mucin environment. FEMS Microbiol Ecol **79:**685-696.

27. **Rajilic-Stojanovic M, Biagi E, Heilig HG, Kajander K, Kekkonen RA, Tims S, de Vos WM.** 2011. Global and deep molecular analysis of microbiota signatures in fecal samples from patients with irritable bowel syndrome. Gastroenterology **141:**1792-1801.

28. **Balamurugan R, Rajendiran E, George S, Samuel GV, Ramakrishna BS.** 2008. Real-time polymerase chain reaction quantification of specific butyrate-producing bacteria, *Desulfovibrio* and *Enterococcus faecalis* in the feces of patients with colorectal cancer. J Gastroenterol Hepatol **23:**1298-1303.

29. **Sobhani I, Tap J, Roudot-Thoraval F, Roperch JP, Letulle S, Langella P, Corthier G, Tran Van Nhieu J, Furet JP.** 2011. Microbial dysbiosis in colorectal cancer (CRC) patients. PLoS One **6:**e16393.

30. **de Goffau MC, Luopajarvi K, Knip M, Ilonen J, Ruohtula T, Harkonen T, Orivuori L, Hakala S, Welling GW, Harmsen HJ, Vaarala O.** 2013. Fecal microbiota composition differs between children with beta-cell autoimmunity and those without. Diabetes **62:**1238-1244.

31. **Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, Peng Y, Zhang D, Jie Z, Wu W, Qin Y, Xue W, Li J, Han L, Lu D, Wu P, Dai Y, Sun X, Li Z, Tang A, Zhong S, Li X, Chen W, Xu R, Wang M, Feng Q, Gong M, Yu J, Zhang Y, Zhang M, Hansen T, Sanchez G, Raes J, Falony G, Okuda S, Almeida M, LeChatelier E, Renault P, Pons N, Batto JM, Zhang Z, Chen H, Yang R, Zheng W, Yang H, Wang J, et al.** 2012. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature **490:**55-60.

32. **Karlsson FH, Tremaroli V, Nookaew I, Bergstrom G, Behre CJ, Fagerberg B, Nielsen J, Backhed F.** 2013. Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature **498:**99-103.

33. **Barcenilla A, Pryde SE, Martin JC, Duncan SH, Stewart CS, Henderson C, Flint HJ.** 2000. Phylogenetic relationships of butyrate-producing bacteria from the human gut. Appl Environ Microbiol **66:**1654-1661.

34. **Duncan SH, Hold GL, Harmsen HJ, Stewart CS, Flint HJ.** 2002. Growth requirements and fermentation products of *Fusobacterium prausnitzii*, and a proposal to reclassify it as *Faecalibacterium prausnitzii* gen. nov., comb. nov. Int J Syst Evol Microbiol **52:**2141-2146.

35. **Lopez-Siles M, Khan TM, Duncan SH, Harmsen HJ, Garcia-Gil LJ, Flint HJ.** 2012. Cultured representatives of two major phylogroups of human colonic

Faecalibacterium prausnitzii can utilize pectin, uronic acids, and host-derived substrates for growth. Appl Environ Microbiol **78:**420-428.

36. **Louis P, Flint HJ.** 2009. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. FEMS Microbiol Lett **294:**1-8.

37. **Martin R, Chain F, Miquel S, Lu J, Gratadoux JJ, Sokol H, Verdu EF, Bercik P, Bermudez-Humaran LG, Langella P.** 2014. The commensal bacterium *Faecalibacterium prausnitzii* is protective in DNBS-induced chronic moderate and severe colitis models. Inflamm Bowel Dis **20:**417-430.

38. **Carlsson AH, Yakymenko O, Olivier I, Hakansson F, Postma E, Keita AV, Soderholm JD.** 2013. *Faecalibacterium prausnitzii* supernatant improves intestinal barrier function in mice DSS colitis. Scand J Gastroenterol **48:**1136-1144.

39. **Wrzosek L, Miquel S, Noordine ML, Bouet S, Chevalier-Curt MJ, Robert V, Philippe C, Bridonneau C, Cherbuy C, Robbe-Masselot C, Langella P, Thomas M.** 2013. *Bacteroides thetaiotaomicron* and *Faecalibacterium prausnitzii* influence the production of mucus glycans and the development of goblet cells in the colonic epithelium of a gnotobiotic model rodent. BMC Biol **11:**61.

40. **Li M, Wang B, Zhang M, Rantalainen M, Wang S, Zhou H, Zhang Y, Shen J, Pang X, Wei H, Chen Y, Lu H, Zuo J, Su M, Qiu Y, Jia W, Xiao C, Smith LM, Yang S, Holmes E, Tang H, Zhao G, Nicholson JK, Li L, Zhao L.** 2008. Symbiotic gut microbes modulate human metabolic phenotypes. Proc Natl Acad Sci U S A **105:**2117-2122.

41. **Tap J, Mondot S, Levenez F, Pelletier E, Caron C, Furet JP, Ugarte E, Munoz-Tamayo R, Paslier DL, Nalin R, Dore J, Leclerc M.** 2009. Towards the human intestinal microbiota phylogenetic core. Environ Microbiol **11:**2574-2584.

42. **Silverberg MS, Satsangi J, Ahmad T, Arnott ID, Bernstein CN, Brant SR, Caprilli R, Colombel JF, Gasche C, Geboes K, Jewell DP, Karban A, Loftus Jr EV, Pena AS, Riddell RH, Sachar DB, Schreiber S, Steinhart AH, Targan SR, Vermeire S, Warren BF.** 2005. 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 **19 Suppl A:**5-36.

43. **Zoetendal EG, von Wright A, Vilpponen-Salmela T, Ben-Amor K, Akkermans AD, de Vos WM.** 2002. Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. Appl Environ Microbiol **68:**3401-3407.

44. **Thompson JD, Higgins DG, Gibson TJ.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **22:**4673-4680.

45. **Maidak BL, Cole JR, Lilburn TG, Parker CT, Jr., Saxman PR, Farris RJ, Garrity GM, Olsen GJ, Schmidt TM, Tiedje JM.** 2001. The RDP-II (Ribosomal Database Project). Nucleic Acids Res **29:**173-174.

46. **Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:**3389-3402.

47. **Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT.** 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem **55:**611-622.
48. **Muyzer G, de Waal EC, Uitterlinden AG.** 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl Environ Microbiol **59:**695-700.
49. **Hall TA.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser **41:**95-98.
50. **Wright ES, Yilmaz LS, Noguera DR.** 2012. DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. Appl Environ Microbiol **78:**717-725.
51. **Saitou N, Nei M.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol **4:**406-425.
52. **Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF.** 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol **75:**7537-7541.
53. **Best WR, Becktel JM, Singleton JW, Kern F, Jr.** 1976. Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study. Gastroenterology **70:**439-444.
54. **Heuer H SK.** 1997. Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) for studying soil microbial communities. *In* van Elsas JD WE, Trevors JT (ed), Modern Soil Microbiology. NY:Marcel Dekker Inc, New York.
55. **Vetrovsky T, Baldrian P.** 2013. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. PLoS One **8:**e57923.
56. **Stoddard SF, Smith BJ, Hein R, Roller BRK, Schmidt TM.** 2014. rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. Nucleic Acids Research.
57. **Nugent SG, Kumar D, Rampton DS, Evans DF.** 2001. Intestinal luminal pH in inflammatory bowel disease: possible determinants and implications for therapy with aminosalicylates and other drugs. Gut **48:**571-577.
58. **Pereira SP, Bain IM, Kumar D, Dowling RH.** 2003. Bile composition in inflammatory bowel disease: ileal disease and colectomy, but not colitis, induce lithogenic bile. Aliment Pharmacol Ther **17:**923-933.
59. **Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, Reyes JA, Shah SA, LeLeiko N, Snapper SB, Bousvaros A, Korzenik J, Sands BE, Xavier RJ, Huttenhower C.** 2012. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol **13:**R79.
60. **Seril DN, Liao J, Yang GY, Yang CS.** 2003. Oxidative stress and ulcerative colitis-associated carcinogenesis: studies in humans and animal models. Carcinogenesis **24:**353-362.
61. **Khan MT, Duncan SH, Stams AJ, van Dijl JM, Flint HJ, Harmsen HJ.** 2012. The gut anaerobe *Faecalibacterium prausnitzii* uses an extracellular electron shuttle to grow at oxic-anoxic interphases. ISME J **6:**1578-1585.

740 **FIGURE LEGENDS**

741 **Figure 1.** Hierarchical distance clustering showing sample relationship based on *F.*
742 *prausnitzii* population as determined by comparing unique sequences (OTU100) by
743 means of the inter-groups joining method based on the Dice coefficient. Scale bar
744 describes similarity between profiles. H, indicates healthy subject; IBS, irritable bowel
745 syndrome; UC, ulcerative colitis; CRC, colorectal cancer; CD, Crohn's disease; ○,
746 colonic CD (C-CD); ●, ileo-colonic CD (IC-CD); and ●, ileal-CD (I-CD). I to V
747 indicate the main groups in which patients have clustered.

748

749 **Figure 2.** Percentage of patients with *F. prausnitzii* populations with one to five unique
750 sequences (OTU100) by group of subjects. H, healthy subject; IBS, irritable bowel
751 syndrome; UC, ulcerative colitis; CD, Crohn's disease; CRC, colorectal cancer.

752

753 **Figure 3.** Heatmap showing the relative abundance of sequences assigned to each
754 operational taxonomic unit at a 99% 16S rRNA gene sequence similarity (OTU 99) by
755 group of subjects. Relative abundace has been calculated as the percentage of sequences
756 in an OTU from the total of sequences recovered in that group of patients. Hierarchical
757 distance clustering showing relationship of OTU 99 and groups of patients based on
758 their distribution has been represented (using the inter-groups joining method based on
759 the Dice coefficient). Scale bar describes similarity between profiles.
760 H, healthy subject; IBS, irritable bowel syndrome; UC, ulcerative colitis; CD, Crohn's
761 disease; CRC, colorectal cancer.

762

763 **Figure 4.** Heatmap showing the relative abundance of sequences assigned to each
764 operational taxonomic unit at a 97% 16S rRNA gene sequence similarity (OTU 97) by

group of subjects. Relative abundace has been calculated as the percentage of sequences in an OTU from the total of sequences recovered in that group of patients. Hierarchical distance clustering showing relationship of OTU 97 and groups of patients based on their distribution has been represented (using the inter-groups joining method based on the Dice coefficient). Scale bar describes similarity between profiles.

H, healthy subject; IBS, irritable bowel syndrome; UC, ulcerative colitis; CD, Crohn's disease; CRC, colorectal cancer.

Figure 5. Prevalence of *F. praunsnitzii* phylogroups (OTU97) in healthy subjects and patients with different intestinal disorders (A) and by Crohn's disease subtype (B). H, healthy subject; IBS, irritable bowel syndrome; UC, ulcerative colitis; CD, Crohn's disease; CRC, colorectal cancer; C-CD, colonic-CD; IC-CD, ileocolonic-CD; I-CD, ileal-CD.

TABLES

Table 1. Sample size and clinical characteristics of subjects.

	Healthy*	Irritable bowel syndrome	IBD		Colorectal cancer	p value§
			Ulcerative colitis	Crohn's disease		
N (patients)	31	6	23	36	22	
Age (mean years ± SD)	49.2±16.3	42.4±11.4	38.4±14.0	34.5±12.8	70.1±10.3	<0.001‡
Male (N, %)	15 (48.4%)	2 (20.0%)	15 (65.2%)	21 (58.3%)	11 (50.0%)	0.538†
Active (N, %)	na	na	17 (73.9%)	23 (63.8%)	na	0.365†
Previous surgery (N, %)	0	nd	2 (8.7%)	7 (19.4%)	nd	0.145†
Smokers (N, %)	0	0	2 (8.7%)	4 (11.1%)	1 (4.5%)	0.386†
Treatment (N, %) **						0.520†
No treatment	na	na	10 (43.5%)	11 (30.6%)	na	
Mesalazine	na	na	2 (16.7%)	3 (8.3%)	na	
Moderate immunosuppressant	na	na	4 (17.4%)	11 (30.6%)	na	
Anti-TNFα (infliximab, adalimumab)	na	na	4 (17.4%)	7 (19.4%)	na	
UC classification (N, %) **						na
Ulcerative proctitis (E1)	na	na	4 (17.4%)	na	na	
Distal UC (E2)	na	na	12 (52.2%)	na	na	
Extensive UC or pancolitis (E3)	na	na	4 (17.4%)	na	na	
CD Montreal classification						
Age of diagnosis (N, %) **						0.309‡
diag < 16y (A1)	na	na	1 (4.3%)	3 (8.3%)	nd	
diag 17-40y (A2)	na	na	11 (47.8%)	23 (63.8%)	nd	
diag >41y (A3)	na	na	7 (30.4%)	7 (19.4%)	nd	
Location (N, %)						na
Ileal-CD (L1)	na	na	na	11 (30.5%)	na	
Colonic-CD (L2)	na	na	na	11 (30.5%)	na	
Ileocolonic-CD (L3)	na	na	na	9 (25.0%)	na	
Behavior (N, %) **						na
Non-stricturing, non-penetrating (B1)	na	na	na	20 (55.6%)	na	
Stricturing (B2)	na	na	na	4 (11.1%)	na	
CRC subtype (N, %) **						na
Sporadic	na	na	na	na	11 (50.0%)	
Hereditary***	na	na	na	na	1 (4.5%)	

IBS, Irritable bowel syndrome; IBD, Inflammatory bowel disease; UC, Ulcerative colitis; CD, Crohn's disease; CRC, Colorectal cancer; nd, not determined; na, not applicable

*Controls consisted of subjects who underwent colonoscopy for different reasons: 9/31 rectorrhagia, 11/31 colorectal cancer familial history and 11/31 abdominal pain.

** Medical treatment at the time of sampling was available in 33/36 CD patients, and 20/23 UC patients; Age of disease onset was available for 33/36 CD patients, and 19/23 UC patients; Disease behavior at last follow-up before the time of sampling was available in 24/36 CD patients, and none had penetrating CD (B3); Maximal disease extent at the time of sampling was available in 20/23 UC patients and 31/36 CD patients; presence or absence of relatives with CRC could only be clearly tracked in 12/22 CRC patients.

***Patients were included within this category if a first grade relative has had also CRC.

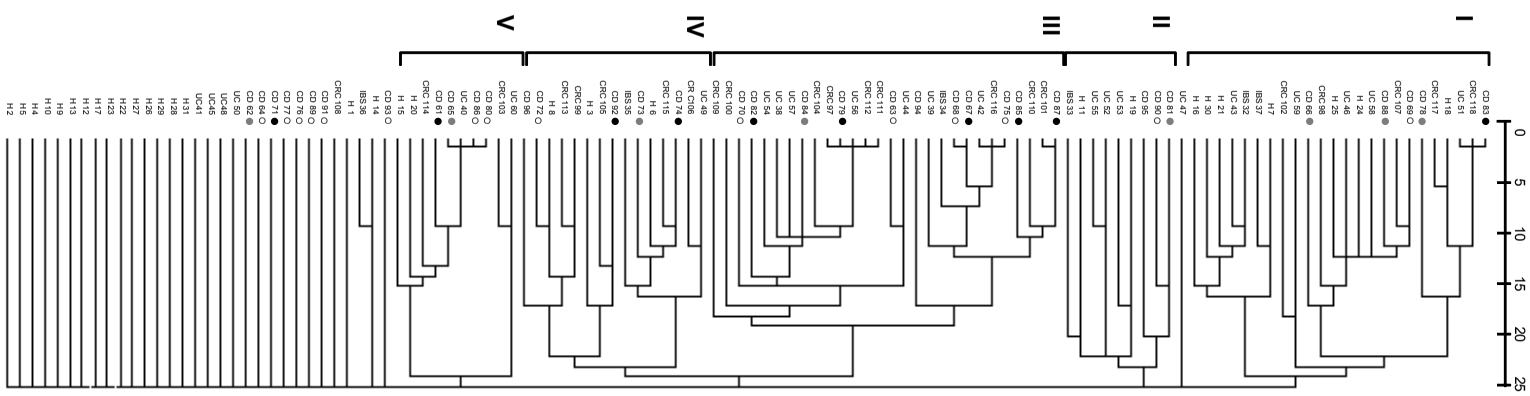
§ Groups were compared by non-parametric statistical tests, and p value ≤0.05 was considered significant

† χ^2 test, ‡ Mann-Whitney U test

788 **Table 2.** Main differences observed at different 16S rRNAgene sequence similarity cutoffs after comparisons of sequences from healthy subjects
789 and patients with several gut disorders.
790

			Nearest *F. prausnitzii* isolate			Number of patients (%)					
Simirlaity cutoff*	**OTU**	**Total sequences**	Strain	Similarity (%)	Accession Number	H (n=31)	IBS (n=6)	UC (n=23)	CD (n=36)	CRC (n=22)	**p-value**
99%	OTU99_1	104	HTF-I	99	HQ457031.1	12 (38.7)	2 (33.3)	17 (73.9)	24 (66.7)	16 (72.7)	**0.020**
	OTU99_10	3	A2-165	99	AJ270469.2	0,0	1 (16.7)	0.0	0.0	2 (9.1)	**0.027**
	OTU99_11	2	S4L/4	99	HQ457025.1	0,0	1 (16.7)	0.0	0.0	0.0	**0.001**
	OTU99_17	1	A2-165	97	AJ270469.2	0,0	1 (16.7)	0.0	0.0	0.0	**0.001**
97%	OTU97_1	161	HTF-I	99	HQ457031.1	16 (51.6)	4 (66.7)	18 (78.3)	31 (86.1)	20 (90.9)	**0.005**
	OTU97_2	118	S3L/3	99	HQ457024.1	27 (87.1)	5 (83.3)	11 (47.8)	16 (44.4)	14 (63.6)	**0.003**
	OTU97_4	1	A2-165	97	AJ270469.2	0	1(16.7)	0	0	0	**0.001**

791 H, healthy subject; IBS, irritable bowel syndrome; UC, ulcerative colitis; CD, Crohn's disease; CRC, colorectal cancer.
792 * Different phylotypes found as calculated by Mothur with the farthest neighbor method using a cutoff of 99% or 97% similarity of the 16S rRNA gene sequence.


I
II
III
IV
V
0
5
10
15
20
25