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 García-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:**

Phyloputype richness of *Faecalibacterium prausnitzii*

*For correspondence:*

Prof. L. Jesús García-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 phylootypes
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≤0.013). 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](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’-TAAAACCTCCTGTTGAGGAAGAT-3’
TTGTCACAAGGCAGTCYKG-3’, flanking a ∼700-bp fragment that includes variable regions V3 to V6.

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

**PCR amplification and DGGE fingerprinting.**

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

PCR reaction was optimized testing different concentrations of MgCl2 (0.25 to 15 mM), deoxyribonucleoside triphosphates (dNTPs; 0.1 to 3 mM), and primers (50 to 900 nM of each primer). The optimized reaction was used in all samples and was composed of: 1× of buffer (II) 10× (Applied Biosystems, Foster City, California), 2.5 mM of MgCl2 (25 mM; Applied Biosystems), 0.2 mM of dNTPs (10 mM; Applied Biosystems), 0.2 µM of GC-Fpra 427F primer, 0.2 µM of Fpra 1127R (10 pmol/mL each), 0.05 U/µL of AmpliTaq DNA polymerase (5 U/mL, Applied Biosystems), and 1 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 (52°C to 70°C). The optimized cycling program was used, and consisted of: 10 minutes at 95°C for initial denaturation and DNA polymerase activation, followed by 35 cycles of 30 seconds at 95°C (denaturation), 30 seconds at 65°C (annealing), and 1 minute at 72°C (extension), with a final extension of 10 minutes at 72°C. 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 60°C 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 Scanner 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 $\chi^2$ 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 $\leq 0.05$. 

10
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
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 through 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
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 a 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 phylotroups 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.
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 rrnDB (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 isolate 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, e.g. 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 horizontal 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 Camprubi-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


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.


FIGURE LEGENDS

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

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

Figure 3. Heatmap showing the relative abundance of sequences assigned to each operational taxonomic unit at a 99% 16S rRNA gene sequence similarity (OTU 99) by group of subjects. Relative abundance 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 99 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 4. Heatmap showing the relative abundance of sequences assigned to each operational taxonomic unit at a 97% 16S rRNA gene sequence similarity (OTU 97) by
group of subjects. Relative abundance 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.
## Table 1. Sample size and clinical characteristics of subjects.

<table>
<thead>
<tr>
<th></th>
<th>Healthy*</th>
<th>Irritable bowel syndrome</th>
<th>Ulcerative colitis</th>
<th>Crohn’s disease</th>
<th>Colorectal cancer</th>
<th>p value&lt;sup&gt;§&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (patients)</td>
<td>31</td>
<td>6</td>
<td>23</td>
<td>36</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Age (mean years ± SD)</td>
<td>49.2±16.3</td>
<td>42.4±11.4</td>
<td>38.4±14.6</td>
<td>34.5±12.8</td>
<td>70.1±10.3</td>
<td>&lt;0.001&lt;sup&gt;‡&lt;/sup&gt;</td>
</tr>
<tr>
<td>Male (N, %)</td>
<td>15 (48.4%)</td>
<td>2 (20.0%)</td>
<td>15 (65.2%)</td>
<td>21 (58.3%)</td>
<td>11 (50.0%)</td>
<td>0.538&lt;sup&gt;†&lt;/sup&gt;</td>
</tr>
<tr>
<td>Active (N, %)</td>
<td>na</td>
<td>na</td>
<td>17 (73.9%)</td>
<td>23 (63.8%)</td>
<td>na</td>
<td>0.365&lt;sup&gt;†&lt;/sup&gt;</td>
</tr>
<tr>
<td>Previous surgery (N, %)</td>
<td>0</td>
<td>na</td>
<td>2 (8.7%)</td>
<td>7 (19.4%)</td>
<td>nd</td>
<td>0.145&lt;sup&gt;†&lt;/sup&gt;</td>
</tr>
<tr>
<td>Smokers (N, %)</td>
<td>0</td>
<td>0</td>
<td>2 (8.7%)</td>
<td>4 (11.1%)</td>
<td>1 (4.5%)</td>
<td>0.386&lt;sup&gt;‡&lt;/sup&gt;</td>
</tr>
<tr>
<td>Treatment (N, %)</td>
<td>0</td>
<td>0</td>
<td>10 (43.5%)</td>
<td>11 (30.6%)</td>
<td>na</td>
<td>0.520&lt;sup&gt;†&lt;/sup&gt;</td>
</tr>
<tr>
<td>No treatment</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>Mesalazine</td>
<td>na</td>
<td>na</td>
<td>2 (16.7%)</td>
<td>3 (8.3%)</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>Moderate immunosuppressant</td>
<td>na</td>
<td>na</td>
<td>4 (17.4%)</td>
<td>11 (30.6%)</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>Anti-TNFα (infliximab, adalimumab)</td>
<td>na</td>
<td>na</td>
<td>4 (17.4%)</td>
<td>7 (19.4%)</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>UC classification (N, %)</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>Ulcerative proctitis (E1)</td>
<td>na</td>
<td>na</td>
<td>4 (17.4%)</td>
<td>na</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>Distant UC (E2)</td>
<td>na</td>
<td>na</td>
<td>12 (52.2%)</td>
<td>na</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>Extensive UC or pancolitis (E3)</td>
<td>na</td>
<td>na</td>
<td>4 (17.4%)</td>
<td>na</td>
<td>na</td>
<td></td>
</tr>
</tbody>
</table>

**CD Montreal classification**

<table>
<thead>
<tr>
<th>Age of diagnosis (N, %) **</th>
<th>0.309&lt;sup&gt;§&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 &lt; 16y (A1)</td>
<td>na</td>
</tr>
<tr>
<td>diag &lt; 16y (A1)</td>
<td>na</td>
</tr>
<tr>
<td>diag 17-40y (A2)</td>
<td>na</td>
</tr>
<tr>
<td>diag 41y (A3)</td>
<td>na</td>
</tr>
<tr>
<td>Location (N, %)</td>
<td>na</td>
</tr>
<tr>
<td>Ileal-CD (L1)</td>
<td>na</td>
</tr>
<tr>
<td>Colonic-CD (L2)</td>
<td>na</td>
</tr>
<tr>
<td>Ileocolonic-CD (L3)</td>
<td>na</td>
</tr>
<tr>
<td>Behavior (N, %) **</td>
<td>na</td>
</tr>
<tr>
<td>Non-stricturing, non-penetrating (B1)</td>
<td>na</td>
</tr>
<tr>
<td>Limiting (B2)</td>
<td>na</td>
</tr>
<tr>
<td>CRC subtype (N, %) **</td>
<td>na</td>
</tr>
<tr>
<td>Sporadic</td>
<td>na</td>
</tr>
<tr>
<td>Hereditary***</td>
<td>na</td>
</tr>
</tbody>
</table>

**CRC** stands for colorectal cancer; IBD, inflammatory bowel disease; UC, ulcerative colitis; CD, Crohn’s disease; nd, not determined; na, not applicable.

*Controls consisted of subjects who underwent colonoscopy for different reasons: 9/31 underwent cancer screening, 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 26/36 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 the category if a first degree relative has had also CRC.

<sup>§</sup> Groups were compared by non-parametric statistical tests, and p value ≤0.05 was considered significant.

<sup>†</sup> χ² test.

<sup>‡</sup> Mann-Whitney U test.
Table 2. Main differences observed at different 16S rRNA gene sequence similarity cutoffs after comparisons of sequences from healthy subjects and patients with several gut disorders.

<table>
<thead>
<tr>
<th>Similarity cutoff*</th>
<th>OTU</th>
<th>Total sequences</th>
<th>Nearest <em>F. prausnitzii</em> isolate</th>
<th>Number of patients (%)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Strain</td>
<td>Similarity (%)</td>
<td>Accession Number</td>
</tr>
<tr>
<td>99%</td>
<td>OTU99_1</td>
<td>104</td>
<td>HTF-I</td>
<td>99</td>
<td>HQ457031.1</td>
</tr>
<tr>
<td></td>
<td>OTU99_10</td>
<td>3</td>
<td>A2-165</td>
<td>99</td>
<td>AJ270469.2</td>
</tr>
<tr>
<td></td>
<td>OTU99_11</td>
<td>2</td>
<td>S4L-4</td>
<td>99</td>
<td>HQ457025.1</td>
</tr>
<tr>
<td></td>
<td>OTU99_17</td>
<td>1</td>
<td>A2-165</td>
<td>97</td>
<td>AJ270469.2</td>
</tr>
<tr>
<td>97%</td>
<td>OTU97_1</td>
<td>161</td>
<td>HTF-I</td>
<td>99</td>
<td>HQ457031.1</td>
</tr>
<tr>
<td></td>
<td>OTU97_2</td>
<td>118</td>
<td>S3L-3</td>
<td>99</td>
<td>HQ457024.1</td>
</tr>
<tr>
<td></td>
<td>OTU97_4</td>
<td>1</td>
<td>A2-165</td>
<td>97</td>
<td>AJ270469.2</td>
</tr>
</tbody>
</table>

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

* 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.
H  IBS  UC  CD  CRC

N=31  N=6  N=23  N=36  N=22

Number of unique sequences

1  2  3  4  5
A

H

IBS

UC

CD

CRC

N=31

N=6

N=23

N=36

N=22

13% 39% 48%

17% 50% 26%

52% 22% 53%

36% 22% 55%

B

C-CD

IC-CD

I-CD

N=11

N=9

N=11

18% 18% 64%

44% 56% 36%

36% 64% 64%

Both phylogroups
Phylogroup I
Phylogroup II