TITLE

Changes in the abundance of *Faecalibacterium prausnitzii* phylogroups I and II in the intestinal mucosa of inflammatory bowel disease and colorectal cancer patients.

RUNNING TITLE

Abundance of mucosa-associated *Faecalibacterium prausnitzii* phylogroups

AUTHORS

MSc. Mireia Lopez-Siles [1], PhD. Margarita Martinez-Medina [1], Romà Surís-Valls [1], PhD. MD. Xavier Aldeguer [2], MD. Miriam Sabat-Mir [3], PhD. Sylvia H. Duncan [4], 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

* **For correspondence and reprints request:**

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

CONFLICTS OF INTEREST AND SOURCES OF FUNDING

Xavier Aldeguer is consultant from AbbVie, and has received honoraria for lectures including services on speakers bureaus from AbbVie and MSD. For the remaining authors none were declared.

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.

ABSTRACT

Background

Faecalibacterium prausnitzii comprises two phylogroups, whose abundance in healthy and diseased gut, and in conjunction with *Escherichia coli* has not yet been studied. This work aims to determine the contribution of *F. prausnitzii* phylogroup I and II in intestinal disease and to assess their potential diagnostic usefulness as biomarkers for gut diseases.

Methods

Total *F. prausnitzii*, its phylogroups and *E. coli* loads were determined by quantitative polymerase chain reaction targeting the 16S rRNA gene on biopsies from 31 healthy controls (H), 45 Crohn's disease (CD), 25 ulcerative colitis (UC), 10 irritable bowel syndrome (IBS), and 20 colorectal cancer (CRC) patients. Data were normalized to total bacterial counts, and analyzed according to patients' disease location and clinical characteristics.

Results

Lower levels of both total *F. prausnitzii* and phylogroup I were found in CD, UC and CRC ($P<0.001$) compared with H subjects. Phylogroup I load was a better biomarker than total *F. prausnitzii* to discriminate subjects with gut disorders from H. Phylogroup II depletion was observed only in CD patients ($P<0.001$), and can be potentially applied to differentiate ulcerative pancolitis from colonic CD. No statistically significant correlation between *E. coli* and any of the two *F. prausnitzii* phylogroups was found in any group of patients or by IBD location. Phylogroup I was lower in active CD patients whereas those CD with intestinal resection showed a

reduction in phylogroup II. Treatments with mesalazine and immunosupressants did not result in the recovery of *F. prausnitzii* phylogroups abundance.

Conclusion

F. prausnitzii phylogroup I was depleted in CD, UC and CRC, whilst phylogroup II was specifically reduced in CD. Quantification of *F. prausnitzii* phylogroups and *E. coli* may help to identify gut disorders, and to classify inflammatory bowel disease location.

KEYWORDS

Faecalibacterium prausnitzii phylogroups, Crohn's disease, ulcerative colitis, colorectal cancer, irritable bowel syndrome.

INTRODUCTION

Faecalibacterium prausnitzii (Ruminococcaceae) is one of the three most abundant species found in the gut, representing between 2–20% of the fecal microbiota in healthy individuals, according to diversity studies of the human gut microbiome based on 16S rRNA gene analysis (1-6). This species has been reported to represent 6% of bacteria in mucosa-associated microbial communities (7), although some studies have indicated that these values can increase to around 20-50% in some individuals (8, 9).

In recent years, there has been increasing interest in *F. prausnitzii* given its potentially important role in promoting gut health (10, 11) through the formation of anti-inflammatory compounds (10-14) and enhancement of intestinal barrier function (15, 16). Many studies have shown that *F. prausnitzii* prevalence and abundance is reduced in different intestinal disorders (for review see (17) and references therein), although the depletion in *F. prausnitzii* numbers has been most extensively reported in inflammatory bowel disease (IBD). Low counts of this species have been observed in both fecal and mucosa-associated communities of adult Crohn's disease (CD) patients (11, 18-21). Variable populations have been reported in ulcerative colitis (UC) patients (7, 18, 19, 22-27), although the reduction of Firmicutes has been repeatedly observed in this disorder (25, 28, 29). A recent study conducted on 127 UC subjects points out that a reduction in *F. prausnitzii* is also involved in UC dysbiosis (25). Interestingly, lower counts of *Faecalibacterium*-related bacteria have also been observed in functional gut disorders such as irritable bowel syndrome (IBS) of alternating type (30), that in turn shares some features with IBD patients (31, 32), and in more severe intestinal disorders as colorectal cancer (CRC)(33). Taken together these findings suggest that shifts in *F. prausnitzii* numbers occur under several pathological disorders but it still remains to

be established if this reduction is equivalent among different conditions, as few studies have considered several gut pathologies simultaneously.

Furthermore, relatively few studies have paid attention to the diversity within the genus *Faecalibacterium*. Recent phylogenetic analysis showed that mainly two different *F. prausnitzii* phylogroups, which include the current cultured representatives, are found in fecal samples of healthy subjects (14) but no data about the abundance of these phylogroups in gut disorders has been reported to date.

Many studies have reported that in addition to *F. prausnitzii* depletion, CD dysbiosis is characterized by an increase in *Escherichia coli* abundance, predominantly in CD patients with ileal involvement (21, 34-36). A possible negative correlation between *F. prausnitzii* and *E. coli* has been observed in I-CD patients (18), suggesting a direct/indirect effect of one population over the other. However it remains to be established whether or not this affects both phylogroups of *F. prausnitzii*.

This work is aimed at determining the variation of mucosa-associated *F. prausnitzii* phylogroups between healthy subjects and patients suffering several intestinal disorders in order to establish whether the imbalance in *F. prausnitzii* includes the overall population or specifically affects a particular phylogroup. Besides correlation between *F. prausnitzii* phylogroups and *E.coli* load has also been analysed. The prevalence and abundance of mucosa-associated *F. prausnitzii* and both phylogroups were determined in samples of CD, UC, IBS and CRC patients and in healthy controls (H) at different locations of the gut. To this end, a novel multiplex quantitative polymerase chain reaction (qPCR) assay was developed for the quantification of the two known phylogroups within this species. Data were analyzed taking into account patients' most relevant clinical characteristics, in order to determine its usefulness to differentially diagnose IBD patients and monitor the evolution of the

disease. Medication at sampling was also considered in order to determine whether any of the current therapies are effective in correcting this species imbalance.

MATERIALS AND METHODS

Patients, clinical data and sampling.

A Spanish cohort consisting of 70 IBD (45 CD and 25 UC), 10 IBS, 20 CRC patients, and 31 H was enrolled (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). Subjects were gender matched for all the groups. Concerning age, CD patients were younger than those in the H group ($P<0.001$), whereas CRC patients were significantly older than all the other groups ($P\leq0.019$). IBD patients were diagnosed according to standard clinical, pathological and endoscopic criteria and categorized according to the Montreal classification (37). IBS patients were diagnosed according to Rome III criteria (available at <http://www.romecriteria.org/criteria/>). The diagnosis of CRC was established by colonoscopy and biopsy, and data correlated with high risk of developing this disease was recorded. The control group consisted of subjects undergoing colonoscopy for different reasons as rectorrhagia (N=9), colorectal cancer familial history (N=11), and abdominal pain (N=11). Clinically relevant data of all the patients was collected. 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 endoscopy, up to three biopsy samples per patient were taken from different locations along the gut (distal ileum, colon, and rectum) following standard procedures. 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.

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 (34). DNA was extracted using the NucleoSpin® Tissue Kit (Macherey-Nagel GmbH &Co., Duren, Germany). The support protocol for Gram positive bacteria and the RNAse treatment step were carried out. Genomic DNA was eluted with 10mM Tris-HCl (pH 7.4) and 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 and hydrolysis probes design, and set up of a qPCR assay for *F. prausnitzii* phylogroups.

In order to simultaneously quantify both *F. prausnitzii* phylogroups, a qPCR assay consisting of a unique pair of species-specific primers for *F. prausnitzii* and two hydrolysis probes targeting each *F. prausnitzii* phylogroup was designed.

Sequences of the 16S rRNA gene from *F. prausnitzii* and from closely related Ruminococcaceae were recovered from GenBank (see Table S1, Supplemental Digital Content 1) and aligned using Clustal W (38). Both primers and hydrolysis probes were manually designed, from consensus sequences (see Table S1, Supplemental Digital Content 1) specifically built for each purpose, following the guidelines set by Applied Biosystems (Foster City, CA, USA) for the design of primers and probes for allelic discrimination, and further checked using the software Primer Express® version 3.0 (Applied Biosystems, Foster City, CA, USA). Oligonucleotides were also evaluated using NetPrimer® software (PREMIER Biosoft International, California, USA) to check for primer-dimer structures, hairpins and possible cross dimer interactions. Resulting primers and probes are listed in Table 2.

Optimization of the reagents for the *F. prausnitzii* phylogroup qPCR assay was performed as described in Supplementary text (Supplemental Digital Content 1). Oligonucleotides specificity was checked against the Ribosomal Database Project II (RDP) (39) and GenBank database through Seqmatch and BLAST (40), respectively. Coverages were evaluated using the SILVA Probe Match and Evaluation Tool - TestProbe 3.0 (available at http://www.arb-silva.de/search/testprobe/). Finally, *in vitro* inclusivity/exclusivity test was performed including 89 bacterial strains, nine of which were *F. prausnitzii* (see Table S2, Supplemental Digital Content 1). Linearity, efficiency and detection limit of the assay were determined as detailed in Supplementary text (Supplemental Digital Content 1).

Quantification standards for qPCR.

Standard DNA templates from *F. prausnitzii* strain S3L/3 (phylogroup I), *F. prausnitzii* DSM 17677 (phylogroup II) and *E. coli* CECT 105 were prepared as genetic constructs after PCR amplification as previously reported (41, 42), and subsequent insertion of the whole 16S rRNA gene into a pCR®4-TOPO® cloning plasmid (Invitrogen, CA, USA) following manufacturer's guidelines. After purification with the NucleoSpin® Plasmid (Macherey-Nagel GmbH&Co., Duren, Germany), plasmids were linearized with *SpeI* (*F. prausnitzii*) or *PstI* (*E. coli*) and quantified using Qubit™ Quantitation Platform (Invitrogen, Carlsbad, USA). Initial target concentration was inferred as previously reported (18). Standard curves were obtained from ten-fold serial dilutions of the titrated suspension of linearized plasmids, and ranged from 20 to 2×10^8 copies/reaction, which correspond to the linear dynamic range span for all the reactions (see Supplementary text, Supplemental Digital Content 1). The standard curve built with *F. prausn*itzii DSM 17677 16S rRNA gene was used for both the total bacteria and the total faecalibacteria 16S rRNA gene quantification, and standard curves

obtained from either phylogroup were intercalibrated using the total *F. prausnitzii* primers and probe set. Total bacteria 16S rRNA gene quantification and the *F. prausnitzii* standard curve were used to check the *E. coli* standard curve quantification in order to make sure that results obtained with both standard curves were comparable. Finally, to demonstrate that the new assay correctly quantifies the appropriate ratios of phylogroups I and II when DNA from both are present in the same sample (as would be expected in vivo), mixtures of both DNA templates were prepared (*i.e.* phylogroup I: phylogroup II mixed at 100:0, 25:75, 50:50, 75:25 0:100), and quantified as unknown samples. Less than 10% of variation was obtained between the experimental qPCR quantification results with that of the expected quantity (see Table S3, Supplemental Digital Content 1).

qPCR assays.

Previously reported 16S rDNA-targeting primers and probe were used for total *F. prausnitzii* (18), *E. coli* (43) and total bacteria (44) quantifications, and amplification reactions were carried out as described elsewhere (18, 44, 45). The novel assay for *F. prausnitzii* phylogroups quantification was carried out in a total volume of 20 µl reactions containing: 1× TaqMan® Universal PCR Master Mix 2× (Applied Biosystems, Foster City, CA, USA), 900 nM of each primer, 300 nM of each probe, and up to 50 ng of genomic DNA template. All primers and probes used in this study as well as PCR conditions are detailed in Table 2. Total *F. prausnitzii*, *E. coli* and total bacteria primers and hydrolysis probes were purchased from Applied Biosystems (Foster City, CA, USA), whereas primers and hydrolysis probes for *F. prausnitzii* phylogroups were acquired from Biomers (Ulm, Germany). The DNA of the internal amplification control (IAC) was synthesized by Bonsai technologies group (Alcobendas, Spain).

Samples were run in duplicate in the same plate. For data analysis, the mean of the duplicate quantifications was used. Duplicates were considered valid if the standard deviation between quantification cycles (C_q) was <0.34 (*i.e.* a difference of <10% of the quantity was tolerated). Quantification controls consisting of at least five reactions with a known number of target genes were performed to assess inter-run reproducibility. Inhibition was controlled on total *F. prausnitzii* quantification by adding 10^3 copies of IAC template to each reaction. It was considered that there was no inhibition if the obtained C_q was <0.34 different from those obtained when quantifying the IAC alone for any of the replicates. A no-template control consisting of a reaction without *F. prausnitzii* DNA as well as a non-amplification control which did not contain any DNA template (either bacterial or IAC) were also included in each run. Negative controls resulted in undetectable C_q values in all cases.

All quantitative PCR were performed using a 7500 Real Time PCR system (Applied Biosystems, Foster City, CA, USA). Data were collected and analyzed using the 7500 SDS system software version 1.4 (Applied Biosystems, Foster City, CA, USA). All quantifications were done under average PCR efficiencies of 89.51±7.06%.

Data normalization and statistical analysis.

As regards to qualitative analyses, absence of *F. prausnitzii* or its phylogroups was considered if no detection was obtained during the qPCR analysis, corresponding to samples that carried *F. prausnitzii* or the phylogroups below the detection limit (*i. e.* 106.6, 1.10 and 2.39 16S rRNA genes per reaction for total *F. prausnitzii*, phylogroup I and phylogroup II, respectively). Pearson's χ^2 test was used to compare the prevalence of *F. prausnitzii* and its phylogroups between groups of patients and by IBD disease location.

Referring to quantitative analyses, total *F. prausnitzii*, phylogroups and *E. coli* copy numbers were normalized to the total bacteria 16S rRNA gene copies. Data is given as the $\log_{10}$ of the ratio between 16S rRNA gene copies of the target microorganism and million of total bacterial 16S rRNA genes detected in the same sample.

The non-parametric Kruskal-Wallis test was used to test differences in variables with more than two categories such as diagnostics, CD and UC disease location, and current medication. Pairwise comparisons of subcategories of these variables were analyzed using a Mann-Whitney U test. This test was also used to compare, within a subgroup of patients, variables with two categories such as activity (active CD and UC patients when CDAI>150 (46) and a Mayo score >3 (47), respectively), and intestinal resection. In addition, the receiver operating characteristic (ROC) curve analysis, a plot of the true positive rate (sensitivity) versus false positive rate (1-specificity), was applied to establish the usefulness of *F. prausnitzii*, and each phylogroup to distinguish amongst different intestinal disorders. The accuracy of discrimination was measured by the area under the ROC curve (AUC). An AUC approaching 1 indicates that the test is highly sensitive as well as highly specific whereas an AUC approaching 0.5 indicates that the test is neither sensitive nor specific.

Spearman correlation coefficient and significance between the phylogroups quantities, and between phylogroups quantities and *E. coli* was calculated. The same statistical method was used to analyze the correlation between each one of the phylogroups with respect to total faecalibacteria quantity, and clinical data such as time (in years) since disease onset.

All the statistical analyses were performed using the SPSS 15.0 statistical package (LEAD Technologies, Inc.). Significance levels were established for P values ≤ 0.05.

ETHICAL CONSIDERATIONS.

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.

RESULTS

Features of the novel multiplex qPCR assay for *F. prausnitzii* phylogroups I and II.

In this study, a novel oligonucleotide set was designed to quantify the two recently described *F. prausnitzii* phylogroups (Table 2, and see Supplemental Digital Content 1). The *in silico* analysis of the oligonucleotide set of choice showed that primer Fpra 136F-Fpra 232R were specific for *F. prausnitzii* and targeted all the isolates available to date, whereas the probes PHG1 180PR and PHG2 180PR specifically matched phylogroups I and II, respectively. These results were confirmed *in vitro* by the inclusivity-exclusivity tests (see Table S2, Supplemental Digital Content 1). Coverage of the Fpra 136F-Fpra 232R primers set was 74.85% of the sequences in the SILVA datasets. PHG1 180PR probe targeted 20.50% of the *Faecalibacterium* sp. sequences whereas PHG2 180PR probe coverage was 38.80% of the *Faecalibacterium* sp. sequences in this database. For both reactions reliable quantification was possible over a linear range span of 7 logarithms, starting at 20 target genes per reaction (R^2=0.998), with an average efficiency of 85.68±3.23 % for phylogroup I and 90.31±3.40% for the phylogroup II. The detection limits were 1.10 and 2.39 target genes for phylogroup I and phylogroup II, respectively.

Prevalence of mucosa-associated *F. prausnitzii* and phylogroups I and II along the gut in health and disease.

Prevalence of *F. prausnitzii* and both phylogroups as calculated from positive determinations over total samples was analyzed both by disease status considering all the data across all sites (Figure 1), and by sample location (Figure 2). *F. prausnitzii* prevalence was lower in CD patients than in H (Figure 1). CD patients with I-CD feature lower *F. prausnitzii* prevalence than those with E1, E2, E3 and C-CD. Prevalence values ranged from 81-100%, except for I-CD whose value was significantly

lower (down to 68%, P≤0.046) regardless of the location (Figure 2). In contrast, reduced prevalence was only evident in ileal and colonic samples in ileocolonic-CD (IC-CD) (75% and 80% respectively) and in rectal samples in colonic-CD (C-CD) (75%) although the differences were not statistically supported (Figure 2).

As far as the phylogroups are concerned, both were found to be less prevalent in CD patients (P<0.001) than in the H and CRC groups, particularly in those with ileal involvement (Figure 1). Of particular interest is the absence of phylogroup I from all five ileal samples of the I-CD patients analyzed (Figure 2). Phylogroup II was less prevalent in I-CD patients regardless of sample location. The same happened in colon and ileum of IC-CD patients, as well as in rectal samples of C-CD patients. For CRC and UC patients, the prevalence remained similar to H. Nevertheless phylogroup I showed a trend of lower values in ulcerative pancolitis, which did not reach statistical significance (P=0.053) probably due to the low number of samples processed. Similarly IBS patients only had reduced prevalence of phylogroup I in comparison to H subjects.

Both phylogroups co-occurred in 85.4% and 85.0% of samples containing *F. prausnitzii* from H and CRC patients, respectively. Phylogroup I was exclusive in 10% of H and CRC subjects, whereas phylogroup II was found as the only representative in 4.2% of H subjects (Figure 3A). In contrast, 16% of IBS, 6% of UC and 22% of CD patients with *F. prausnitzii* carried neither phylogroup I nor II, which suggests the existence of other phylogroups. Differences in prevalences were observed between IBD disease location. All the patients with less severe UC (*i.e.* E1 and E2) had one or both *F. prausnitzii* phylogroups, resembling H subjects, whereas none of the phylogroups were detected in 23.1% of ulcerative pancolitis patients despite having *F. prausnitzii* (Figure 3B). Similarly, 22.2% of all CD patients did not show either of the phylogroups. Within CD patients, 47.1% of C-CD patients had both *F. prausnitzii*

phylogroups whereas the presence of a unique phylogroup was more frequent (44.4% of IC-CD and 28.0% of I-CD patients) in those with ileal involvement. Remarkably whenever a single phylogroup was found in I-CD it always was the phylogroup II.

Abundances of mucosa-associated *F. prausnitzii* and phylogroups in health and disease.

The abundance of *F. prausnitzii* and its phylogroups from all the biopsies pooled together was compared amongst patients with different intestinal disorders and H subjects (Table 3). *F. prausnitzii* was less abundant in IBD and CRC patients as compared to healthy subjects ($P<0.001$), whereas IBS patients closely resembled the H group. As previously reported (18), within UC patients, those with E1 and E3 presented *F. prausnitzii* loads similar to H subjects, whereas those with E2 had abundances between CD patients and H subjects. In CD patients, those with ileal involvement presented the lowest levels of this bacterium, whereas C-CD patients were similar to UC (Table 3).

F. prausnitzii phylogroup I load was reduced in all the intestinal diseases analyzed in comparison to H subjects, except for IBS patients, probably due to the low number of patients included and the high dispersion of data. This reduction was particularly conspicuous in CD patients, who had values 1000 times lower than H subjects ($P<0.001$). When analyzing data by disease location, all CD patients showed this marked reduction of phylogroup I abundance, as well as those UC patients with E3 that resembled more to CD patients than to those with other UC disease location. In contrast, *F. prausnitzii* phylogroup II abundance was only significantly reduced in CD patients in comparison to H ($P<0.001$) (Table 3), particularly in those with ileal involvement (either I-CD or IC-CD), suggesting that in these patients the depletion of *F. prausnitzii* affects the overall faecalibacteria community.

Interestingly, in H, CRC and IBS subjects the abundance of the two phylogroups was similar, whereas in IBD patients phylogroup II outnumbered phylogroup I (Table 3). In UC patients, *F. prausnitzii* phylogroup II abundance was twice that of phylogroup I, whereas in CD patients the imbalance between the two phylogroups was more marked, with *F. prausnitzii* phylogroup II the abundance was 6.76 times higher than that of phylogroup I. Notably, patients with E3 also featured a marked imbalance in phylogroup abundances which resembled that found in CD.

Usefulness of mucosa-associated *F. prausnitzii* and phylogroup abundance as diagnostic biomarkers.

ROC curve analysis, applied to test the putative accuracy of total *F. prausnitzii* abundance as an indicator to differentiate between two groups of patients, confirmed that the reduction of this species load is a good discriminator for CRC patients from H and IBS patients, with AUC values greater than 0.8 (Figure 4) with an 80% of specificity and above 70% of sensitivity at a set threshold. Good discrimination was also observed between CD and H patients, although for the same specificity values, sensitivity was reduced to 62%. Interestingly, phylogroup I abundance was a more accurate indicator to distinguish H from IBD subjects, than total *F. prausnitzii* abundance (Figure 4). When comparing H subjects with UC more than 76.60% of sensitivity and above 57.14% of specificity at a set threshold were reached for all the disease locations but with the exception of ulcerative proctitis (E1). Specificity was improved up to 70% when considering exclusively E3 patients. In addition, phylogroup I abundance was a particularly accurate biomarker to distinguish H and CD patients (91.48% sensitivity, 73.02% specificity), especially those with I-CD in which 91.48% sensitivity and up to 88.00% of specificity could be reached. Although phylogroup II abundance can accurately discriminate H and CD subjects, AUC values were slightly

lower than those obtained for phylogroup I, thus indicating that the latter is a more suitable biomarker for H status. In contrast, phylogroup II was a useful biomarker to discriminate within IBD subtypes as the best AUC values were obtained to distinguish between ulcerative pancolitis patients and those with CD with colonic involvement (phylogroup II AUC E3vsC-CD=0.817).

Correlation of total *F. prausnitzii* with phylogroups, between phylogroups, and between phylogroups and *E. coli* abundances.

Correlations between total *F. prausnitzii* and phylogroups abundances were conducted in order to determine if the depletion in *F. prausnitzii* abundance could be attributed to the depletion of one of the phylogroups in certain intestinal disorders. In H and IBD patients a positive correlation exists between the two phylogroups and total *F. prausnitzii* abundance, suggesting that they are key contributors to *F. prausnitzii* abundance in the gut of these groups of patients (Table 4). In contrast, in CRC patients a significant correlation was found only for phylogroup I and total faecalibacteria abundance, which suggests that phylogroup II subpopulation is not particularly influencing total *F. prausnitzii* load in this clinical condition. Similarly, no significant correlation was found in IBS, probably because of the low cohort of these patients included in this study.

Abundances of the two phylogroups were positively correlated in H and IBD patients. In contrast, no significant correlation between the two phylogroup loads was found in IBS and CRC patients (Table 4), suggesting that in these disorders the gut environmental conditions differentially impact on each phylogroup.

Finally, correlation between *F. prausnitzii* phylogroups and *E. coli* was determined in order to establish if they were positively or negatively correlated, and whether this could provide supporting evidence about a putative common factor

affecting negatively/positively both bacterial populations in a given patient or about a direct/indirect effect of one population over the other. No statistically significant correlation between *E. coli* and any of the two *F. prausnitzii* phylogroups was found in any group of patients or by IBD location however, it is intriguing that phylogroup II load negatively correlated with *E.coli* in all the groups of gut disease (Table S4). A significant negative correlation between phylogroup II and *E. coli* across all disease groups together was observed (ρ=-0.196, P=0.016).

F. prausnitzii and phylogroup abundances in relation to patients clinical and treatment data.

F. prausnitzii and the abundance of the phylogroups did not differ between active and inactive UC patients (see Table S5, Supplemental Digital Content 1). Although no statistical significance was reached, active CD patients showed a marked reduction on phylogroup I abundance with respect to CD patients in remission (P=0.106).

F. prausnitzii abundance was reduced in those CD patients that underwent intestinal resection (see Table S6, Supplemental Digital Content 1). Interestingly, this could be attributable to lower phylogroup II abundance, that was 10-fold lower in resected CD patients than in those without intestinal surgery (P=0.001) whereas the phylogroup I load was only slightly lower between resected and non-resected patients.

Concerning disease duration, no statistically significant correlation was found between time from disease onset and *F. prausnitzii* and phylogroup abundances (data not shown).

Finally, as far as therapies are concerned, data were analyzed taking into account the medication of the patients at the time of sampling (see Table S7, Supplemental

Digital Content 1). No differences in *F. prausnitzii* or in phylogroup abundances were observed between medications within any IBD. However, those CD patients who received no treatment or mesalazine had higher *F. prausnitzii* loads than those patients under moderate immunosupressants or anti-tumor necrosis factor. No medication was associated with the recovery of normal levels of these bacterial indicators.

DISCUSSION

In the present study we have analyzed the prevalence and abundance of mucosa associated *F. prausnitzii* and its two phylogroups in H, IBS, CRC and IBD subjects, taking into account both the diversity of disease locations and the clinical features of patients. *F. prausnitzii* abundance is reduced in several intestinal disorders, and for the first time we describe how the abundance of its phylogroups differ between intestinal conditions, and in relation to *E.coli*. New data on phylogroup distribution along the gut and in relation to clinical data are revealed.

Our data show that mucosa-associated *F. prausnitzii* loads are markedly reduced in CRC and CD patients, especially in those with ileal involvement. *F. prausnitzii* was below detection limits of the method (106.6 16S rRNA genes of *F. prausnitzii* per reaction) in 5% of CRC and 20% of CD patients. UC patients also featured a lower *F. prausnitzii* abundance than H subjects, but this depletion was four-times less prominent than the depletion observed in CD and CRC patients. Finally, abundance in IBS patients was similar to H subjects. Our study is in agreement with previous reports which found *F. prausnitzii* to be less abundant and/or prevalent in adult CD (11, 17-21, 28, 34), UC (7, 19, 24-27) and CRC (33). Intriguingly, a recent study has reported increased *F. prausnitzii* abundance in *de-novo* pediatric CD patients (22) which is not in line with our results and suggests that dysbiosis in adult and pediatric CD may be different, which merits further investigation. Contradictory data can also be found in the literature concerning lower *F. prausnitzii* numbers in CRC (33, 48, 49). Controversy also exists with respect to IBS patients. Some previous studies report normal counts (7, 20, 23, 50-53), whereas others found lower numbers, but exclusively in those patients with IBS of alternating type (30). We have not observed depletion in *F. prausnitzii* load

in IBS patients, although this observation could be biased by the small cohort size which also had not been classified by disease type.

Among many intestinal disorders (IBS, diarrhea, upper gut disorder, colonic disorder, UC, CD, ischemic colitis, celiac disease and self-limiting colitis) CD patients have been shown to possess the lowest abundance of *F. prausnitzii* in feces (20, 23). These results are now corroborated in intestinal mucosa by our study, which reveals for the first time that at mucosa level, the abundance of *F. prausnitzii* in CRC is similar to that found in patients with CD. Altogether, these findings suggest that down-shifts in *F. prausnitzii* numbers occur under several pathological disorders although the numbers are especially compromised in severe diseases such as CRC and CD. Our study supports the view of *F. prausnitzii* as an indicator of healthy gut status. It has been reported that *F. prausnitzii* is seriously affected by the changes that occur in gut environmental conditions during disease such as changes in pH, bile salt or oxygen (13, 14). This suggests that its decrease may be regarded as an indicator of an altered gut environment, which can be associated with worse disease prognostics, and that changes in the abundance of this species is not necessarily indicating a pathogenic role but rather that yet some environmental factors of the gut compromising its presence remain altered. Besides, recent studies (54) have suggested that the beneficial effect of enteral nutrition in pediatric CD are not mediated by an increase in this species. The fact that mucosal healing can be achieved in CD with enteral nutrition whilst *F. prausnitzii* decreases suggests that the effect of this species may be relatively modest compared with some other factor(s) that are improved by enteral nutrition.

We have further analyzed the prevalence and abundance of *F. prausnitzii* phylogroups I and II, by developing a new quantitative assay. Approximately 25% of all *Faecalibacterium* sequences available in SILVA dataset are not targeted *in silico* by any

of these assays. This discrepancy could be due to the existence of other phylogroups and/or because different phylogroup probes do not include all members within each phylogroup. Our results are still valid however to compare between diseases in our study, as all have been analyzed with the same technique. The majority of H and CRC subjects harbored both phylogroups far higher than the detectable level whereas IBS, and IBD patients feature a reduced prevalence of one of the phylogroups, particularly those with CD. Furthermore, phylogroup I and II were undetected in 16% of IBS and 22% of CD patients with *F. praunsitzii*. These results suggest an imbalance within the *F. prausnitzii* population in these diseases and suggest the existence of at least one more phylogroup.

Quantitative analysis demonstrated that, while the depletion in phylogroup I abundance is a general feature in abnormal gut conditions, the depletion of *F. prausnitzii* phylogroup II seems to be specific to CD patients with ileal disease location. At this stage we cannot determine whether or not this is involved in the pathogenesis of this disease location, or if it is a consequence. It does however indicate that the overall *Faecalibacteria* community is depleted in CD patients and supports the hypothesis that patients with ileal disease location constitute a differentiated pathological entity (21). Previous work based on inferring *F. prausnitzii* subgroup quantities from PCR band intensity on agarose gels already suggested that the levels of M21/2 subgroup (phylogroup I) in CD patients were lower than those in the control group, and that the levels of the A2-165 subgroup (phylogroup II) were the lowest for CD patients (23). These observations have now been quantitatively confirmed on mucosal samples by our current study, which in addition reveals differences between IBD subtypes. Currently there is no phenotypic trait that consistently distinguishes *F. prausnitzii* members from one or other phylogroup (14) which can undoubtedly

explain their differential load in specific disease phenotypes, although the effect of host factors differentially influencing *F. praunitzii* subpopulations has not yet been explored. Another hypothesis could be that *F. prausnitzii* phylogroups interact in a different manner with other members of the microbiome. We have observed that in all patients with gut disease phylogroup II tends to negatively correlate with *E. coli* whereas correlation between this species and phylogroup I depends on the patient group. Our data does not allow us to decipher whether or not one population is directly influencing the other, but suggests that interaction between these two species varies between different gut conditions.

The potential use of *F. prausnitzii* and its phylogroup quantification to assist in IBD diagnostics or to monitor disease progression is of interest in clinical management. It has been reported that CD and UC could be differentiated through monitoring *F. prausnitzii* abundance in conjunction with fecal leucocyte counts (20). Furthermore, the usefulness of *F. prausnitzii* abundance in biopsy samples as a biomarker to distinguish IBD patients from IBS and H subjects has been demonstrated recently (18). Adding *Escherichia coli* counts as a complementary contrasting indicator improved the discrimination power and allowed for good differentiation of IBD locations difficult to discriminate such as I-CD from IC-CD and C-CD from extensive UC. *F. prausnitzii* phylogroups I and II could be novel biomarkers to improve differential diagnosis of those IBD subtypes which are usually difficult to distinguish. For instance, we have observed that phylogroup II is reduced in IC-CD with respect to C-CD, whereas phylogroup I is less abundant in extensive UC than in distal UC. Moreover, phylogroup I proved to be a more accurate marker than total *F. prausnitzii* counts to discriminate between H subjects and those with IBD. However, prospective studies to support the applicability of *F. prausnitzii* phylogroup abundance as biomarkers by comparing with,

for example, established measures such as CRP, albumin, and fecal calprotectin would be necessary to truly determine their ability to distinguish between intestinal disorders and IBD subtypes. In addition, further validation of our results in feces would provide a non-invasive approach to identify CD and UC, which is more likely to be used as diagnostics test.

The fact that *F. prausnitzii* abundance, including both phylogroups, seems to remain lower under remission suggest that this depletion may be occurring at early disease stages or even prior to disease onset, and remains altered over time even if there is endoscopic and clinical remission. Previous studies based on biopsies from CD patients with both active and in remission carry lower *F. prausnitzii* numbers in comparison to H subjects (18, 21). Our data confirm that this feature is shared by both phylogroups. However, despite no statistically significant differences being observed, active CD patients presented a reduction of phylogroup I levels in comparison with inactive patients. Therefore, subsequent studies on larger cohorts of patients are needed to corroborate this trend, and follow up studies would also be interesting to determine how disease status may be specifically compromising this subpopulation and to irrefutably rule out its potential usefulness as a prognostic biomarker. In agreement with previous studies (11, 18) lower numbers of *F. prausnitzii* were detected in resected CD patients. This reduction is also replicated with phylogroups counts. In this case nevertheless, statistical significant differences were only achieved for phylogroup II, probably because the depletion is more striking. However, whether this shift is a consequence of the surgery is still unclear.

In general terms, we have observed that current medication does not restore the levels of mucosa-associated *F. prausnitzii* or its phylogroups, which is in agreement with a previous report (18) although little attention has been paid in the literature to the

effect of medication on *F. prausnitzii* abundance. Some specific therapies not included in the present study such as chemotherapy with somatostatin and interferon α-2b treatment in patients with mid-gut neuroendocrine tumor (55), and rifaximin (56), high-dose cortisol and infliximab (20) in CD patients have proven useful to restore the level of this species. Altogether these data suggest that such therapies will be more useful in terms of counterbalancing *F. prausnitzii* depletion, follow-up studies monitoring this species load in patients starting these treatments will be necessary to demonstrate their effect on modulating this species and its phylogroups abundance.

Finally, in IBD patients, *F. prausnitzii* abundance correlated positively with both phylogroup I and II. A positive correlation was also found between phylogroups. This indicates that environmental changes in the gut ecosystem of IBD patients have a similar effect on both phylogroups and that a reduction in both phylogroups is an indication of the total *F. prausnitzii* population decrease. In line with this observation, all *F. prausnitzii* representatives cultured so far, regardless of their phylogroup, are sensitive to small physico-chemical changes in the gut occurring as a consequence of disease status, such as lower pH or increased bile salts content (14). However, the depletion of phylogroup II was specifically observed in CD patients with ileal involvement. This suggests that specific phenomena in particular gut diseases can compromise one group more than the other. Therefore the use of *Faecalibacterium* as a fine indicator of different gut environmental alterations, which would be characteristic of each intestinal disease, could be the subject for further research. In addition, assessing whether or not *F. prausnitzii* populations hosted by patients with different intestinal disorders are different from those found in H subjects at the level of subtype composition may shed light on the role of this species in gut health maintenance.

CONCLUSIONS

Mucosa-associated *F. prausnitzii* is significantly depleted in patients with gut disorders. Populations of phylogroups I and II of this species however depend on the disease condition. Thus, while *F. prausnitzii* phylogroup I is generally depleted in most intestinal diseases, phylogoup II numbers are specifically reduced in CD. Phylogroup loads can be potentially applied to assist in gut disease diagnostics and in IBD location classification.

ACKNOWLEDGEMENTS

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. Thanks are due to PhD. Teresa Mas-de-Xaxars for assistance in CRC samples recruitment and Ms. Natàlia Adell from the Serveis Tècnics de Recerca of the Universitat de Girona for statistical assistance.

FIGURE LEGENDS

Figure 1. Prevalence of *F. prausnitzii* (black), phylogroup I (grey) and phylogroup II (white) by disease (left) and inflammatory bowel disease location (right) considering all the biopsies from different gut locations. H, control subjects; CRC, colorectal cancer; IBS, irritable bowel syndrome; UC, ulcerative colitis; CD, Crohn's disease; E1, ulcerative proctitis; E2, ulcerative left-sided colitis; E3, ulcerative pancolitis; C-CD, colonic CD; IC-CD, ileocolonic CD; I-CD, ileal CD, IBD, inflammatory bowel disease. Numbers in the bars indicate the number of patients (*biopsies*) analysed to calculate the prevalence. Statistics was calculated separately for each panel. Homogeneous subgroups ($P>0.05$) within each panel are indicated with the same symbols above the bars, whereas

groups of patients with statistically different prevalences (P<0.05) do not share any superscript.

Figure 2. *F. prausnitzii* (black), phylogroup I (grey) and phylogroup II (white) prevalence at ileum, colon and rectum by disease (left panels) and inflammatory bowel disease location (right panels). H, control subjects; CRC, colorectal cancer; IBS, irritable bowel syndrome; UC, ulcerative colitis; CD, Crohn's disease; E1, proctitis; E2, left-sided colitis; E3, pancolitis; C-CD, colonic CD; IC-CD, ileocolonic CD; I-CD, ileal CD, IBD, inflammatory bowel disease; nd, not determined. Numbers in the bars indicate the number of biopsies analysed to calculate the prevalence. Homogeneous subgroups (P>0.05) within each panel are indicated with the same symbols above the bars, whereas groups of patients with statistically different prevalences (P<0.05) do not share any superscript.

Figure 3. Prevalence of *F. prausnitzii*, phylogroup I and phylogroup II in each group of patients (A) and by disease subtype in IBD patients (B). Prevalences along the gut (from inner to outer circles-ileum, colon and rectum) and pooling all the samples (outer circle) have been represented. H, control subjects; CRC, colorectal cancer; IBS, irritable bowel syndrome; UC, ulcerative colitis; CD, Crohn's disease; E1, ulcerative proctitis; E2, ulcerative left-sided colitis; E3, ulcerative pancolitis; C-CD, colonic CD; IC-CD, ileocolonic CD; I-CD, ileal CD. Numbers in the sectors indicate the number of biopsies analysed. * Samples with uncertain location have been included in the average analysis of IBS patients.

Figure 4. Suitability of *F. prausnitzii* and phylogroups abundances as biomarkers to distinguish amongst different intestinal disorders and inflammatory bowel disease

635 locations determined by the area under the curve (AUC) obtained by receiver operating
636 characteristic analysis (ROC curve). A test is considered to be suitable if the AUC range
637 from 0.6 to 0.75, and to have good sensitivity and specificity if the AUC range from
638 0.75 to 0.9. H, controls; IBD, inflammatory bowel disease; IBS, irritable bowel
639 syndrome; UC, ulcerative colitis; CD, Crohn's disease; CRC, colorectal cancer; I-CD,
640 ileal CD; IC-CD, ileocolonic CD, C-CD, colonic CD; E1, ulcerative proctitis, E2, distal
641 UC; E3, extensive UC or ulcerative pancolitis.
642
643 **REFERENCES**
644 1. Arumugam M, Raes J, Pelletier E, et al. Enterotypes of the human gut
645 microbiome. Nature. 2011;473:174-180
646 2. Eckburg PB, Bik EM, Bernstein CN, et al. Diversity of the human intestinal
647 microbial flora. Science. 2005;308:1635-1638
648 3. Hold GL, Schwiertz A, Aminov RI, et al. Oligonucleotide probes that detect
649 quantitatively significant groups of butyrate-producing bacteria in human feces. Appl
650 Environ Microbiol. 2003;69:4320-4324
651 4. Schwiertz A, Jacobi M, Frick JS, et al. Microbiota in pediatric inflammatory
652 bowel disease. J Pediatr. 2010;157:240-244
653 5. Suau A, Rochet V, Sghir A, et al. *Fusobacterium prausnitzii* and related species
654 represent a dominant group within the human fecal flora. Systematic and Applied
655 Microbiology. 2001;24:139-145
656 6. Walker AW, Ince J, Duncan SH, et al. Dominant and diet-responsive groups of
657 bacteria within the human colonic microbiota. ISME J. 2011:220-230

7. Swidsinski A, Loening-Baucke V, Lochs H, et al. Spatial organization of bacterial flora in normal and inflamed intestine: a fluorescence in situ hybridization study in mice. World J Gastroenterol. 2005;11:1131-1140

8. Nava GM, Stappenbeck TS. Diversity of the autochthonous colonic microbiota. Gut Microbes. 2011;2:99-104

9. Baumgart M, Dogan B, Rishniw M, et al. Culture independent analysis of ileal mucosa reveals a selective increase in invasive *Escherichia coli* of novel phylogeny relative to depletion of Clostridiales in Crohn's disease involving the ileum. ISME J. 2007;1:403-418

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

11. Sokol H, Pigneur B, Watterlot L, et al. *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. Proc Natl Acad Sci U S A. 2008;105:16731-16736

12. Barcenilla A, Pryde SE, Martin JC, et al. Phylogenetic relationships of butyrate-producing bacteria from the human gut. Appl Environ Microbiol. 2000;66:1654-1661

13. Duncan SH, Hold GL, Harmsen HJ, et al. 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 2002;52:2141-2146

14. Lopez-Siles M, Khan TM, Duncan SH, et al. 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. 2012;78:420-428

15. Carlsson AH, Yakymenko O, Olivier I, et al. *Faecalibacterium prausnitzii* supernatant improves intestinal barrier function in mice DSS colitis. Scand J Gastroenterol. 2013;48:1136-1144

16. Wrzosek L, Miquel S, Noordine ML, et al. *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. 2013;11:61

17. Miquel S, Martin R, Rossi O, et al. *Faecalibacterium prausnitzii* and human intestinal health. Curr Opin Microbiol. 2013;16:255-261

18. Lopez-Siles M, Martinez-Medina M, Busquets D, et al. Mucosa-associated *Faecalibacterium prausnitzii* and *Escherichia coli* co-abundance can distinguish Irritable Bowel Syndrome and Inflammatory Bowel Disease phenotypes. International Journal of Medical Microbiology. 2014;304:464-475

19. Sokol H, Seksik P, Furet JP, et al. Low counts of *Faecalibacterium prausnitzii* in colitis microbiota. Inflamm Bowel Dis. 2009;15:1183-1189

20. Swidsinski A, Loening-Baucke V, Vaneechoutte M, et al. Active Crohn's disease and ulcerative colitis can be specifically diagnosed and monitored based on the biostructure of the fecal flora. Inflamm Bowel Dis. 2008;14:147-161

21. Willing B, Halfvarson J, Dicksved J, et al. Twin studies reveal specific imbalances in the mucosa-associated microbiota of patients with ileal Crohn's disease. Inflamm Bowel Dis. 2009;15:653-660

22. Hansen R, Russell RK, Reiff C, et al. Microbiota of de-novo pediatric IBD: increased *Faecalibacterium prausnitzii* and reduced bacterial diversity in Crohn's but not in ulcerative colitis. Am J Gastroenterol. 2012;107:1913-1922

706 23. Jia W, Whitehead RN, Griffiths L, et al. Is the abundance of *Faecalibacterium*
707 *prausnitzii* relevant to Crohn's disease? FEMS Microbiol Lett. 2010;310:138-144
708 24. Kabeerdoss J, Sankaran V, Pugazhendhi S, et al. *Clostridium leptum* group
709 bacteria abundance and diversity in the fecal microbiota of patients with inflammatory
710 bowel disease: a case-control study in India. BMC Gastroenterol. 2013;13:20
711 25. Machiels K, Joossens M, Sabino J, et al. A decrease of the butyrate-producing
712 species *Roseburia hominis* and *Faecalibacterium prausnitzii* defines dysbiosis in
713 patients with ulcerative colitis. Gut. 2013
714 26. McLaughlin SD, Clark SK, Tekkis PP, et al. The bacterial pathogenesis and
715 treatment of pouchitis. Therap Adv Gastroenterol. 2010;3:335-348
716 27. Vermeiren J, Van den Abbeele P, Laukens D, et al. Decreased colonization of
717 fecal *Clostridium coccoides/Eubacterium rectale* species from ulcerative colitis patients
718 in an in vitro dynamic gut model with mucin environment. FEMS Microbiol Ecol.
719 2012;79:685-696
720 28. Frank DN, St Amand AL, Feldman RA, et al. Molecular-phylogenetic
721 characterization of microbial community imbalances in human inflammatory bowel
722 diseases. Proc Natl Acad Sci U S A. 2007;104:13780-13785
723 29. Nagalingam NA, Lynch SV. Role of the microbiota in inflammatory bowel
724 diseases. Inflamm Bowel Dis. 2012;18:968-984
725 30. Rajilic-Stojanovic M, Biagi E, Heilig HG, et al. Global and deep molecular
726 analysis of microbiota signatures in fecal samples from patients with irritable bowel
727 syndrome. Gastroenterology. 2011;141:1792-1801
728 31. Ghoshal UC, Shukla R, Ghoshal U, et al. The gut microbiota and irritable bowel
729 syndrome: friend or foe? Int J Inflam. 2012;2012:151085

730 32. Spiller RC. Inflammation as a basis for functional GI disorders. Best Practice &
731 Research Clinical Gastroenterology. 2004;18:641-661
732 33. Balamurugan R, Rajendiran E, George S, et al. Real-time polymerase chain
733 reaction quantification of specific butyrate-producing bacteria, *Desulfovibrio* and
734 *Enterococcus faecalis* in the feces of patients with colorectal cancer. J Gastroenterol
735 Hepatol. 2008;23:1298-1303
736 34. Martinez-Medina M, Aldeguer X, Gonzalez-Huix F, et al. Abnormal microbiota
737 composition in the ileocolonic mucosa of Crohn's disease patients as revealed by
738 polymerase chain reaction-denaturing gradient gel electrophoresis. Inflamm Bowel Dis.
739 2006;12:1136-1145
740 35. Mondot S, Kang S, Furet JP, et al. Highlighting new phylogenetic specificities
741 of Crohn's disease microbiota. Inflamm Bowel Dis. 2011;17:185-192
742 36. Seksik P, Rigottier-Gois L, Gramet G, et al. Alterations of the dominant faecal
743 bacterial groups in patients with Crohn's disease of the colon. Gut. 2003;52:237-242
744 37. Silverberg MS, Satsangi J, Ahmad T, et al. Toward an integrated clinical,
745 molecular and serological classification of inflammatory bowel disease: Report of a
746 Working Party of the 2005 Montreal World Congress of Gastroenterology. Can J
747 Gastroenterol. 2005;19 Suppl A:5-36
748 38. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity
749 of progressive multiple sequence alignment through sequence weighting, position-
750 specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22:4673-4680
751 39. Maidak BL, Cole JR, Lilburn TG, et al. The RDP-II (Ribosomal Database
752 Project). Nucleic Acids Research. 2001;29:173-174

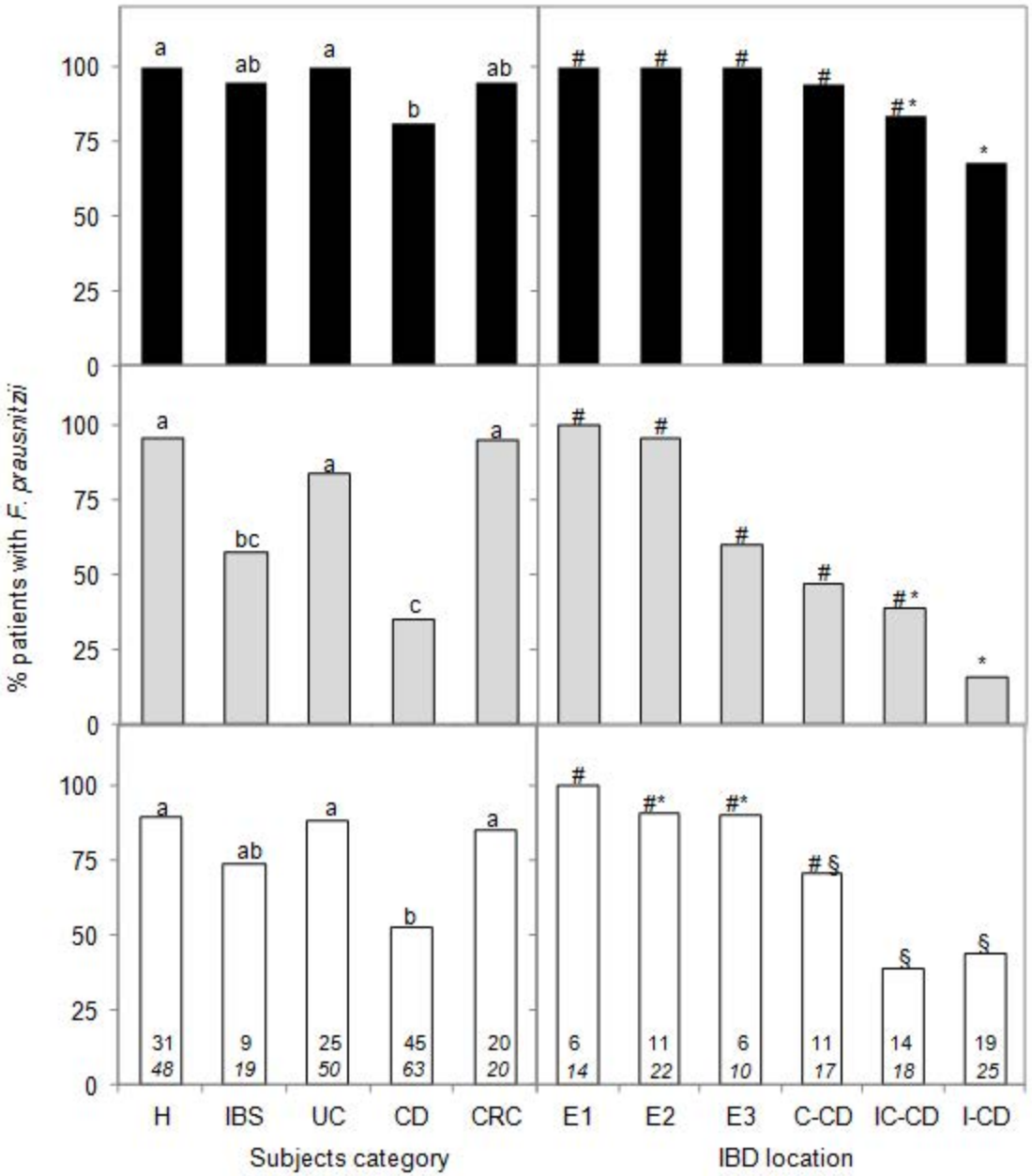
753 40. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST:
754 a new generation of protein database search programs. Nucleic Acids Research.
755 1997;25:3389-3402
756 41. Lane DJ. 16S/23S rRNA sequencing. New York: In E. Stackebrandt and M.
757 Goodfellow (ed.)., John Willy and Sons; 1991
758 42. Weisburg WG, Barns SM, Pelletier DA, et al. 16S ribosomal DNA amplification
759 for phylogenetic study. J Bacteriol. 1991;173:697-703
760 43. Huijsdens XW, Linskens RK, Mak M, et al. Quantification of Bacteria Adherent
761 to Gastrointestinal Mucosa by Real-Time PCR. J Clin Microbiol. 2002;40:4423-4427
762 44. Furet J-P, Firmesse O, Gourmelon M, et al. Comparative assessment of human
763 and farm animal faecal microbiota using real-time quantitative PCR. FEMS
764 Microbiology Ecology. 2009;68:351-362
765 45. Martinez-Medina M, Aldeguer X, Lopez-Siles M, et al. Molecular diversity of
766 Escherichia coli in the human gut: new ecological evidence supporting the role of
767 adherent-invasive E. coli (AIEC) in Crohn's disease. Inflamm Bowel Dis. 2009;15:872-
768 882
769 46. Best WR, Becktel JM, Singleton JW, et al. Development of a Crohn's disease
770 activity index. National Cooperative Crohn's Disease Study. Gastroenterology.
771 1976;70:439-444
772 47. Pineton de Chambrun G, Peyrin-Biroulet L, Lemann M, et al. Clinical
773 implications of mucosal healing for the management of IBD. Nat Rev Gastroenterol
774 Hepatol. 2010;7:15-29
775 48. Wang T, Cai G, Qiu Y, et al. Structural segregation of gut microbiota between
776 colorectal cancer patients and healthy volunteers. ISME J. 2012;6:320-329

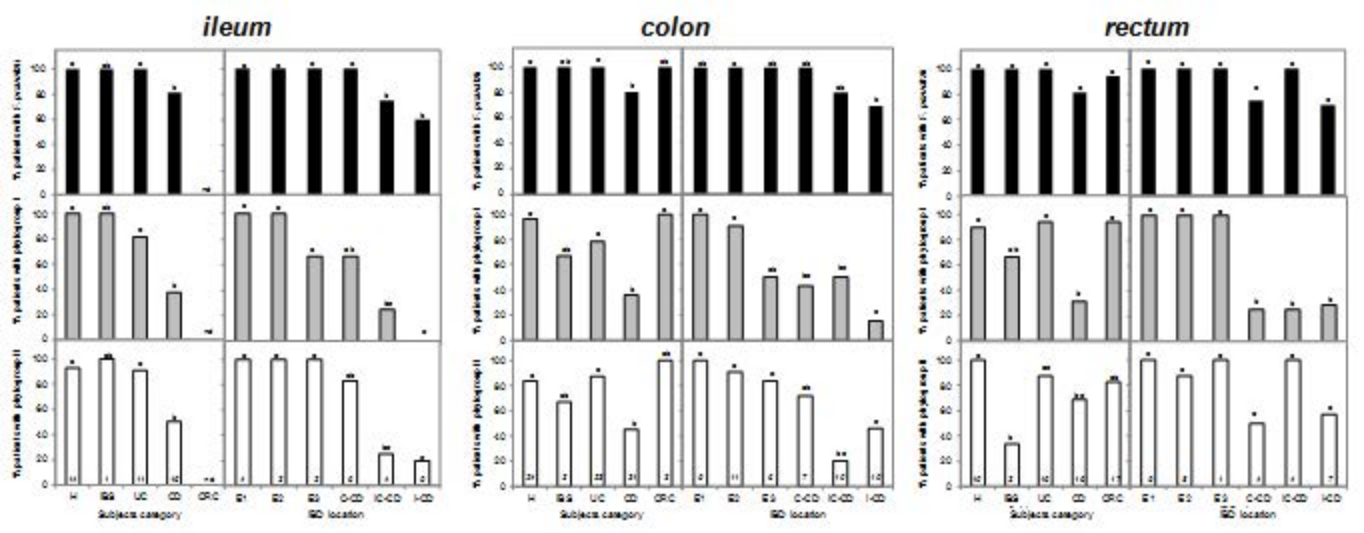
777 49. Sobhani I, Tap J, Roudot-Thoraval F, et al. Microbial dysbiosis in colorectal
778 cancer (CRC) patients. PLoS One. 2011;6:e16393
779 50. Duboc H, Rainteau D, Rajca S, et al. Increase in fecal primary bile acids and
780 dysbiosis in patients with diarrhea-predominant irritable bowel syndrome.
781 Neurogastroenterol Motil. 2012;24:513-520, e246-517
782 51. Kassinen A, Krogius-Kurikka L, Makivuokko H, et al. The fecal microbiota of
783 irritable bowel syndrome patients differs significantly from that of healthy subjects.
784 Gastroenterology. 2007;133:24-33
785 52. Malinen E, Rinttila T, Kajander K, et al. Analysis of the Fecal Microbiota of
786 Irritable Bowel Syndrome Patients and Healthy Controls with Real-Time PCR. Am J
787 Gastroenterol. 2005;100:373-382
788 53. Rigsbee L, Agans R, Shankar V, et al. Quantitative Profiling of Gut Microbiota
789 of Children With Diarrhea-Predominant Irritable Bowel Syndrome. Am J Gastroenterol.
790 2012;107:1740-1751
791 54. Gerasimidis K, Bertz M, Hanske L, et al. Decline in presumptively protective
792 gut bacterial species and metabolites are paradoxically associated with disease
793 improvement in pediatric Crohn's disease during enteral nutrition. Inflamm Bowel Dis.
794 2014;20:861-871
795 55. Dorffel Y, Swidsinski A, Loening-Baucke V, et al. Common biostructure of the
796 colonic microbiota in neuroendocrine tumors and Crohn's disease and the effect of
797 therapy. Inflamm Bowel Dis. 2012;18:1663-1671
798 56. Maccaferri S, Vitali B, Klinder A, et al. Rifaximin modulates the colonic
799 microbiota of patients with Crohn's disease: an in vitro approach using a continuous
800 culture colonic model system. J Antimicrob Chemother. 2010;65:2556-2565
801
802

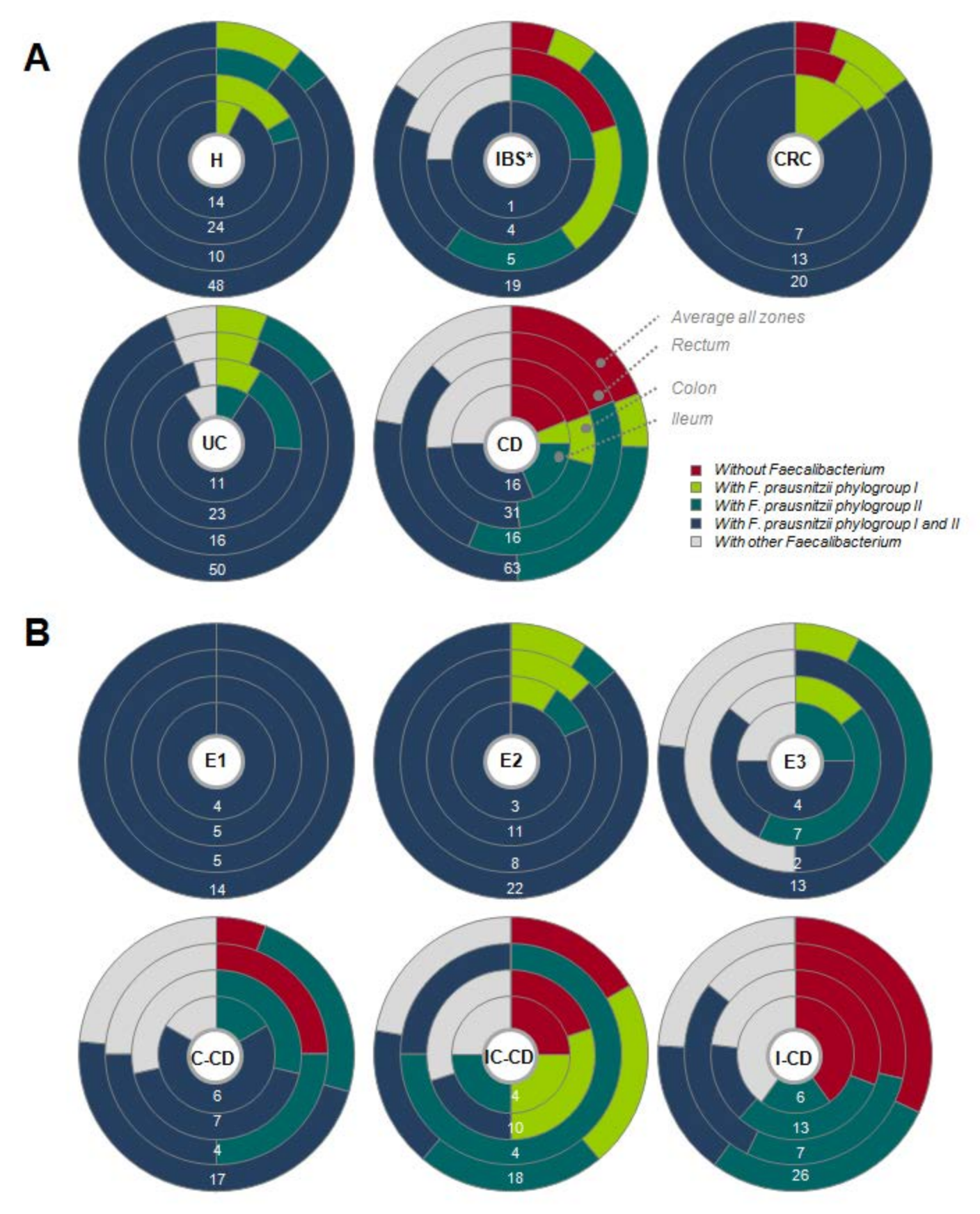
803 **LIST OF SUPPLEMENTAL DIGITAL CONTENT (SCD)**

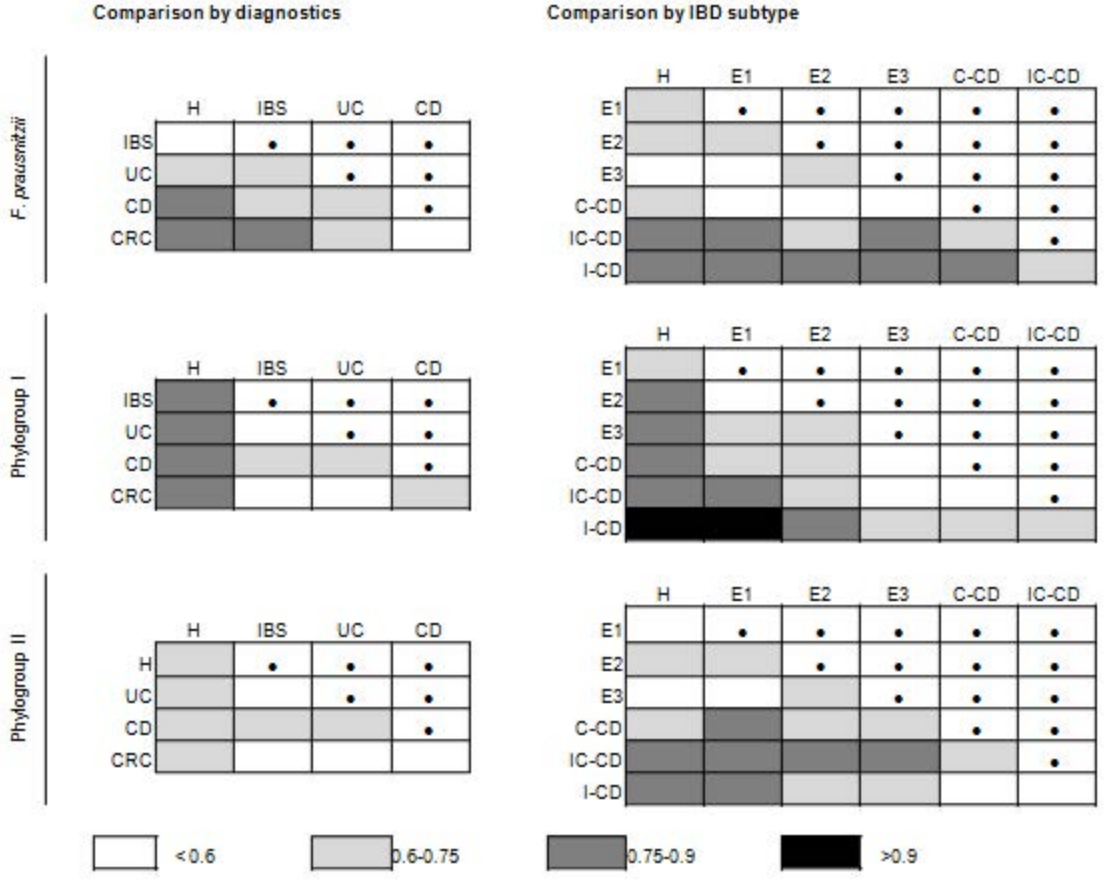
SCD	SCD NUMBER	FILE TYPE
Supplemental Digital Content	1	Text

804


% patients with F. prausnitzii
100
75
50
25
0
a
ab
a
b
ab
#
#
#
#
#*
*
a
bc
a
c
a
#
#
#
#
#*
*
a
ab
a
b
a
#
#*
#*
§
§
§
31
48
9
19
25
50
45
63
20
20
6
14
11
22
6
10
11
17
14
18
19
25
H
IBS
UC
CD
CRC
E1
E2
E3
C-CD
IC-CD
I-CD
Subjects category
IBD location


ileum
colon
rectum
Subjects category
IBD location


A
H
14
24
10
48
IBS*
1
4
5
19
CRC
7
13
20
UC
11
23
16
50
CD
16
31
16
63
Average all zones
Rectum
Colon
Ileum
Without Faecalibacterium
With F. prausnitzii phylogroup I
With F. prausnitzii phylogroup II
With F. prausnitzii phylogroup I and II
With other Faecalibacterium
B
E1
4
5
5
14
E2
3
11
8
22
E3
4
7
2
13
C-CD
6
7
4
17
IC-CD
4
10
4
18
I-CD
6
13
7
26


Comparison by diagnostics
Comparison by IBD subtype
F. prausnitzii
Phylogroup I
Phylogroup II
H
IBS
UC
CD
CRC
E1
E2
E3
C-CD
IC-CD
I-CD
< 0.6
0.6-0.75
0.75-0.9
>0.9

TABLES

Table 1. Sample size and clinical characteristics of subjects.

	Healthy*	Irritable bowel syndrome	IBD		Colorectal cancer	p value[§]
			Crohn's disease	Ulcerative colitis		
N (patients)	31	10	45	25	20	
Age (mean years ± SD)	48.1±16.3	42.4±11.4	33.5±11.1	40.1±15.8	58.6±7.52	<0.001[‡]
Male (N, %)	16 (51.6%)	2 (20.0%)	26 (57.7%)	16 (64.0%)	14 (70.0%)	0.605[†]
Active (N, %)	na	na	28 (62.2%)	20 (80.0%)	na	0.059[†]
Previous surgery (N, %)	0	nd	9 (20.0%)	1 (4.0%)	nd	0.049[†]
Smokers (N, %)	0	0	8 (17.8%)	2 (8.0%)	5 (25.0%)	0.327[†]
Treatment (N, %) **						0.087[†]
No treatment			12 (26.7%)	13 (52.0%)		
Mesalazine	na	na	3 (6.7%)	3 (12.0%)	na	
Moderate immunosuppressant	na	na	16 (35.5%)	3 (12.0%)	na	
Anti-TNFα (infliximab, adalimumab)	na	na	10 (22.2%)	4 (16.0%)	na	
CD Montreal classification						
Age of diagnosis (N, %) **						0.257[‡]
diag < 16y (A1)	na	na	5 (11.1%)	1 (4.0%)	nd	
diag 17-40y (A2)	na	na	33 (73.3%)	13 (52.0%)	nd	
diag >41y (A3)	na	na	5 (11.1%)	8 (32.0%)	nd	
Location (N, %)						na
Ileal-CD (L1)	na	na	19 (42.2%)	na	na	
Colonic-CD (L2)	na	na	11 (24.4%)	na	na	
Ileocolonic-CD (L3)	na	na	14 (31.1%)	na	na	
Behavior (N, %) **						na
Non-stricturing, non-penetrating (B1)	na	na	30 (66.7%)	na	na	
Stricturing (B2)	na	na	9 (20.0%)	na	na	
UC classification (N, %) **						na
Ulcerative proctitis (E1)	na	na	na	6 (24.0%)	na	
Distal UC (E2)	na	na	na	11 (44.0%)	na	
Extensive UC or ulcerative pancolitis (E3)	na	na	na	6 (24.0%)	na	
IBS subtype (N, %) **						na
Diarrhea predominant type	na	2 (20.0%)	na	na	na	
Constipation predominant type	na	2 (20.0%)	na	na	na	
CRC subtype (N, %) **						na
Sporadic	na	na	na	na	14 (70.0%)	
Hereditary***	na	na	na	na	3 (15.0%)	

IBD, Inflammatory bowel disease; IBS, Irritable bowel syndrome; CRC, colorectal cancer; TNF, tumor necrosis factor; 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 41/45 CD patients, and 23/25 UC patients; Age of disease onset was available for 43/45 CD patients, and 22/25 UC patients; Disease behavior at last follow-up before the time of sampling was available in 39/45 CD patients, and none had penetrating CD (B3); Maximal disease extent at the time of sampling was available in 23/25 UC patients; disease subtype was available in 4/10 Irritable bowel syndrome patients, and none had alternating predominant type; presence or absence of relatives with CRC could only be clearly tracked in 17/20 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

Table 2. 16S rRNA-targeted primers and probes used in this study.

Target	Primer and Probes [a]			PCR conditions [c]		
	Name	**Sequence 5'-3'**	**Reference**	**Total cycles**	**Denaturing (ºC; s)**	**Annealing and extension (ºC; s)**
Bacteria	F_Bact 1369	CGGTGAATACGTTCCCGG	(44)	50	95; 15	60; 60
	R_Prok_1492	TACGGCTACCTTGTTACGACTT				
	P_TM_1389F	**6FAM-CTTGTACACACCGCCCGTC-TAMRA**				
F. prausnitzii (total)	Fpra 428 F	TGTAAACTCCTGTTGTTGAGGAAGATAA	(18)	40	95; 15	60; 60
	Fpra 583 R	GCGCTCCCTTTACACCCA				
	Fpra 493 PR	**6FAM-CAAGGAAGTGACGGCTAACTACGTGCCAG-TAMRA**				
DNA IAC [b]	IAC F	TACGGATGAGGAGGACAAAGGA	(18)	40	95; 15	60; 60
	IAC R	CACTTCGCTCTGATCCATTGG				
	IAC PR	**VIC®-CGCCGCTATGGGCATCGCA-TAMRA**				
E. coli	E.coli 395 F	CATGCCGCGTGTATGAAGAA	(43)	40	95; 15	60; 60
	E.coli 490 R	CGGGTAACGTCAATGAGCAAA				
	E.coli 437 PR	**6FAM-TATTAACTTTACTCCCTTCCTCCCCGCTGAA-TAMRA**				
F. prausnitzii (phylogroups)	Fpra 136F	CTCAAAGAGGGGGACAACAGTT	this study	50	95; 15	64; 60
	Fpra 232R	GCCATCTCAAAGCGGATTG				
	PHG1 180PR	**6FAM-TAAGCCCACGACCCGGCATCG-BHQ1**				
	PHG2 180PR	**JOE-TAAGCCCACRGCTCGGCATC-BHQ1**				

[a] Probe sequences are in bold. FAM™ (6-carboxyfluorescin), VIC® (6-carboxyrhodamine), JOE (4',5'-dichloro-2',7'-dimethoxy-5(6)-carboxyfluorescein), TAMRA™ (tetramethylrhodamin) BHQ1 (Black Hole Quencher1).

[b] IAC, Internal Amplification Control; DNA IAC sequence: 5'-TACGGATGAGGAGGACAAAGGACGCCGCTATGGGCATCGCACCAATGGATCAGAGCGAAGTG-3' (according to Lopez-Siles et al 2014).

[c] For all quantitative PCR, an initial step at 50ºC during 2 min was performed for amperase treatment. Also an initial denaturation step was set at 95ºC for 10 min. In quantitative PCR, annealing and extension steps were performed simultaneously.

Table 3. Abundances of mucosa-associated *F. prausnitzii* and its phylogroups in controls (H), Irritable Bowel Syndrome (IBS), Ulcerative Colitis (UC), and Crohn's disease (CD) patients. Disease locations of UC and CD patients are analyzed as independent groups.

	n patients ***(n biopsies)***	***F. prausnitzii**§**	**Phylogroup I*§**	**Phylogroup II*§**
H	**31** ***(48)***	**5.33±0.58** [a]	**3.39±0.87** [a]	**3.39±1.51** [a]
IBS	**9** ***(19)***	**5.29±0.54** [a,b]	**2.53±1.22** [a,b]	**2.72±1.06** [a,b]
CRC	**20** ***(20)***	**4.42±0.58** [c]	**2.66±0.91** [b]	**2.56±1.14** [a,b]
UC	**25** ***(50)***	**5.00±0.62** [b]	**2.59±1.24** [b]	**2.93±0.99** [a]
Location				
Ulcerative proctitis (E1)	6 *(14)*	5.09±0.29 [a]	2.76±0.38 [a,b]	3.22±0.43 [a]
Distal UC (E2)	11 *(22)*	4.49±0.59 [b]	2.58±1.15 [a, b]	2.84±0.93 [a,b]
Extensive UC or ulcerative pancolitis (E3)	6 *(10)*	5.34±0.69 [a]	0.95±1.60 [b,c]	3.13±1.02 [a,b]
CD	**45** ***(63)***	**4.26±1.34** [c]	**0.71±1.65** [c]	**1.54±1.47** [c]
Location				
Ileal-CD (L1)	19 *(25)*	3.97±1.42 [c]	0.43±1.33 [c]	1.14±1.54 [b]
Colonic-CD (L2)	11 *(17)*	5.06±1.07 [a,c]	1.54±1.71 [b c]	2.63±1.51 [a,b]
Ileocolonic-CD (L3)	14 *(18)*	4.30±1.12 [b, c]	1.06±1.72 [b,c]	1.38±1.54 [b]

* Statistics was calculated separately for each variable (column). Groups of patients with similar abundances of *F. prausnitzii* or its phylogroups are indicated with the same superscript (a,b or c) whereas groups not sharing superscript are those with statistically different median abundance values ($P<0.05$)

§ Median $\log_{10}$ 16S rRNA gene copies/ million bacterial 16S rRNA gene copies ± standard deviations

Table 4. Correlation between *F. prausnitzii* and its phylogroups abundances, and between phylogroups abundances in controls (H), Irritable Bowel Syndrome (IBS), Ulcerative Colitis (UC), and Crohn's disease (CD) patients.

		***F. prausnitzii* vs phylogroup I**		***F. prausnitzii* vs phylogroup II**		**Phylogroup I vs phylogroup II**	
Diagnostics	**N patients (*N biopsies*)**	ρ	**P**	ρ	**P**	ρ	**P**
H	**31 *(48)***	**0.573**	**<0.001**	**0.741**	**<0.001**	**0.716**	**<0.001**
CRC	**20 *(20)***	**0.626**	**0.003**	**0.177**	**0.456**	**0.190**	**0.422**
IBS	**9 *(19)***	**0.327**	**0.172**	**0.284**	**0.239**	**0.217**	**0.373**
UC	**25 *(50)***	**0.671**	**<0.001**	**0.677**	**<0.001**	**0.667**	**<0.001**
CD	**45 *(63)***	**0.618**	**<0.001**	**0.743**	**<0.001**	**0.565**	**<0.001**

Supplemental materials

Changes in the abundance of *Faecalibacterium prausnitzii* phylogroups I and II in the intestinal mucosa of inflammatory bowel disease and colorectal cancer patients.

Mireia Lopez-Siles, Margarita Martinez-Medina, Romà Surís-Valls, Xavier Aldeguer, Miriam Sabat-Mir, Sylvia H. Duncan, Harry J. Flint, and L. Jesús Garcia-Gil.

Supplementary text

Optimisation and characterization of the multiplex qPCR assay for *F. prausnitzii* phylogroups

To determine the best reagent concentrations for the qPCR assay, experiments were performed using different primer and probe concentrations ranging from 50 to 900 nM. Those reagents concentrations that yield the maximum fluorescent signal and the lowest quantification cycle (C_q) value for 10^6 copies/reaction of the target DNA were chosen as optimal, and have therefore been used for further quantification in samples (as described in the main text section Quantitative PCR conditions).

Inclusivity and exclusivity tests

For the multiplex quantification of *F. prausnitzii* phylogroups using a qPCR assay, specificity was also tested *in vitro* by comparing the quantification of pure *F. prausnitzii* DNA (10 ng) recovered from nine isolates, representative of both phylogroups. DNA from 80 additional representative bacterial species (see list in Table S2) which are either close relatives of *F. prausnitzii* or belong to the major groups of bacteria present in the colon were also included.

F. prausnitzii strains were from stocks held by the authors (Rowett Institute of Nutrition and Health, Aberdeen, United Kingdom) and several came from previous studies (1-5). Additional bacterial strains were either available in our laboratory collection or were otherwise obtained from several biological resource centers specified in Table S2. When possible, bacteria were cultured aerobically or anaerobically on the recommended medium. DNA was extracted and purified by using the Wizard™ Genomic Purification Kit (Promega Corporation, USA) following the manufacturer's guidelines.

The qPCR was carried out as described in the section Quantitative PCR conditions of the main text. Negative results were cross checked by alternative amplification by end-point, conventional PCR with universal bacterial primers Bac27F and Uni1492R as previously reported (6, 7). Results from the specificity test are also shown in Table S2. The assay was totally specific. All the *F. prausnitzii* isolates were only detected for the phylogroup they belong to, and no statistically significant differences in C_q values between isolates were observed. There was no cross-reaction with any of the non-target microorganisms, and negative results were validated by a positive amplification by conventional PCR.

Linear quantification range and efficiency of the qPCR

To determine the confident quantification range of the assay, decaplicate tenfold dilutions (ranging from 2×10^8 to 2 target gene copies per reaction) of a linearized plasmid containing either a single copy of the 16S rRNA gene of *F. prausnitzii* S3L/3 (phylogroup I) or *F. prausnitzii* DSM 17677 (phylogroup II) were used. The linear range for quantification was considered for those concentrations having a SD value lower than 0.34 between replicates. Regression analysis plotting the obtained Cq against the logarithm of the number of target genes in the reaction was also performed. The efficiency of the qPCR assay was calculated using the formula: Efficiency= $[10^{(-1/slope)}]$-1.

Detection limit of the assay.

A calibration curve of two-fold serial dilutions between 1 and 100 target copies of *F. prausnitzii* 16S rRNA gene was performed. Eight replicas of each dilution were assayed. Data was analyzed by a Probit test (Minitab® 14 Statistical Software, Pennsylvania, US), in which the ratio of positive/negative amplification events was plotted against the amount of target genes present per reaction.

Table S1. 16S rRNA gene sequences used to perform oligonucleotide design. GenBank accession numbers have been indicated. Sequences from *F. prausnitzii* isolates, related sequences recovered via molecular methods and sequences of the same gene from *F. prausnitzii* close relatives have been included.

Accession number	Characteristics
AJ413954*[1]	*Faecalibacterium prausnitzii* 16S rRNA gene, strain ATCC 27768
X85022*[1]	*F. prausnitzii* DNA for 16S ribosomal RNA, strain ATCC 27766
AY305307*[1]	Butyrate-producing bacterium M21/2 16S ribosomal RNA gene
HQ457025*[1]	*F. prausnitzii* strain S4L/4 16S ribosomal RNA gene
HQ457024*[1]	*F. prausnitzii* strain S3L/3 16S ribosomal RNA gene
AJ270469*[2]	Butyrate-producing bacterium A2-165 16S rRNA gene
AJ270470*[2]	Butyrate-producing bacterium L2-6 16S rRNA gene
JN037415*[2]	*F. prausnitzii* strain L2-15 16S ribosomal RNA gene
JN037416*[2]	*F. prausnitzii* strain L2-39 16S ribosomal RNA gene
JN037417*[2]	*F. prausnitzii* strain L2-61 16S ribosomal RNA gene
HQ457026*[2]	*F. prausnitzii* strain HTF-A 16S ribosomal RNA gene
HQ457027*[2]	*F. prausnitzii* strain HTF-B 16S ribosomal RNA gene
HQ457028*[2]	*F. prausnitzii* strain HTF-C 16S ribosomal RNA gene
HQ457029*[2]	*F. prausnitzii* strain HTF-E 16S ribosomal RNA gene
HQ457030*[2]	*F. prausnitzii* strain HTF-F 16S ribosomal RNA gene
HQ457031*[2]	*F. prausnitzii* strain HTF-I 16S ribosomal RNA gene
HQ457032*[2]	*F. prausnitzii* strain HTF-60C 16S ribosomal RNA gene
HQ457033*[2]	*F. prausnitzii* strain HTF-75H 16S ribosomal RNA gene
AY169429*	*Faecalibacterium prausnitzii* clone 1-84 16S ribosomal RNA gene, partial sequence
AY169430*	*Faecalibacterium prausnitzii* clone 1-88 16S ribosomal RNA gene, partial sequence
AY169427*	*Faecalibacterium prausnitzii* clone 1-79 16S ribosomal RNA gene, partial sequence
AF132237*	Uncultured bacterium adhufec13 16S ribosomal RNA gene, partial sequence[§]
AF132236*	Uncultured bacterium adhufec113 16S ribosomal RNA gene, partial sequence[§]
AF132246*	Uncultured bacterium adhufec218 16S ribosomal RNA gene, partial sequence[§]
AF132265*	Uncultured bacterium adhufec365 16S ribosomal RNA gene, partial sequence[§]
AY494671*	Uncultured *Faecalibacterium* sp. clone FIRM8 16S ribosomal RNA gene, partial sequence
EF205929*	Uncultured bacterium clone 46706[§]
EF205662*	Uncultured bacterium clone 58014[§]
EF206222*	Uncultured bacterium clone 56806[§]
EF206249*	Uncultured bacterium clone 57601[§]
EF205881*	Uncultured bacterium clone 35509[§]
EF205761*	Uncultured bacterium clone 59415[§]
EF205681*	Uncultured bacterium clone 58033[§]
X98011	*Anaerofilum agile* 16S rRNA gene
X97852	*Anaerofilum pentosovorans* 16S rRNA gene
L09177	*Clostridium cellulosi* 16S ribosomal RNA (16S rRNA) gene
M59095	*Clostridium leptum* 16S ribosomal RNA
AJ305238	*Clostridium leptum*; DSM 753T
M59116	*Clostridium sporosphaeroides* 16S ribosomal RNA
X66002	*Clostridium sporosphaeroides*; DSM 1294
X81125	*Clostridium viride* 16S rRNA gene
L34618	*Eubacterium desmolans* 16S ribosomal RNA
L34625	*Eubacterium siraeum* 16S ribosomal RNA
AY445600	*Ruminococcus albus* strain 7 16S ribosomal RNA gene, complete
AY445594	*Ruminococcus albus* strain 8 16S ribosomal RNA gene, complete
AY445592	*Ruminococcus albus* strain B199 16S ribosomal RNA gene, complete
AY445596	*Ruminococcus albus* strain KF1 16S ribosomal RNA gene, complete
AY445602	*Ruminococcus albus* strain RO13 16S ribosomal RNA gene, complete
X85099	*Ruminococcus bromii* 16S rRNA gene
L76600	*Ruminococcus bromii* small subunit ribosomal RNA (16S rDNA) gene
X85100	*Ruminococcus callidus* 16S rRNA gene
L76596	*Ruminococcus callidus* small subunit ribosomal RNA (16S rDNA)
AM915269	*Ruminococcus flavefaciens* partial 16S rRNA gene, type strain C94T=ATCC19208
AF030449	*Ruminococcus flavefaciens* strain ATCC 49949 16S ribosomal RNA, partial sequence
AY445599	*Ruminococcus flavefaciens* strain B146 16S ribosomal RNA gene, complete sequence
AY445597	*Ruminococcus flavefaciens* strain FD1 16S ribosomal RNA gene, complete sequence
AY445595	*Ruminococcus flavefaciens* strain JM1 16S ribosomal RNA gene, complete sequence
AY445593	*Ruminococcus flavefaciens* strain C94 16S ribosomal RNA gene, complete sequence
AY445603	*Ruminococcus flavefaciens* strain LB4 16S ribosomal RNA gene, complete sequence
AY445601	*Ruminococcus flavefaciens* strain JF1 16S ribosomal RNA gene, complete sequence
AY445598	*Ruminococcus flavefaciens* strain R13e2 16S ribosomal RNA gene, complete sequence

* Sequences used to obtain the *F. prausnitzii* 16S rRNA gene consensus sequence for oligonucleotides design

[1] Sequences used to obtain the *F. prausnitzii* phylogroup I 16S rRNA gene consensus sequence for specific hydrolysis probe design

Accession number	Characteristics

[2] Sequences used to obtain the *F. prausnitzii* phylogroup II 16S rRNA gene consensus sequence for specific hydrolysis probe design
§ Sequence of the genus *Faecalibacterium*

Table S2. Growth conditions and source of the bacterial strains used in this study. The results obtained from the specificity tests are also included.

Source of DNA information*		Growth [2]		Specificity test information			
Phylogeny	**Strain/source [1]**	**Media**	**T(ºC)**	**ng [3]**	**cnPCR**	**qPHG1**	**qPHG2**
Firmicutes							
Faecalibacterium prausnitzii ATCC 27768^T	ATCC 27768	M2GSC	37	10	+	+	-
F. prausnitzii M21/2	nd	M2GSC	37	10	+	+	-
F. prausnitzii S3L/3	nd	M2GSC	37	10	+	+	-
F. prausnitzii S4L/4	nd	M2GSC	37	10	+	+	-
F. prausnitzii A2-165	DSM17677	M2GSC	37	10	+	-	+
F. prausnitzii L2-15	nd	M2GSC	37	10	+	-	+
F. prausnitzii L2-39	nd	M2GSC	37	10	+	-	+
F. prausnitzii L2-6	nd	M2GSC	37	10	+	-	+
F. prausnitzii L2-61	nd	M2GSC	37	10	+	-	+
Anaerofilum agile	DSM4272	nc	nc	1.6	+	-	-
Eubacterium siraeum	DSM15702	nc	nc	6.9	+	-	-
Eubacterium halii	DSM17630	nc	nc	1	+	-	-
Clostridium viride	DSM6836	nc	nc	10	+	-	-
Clostridium leptum	DSM753	nc	nc	10	+	-	-
Ruminococcus albus	DSM20455	nc	nc	10	+	-	-
Clostridium acetobutylicum	CECT 979	AN	37	3.7	+	-	-
Clostridium botulinum type E	CECT4611	LiB	37	10	+	-	-
Bacillus cereus	NCTC11145	AN	30	10	+	-	-
Bacillus megaterium	DSM319	AN	30	10	+	-	-
Bacillus sp.	CECT 40	AN	30	10	+	-	-
Bacillus subtilis	NCTC10400	AN	30	2.3	+	-	-
Bacillus subtilis sups. spizizwnii	CECT 482	AN	30	10	+	-	-
Listeria grayi	CECT931	BHI	37	10	+	-	-
Listeria innocua	CECT 910	BHI	37	10	+	-	-
Paenibacillus polymyxa	DSM372	BHI	37	2.1	+	-	-
Staphylococcus aureus	ATCC9144	AN	37	10	+	-	-
Staphylococcus epidermidis	CECT 231	AN	37	10	+	-	-
Enterococcus avium	CECT 968	BHI	37	10	+	-	-
Enterococcus columbae	CECT 4798	BHI	37	10	+	-	-
Enterococcus durans	CECT 411	BHI	37	10	+	-	-
Enterococcus faecalis	CECT 481	BHI	37	10	+	-	-
Enterococcus faecium	CECT 410	BHI	37	10	+	-	-
Enterococcus gallinarum	CECT 970	BHI	37	10	+	-	-
Enterococcus mundtii	CECT 972	BHI	37	10	+	-	-
Lactobacillus acidophilus	CECT 903	MRS	30	6.3	+	-	-
Lactococcus lactis	CECT 185	MRS	30	3.8	+	-	-
Streptococcus agalactiae	CECT 183	BHI	37	7.2	+	-	-
Streptococcus anginosus	CECT 948	BHI	37	10	+	-	-
Streptococcus equi subsp. *equi*	CECT 989	BHI	37	10	+	-	-
Streptococcus equinus	CECT 213	BHI	37	10	+	-	-
Streptococcus intermedius	CECT 803	BHI	37	10	+	-	-
Streptococcus mutans	CECT 479	BHI	37	3.8	+	-	-
Streptococcus oralis	CECT 907	BHI	37	10	+	-	-
Streptococcus pneumoniae	CECT 993	BHI	37	10	+	-	-
Streptococcus pyogenes	CECT 598	BHI	37	10	+	-	-
Streptococcus salivarus	CECT 805	BHI	37	10	+	-	-
Streptococcus sanguinis	CECT 480	BHI	37	5.5	+	-	-
Streptococcus sobrinus	CECT 4034	BHI	37	6.5	+	-	-
Streptococcus suis	CECT 958	BHI	37	10	+	-	-
Streptococcus thermophilus	CECT 986	BHI	37	10	+	-	-
Streptococcus uberis	CECT 994	BHI	37	10	+	-	-
Actinobacteria							
Corynebacterium bovis	DSM20582	MRS	37	4.8	+	-	-
Kocuria rhizophila	DSM348	AN	30	2.3	+	-	-

Source of DNA information*		Growth [2]		Specificity test information			
Phylogeny	**Strain/source [1]**	**Media**	**T(ºC)**	**ng [3]**	**cnPCR**	**qPHG1**	**qPHG2**
Micrococcus luteus	CECT 241	AN	30	2.6	+	-	-
Mycobacterium phlei	CECT 3009	BHI	37	10	+	-	-
Streptomyces griseus	DSM40236	PDA	30	10	+	-	-
Bifidobacterium adolescentis	CECT 5781	AN	37	0.4	+	-	-
Bifidobacterium breve	CECT 4839	AN	37	2.0	+	-	-
Bacteroidetes							
Bacteroides fragilis	DSM2151	nc	nc	10	+	-	-
Bacteroides uniformis	DSM6597	nc	nc	10	+	-	-
Bacteroides vulgatus	DSM1447	nc	nc	10	+	-	-
Proteobacteria							
Methylophilus methylotrophus	DSM5691	CZ	30	10	+	-	-
Campylobacter jejuni	DSM4688	BA	37	10	+	-	-
Citrobacter freundii	CECT 401	AN	30	10	+	-	-
Enterobacter aerogenes	CECT 684	AN	30	10	+	-	-
Enterobacter cloacae	CECT 194	AN	30	10	+	-	-
Enterobacter sakazakii	CECT 858	AN	30	10	+	-	-
Enterobacter sakazakii	ATCC51329	AN	30	0.4	+	-	-
Enterobacter amnigenus (Sakazakii)	CECT 4078	AN	37	10	+	-	-
Enterobacter gergoviae (Sakazakii)	CECT 857	AN	37	10	+	-	-
Escherichia coli	CECT 100	AN	37	10	+	-	-
Escherichia coli	CECT 101	AN	37	10	+	-	-
Escherichia coli	CECT 105	AN	37	10	+	-	-
Escherichia coli	CECT 12242	AN	37	10	+	-	-
Escherichia coli	CECT 831	AN	37	10	+	-	-
Escherichia coli	CECT 4201	AN	37	10	+	-	-
Escherichia coli	CECT 4084	AN	37	10	+	-	-
Escherichia coli	CECT 405	AN	37	10	+	-	-
Escherichia coli	ATCC10536	AN	37	10	+	-	-
Klebsiella pneumoniae ssp. pneumoniae	CECT 143	AN	37	10	+	-	-
Proteus mirabilis	CECT 170	AN	37	10	+	-	-
Salmonella LT2	CECT878	AN	37	10	+	-	-
Salmonella TA98	CECT880	AN	37	10	+	-	-
Serratia marcescens	CECT846	AN	25	10	+	-	-
Shigella sonnei	CECT457	AN	37	10	+	-	-
Pseudomonas aeruginosa	CECT 532	AN	30	10	+	-	-
Pseudomonas fluorescens	CECT 378	AN	30	10	+	-	-
Pseudomonas mendocina	CECT320	AN	30	10	+	-	-
Pseudomonas putida	CECT 324	AN	30	4.1	+	-	-

* Specificity test with human Xsomal DNA (Eurogentec, Belgium) was also performed

[1] ATCC: American Type Culture Collection (Manassas, VA, USA); CECT: Colección Española de Cultivos Tipo (Valencia, Spain); DSMZ: Deutche Sammlung von Mikroorganismen and Zellkulturren (Braunschweig, Germany), NCTC: National Collection of Type Cultures (London,UK), nd: not deposited (stocks held by the authors, Rowett Institute of nutrition and Health, Aberdeen, United Kingdom).

[2] nc: not cultured. BHI (Brain Heart Infusion Broth), AN (Nutrient Agar), BA (Blood Agar), MRS (Man, Rogosa and Sharpe medium), LiB(Liver Broth, CECT medium #15), CZ (Colby and Zathman medium, DSMZ medium #606), PDA (Potato Dextrose Agar), M2GSC (modified Med2 of Hobson, (1))

[3] ng of genomic DNA used for the inclusivity/exclusivity test. When possible, 10ng was used. The DNA was obtained from 1ml of bacterial culture at the stationary growth phase or for nc strains, the dried culture directly obtained from the culture type collection was rehydrated with the appropriate buffer for DNA extraction and used for DNA purification.

Table S3. *F. prausnitzii* and its phylogroups abundance in inflammatory bowel disease patients by disease activity status. Active CD and UC were defined by a CDAI of >150 (47) and a Mayo score >3, respectively.

Diagnostics[§]	**N**	***F. prausnitzii****	p-value	**Phylogroup I***	p-value	**Phylogroup II***	p-value
UC							
active	41	4.80±0.41	0.344	2.62±1.32	0.720	2.92±1.02	0.623
inactive	8	5.02±0.66		2.69±0.78		3.18±0.87	
CD							
active	41	4.31±1.10	0.507	0.61±1.51	0.106	1.50±1.63	0.624
inactive	22	4.25±1.46		1.36±1.80		1.69±1.14	

[*] Median $\log_{10}$ 16S rRNA gene copies/ million bacterial 16S rRNA gene copies ± standard deviations
[§]UC, ulcerative colitis; CD, Crohn's disease

Table S4. *F. prausnitzii* and its phylogroups abundance in inflammatory bowel disease patients depending on whether or not they have had intestinal resection during the course of the disease.

Diagnostics[§]	**N**	***F. prausnitzii****	p-value	**Phylogroup I***	p-value	**Phylogroup II***	p-value
UC							
non-resected	43	4.85±0.61	1.000	2.51±1.21	0.136	2.92±0.96	0.727
resected	1	4.91		3.45		2.68	
CD							
non-resected	41	4.86±1.43	0.016	1.52±1.84	0.379	2.11±1.46	0.001
resected	13	3.74±0.78		0.45±1.07		0.65±0.84	

[*] Median $\log_{10}$ 16S rRNA gene copies/ million bacterial 16S rRNA gene copies ± standard deviations
[§]UC, ulcerative colitis; CD, Crohn's disease

Table S5. *F. prausnitzii* and its phylogroups abundances (median $\log_{10}$ 16S rRNA gene copies/ million bacterial 16S rRNA gene copies ± standard deviations) in inflammatory bowel disease by medication at sampling.

Diagnostics[§]	**N**	***F. prausnitzii****	p-value	**Phylogroup I***	p-value	**Phylogroup II***	p-value
UC							
No treatment	25	4.95±0.65		2.51±1.32		2.93±1.03	
Mesalazine	6	5.02±0.33	0.904	2.53±0.84	0.806	3.31±0.98	0.832
moderate immunosuppresants	9	4.56±0.58		2.75±0.41		2.85±0.71	
Anti-tumor necrosis factor	7	4.44±0.83		3.16±1.93		2.92±1.07	
CD							
No treatment	21	4.86±1.66		0.69±2.04		2.70±1.71	
Mesalazine	3	5.10±0.41	0.225	1.71±1.67	0.854	2.63±1.89	0.738
moderate immunosuppresants	19	4.01±0.95		0.71±1.45		1.23±1.48	
Anti-tumor necrosis factor	16	4.01±1.43		0.67±1.48		1.49±1.18	

[*] Median $\log_{10}$ 16S rRNA gene copies/ million bacterial 16S rRNA gene copies ± standard deviations
[§]UC, ulcerative colitis; CD, Crohn's disease

Supplemental references

1. Lopez-Siles M, Khan TM, Duncan SH, et al. 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. 2012;78:420-428
2. Barcenilla A, Pryde SE, Martin JC, et al. Phylogenetic relationships of butyrate-producing bacteria from the human gut. Appl Environ Microbiol. 2000;66:1654-1661
3. Cato EP, Salmon, Carolyn W, and W.E.C. Moore. *Fusobacterium prausnitzii* (Hauduroy et al.) Moore and Holdeman: emended description and designation of neotype strain. Int J Syst Bacteriol. 1974;24:225-229
4. Duncan SH, Hold GL, Harmsen HJ, et al. 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 2002;52:2141-2146
5. Louis P, Duncan SH, McCrae SI, et al. Restricted distribution of the butyrate kinase pathway among butyrate-producing bacteria from the human colon. J Bacteriol. 2004;186:2099-2106
6. Lane DJ. 16S/23S rRNA sequencing. New York: In E. Stackebrandt and M. Goodfellow (ed.)., John Willy and Sons; 1991
7. Weisburg WG, Barns SM, Pelletier DA, et al. 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol. 1991;173:697-703