# Quantitative real-time PCR analysis of bacterial biomarkers enable fast and accurate monitoring in inflammatory bowel disease

**Efe Sezgin** [Corresp., 1], **Gamze Terlemez** [1], **Berkay Bozkurt** [2], **Göksel Bengi** [3], **Hale Akpinar** [3], **İlker Büyüktorun** [3]

[1] Izmir Institute of Technology, Izmir, Turkey

[2] Izmir Institute of Technology, Izmır, Turkey

[3] Dokuz Eylül University, Izmır, Turkey

Corresponding Author: Efe Sezgin
Email address: efesezgin@iyte.edu.tr

Inflammatory bowel diseases (IBD) affect millions of people worldwide with increasing incidence. Ulcerative colitis (UC) and Crohn's disease (CD) are the two most common IBDs. There is no definite cure for IBD, and response to treatment greatly vary among patients. Therefore, there is urgent need for biomarkers to monitor therapy efficacy, and disease prognosis. We aimed to test whether qPCR analysis of common candidate bacteria identified from a patient's individual fecal microbiome can be used as a fast and reliable personalized microbial biomarker for efficient monitoring of disease course in IBD. Next generation sequencing (NGS) of 16S rRNA gene region identified species level microbiota profiles for a subset of UC, CD, and control samples. Common high abundance bacterial species observed in all three groups, and reported to be associated with IBD are chosen as candidate marker species. These species, and total bacteria amount are quantified in all samples with qPCR. Relative abundance of anti-inflammatory, beneficial *Faecalibacterium prausnitzii, Akkermansia muciniphila*, and *Streptococcus thermophilus* was significantly lower in IBD compared to control samples. Moreover, the relative abundance of the examined common species was correlated with the severity of IBD disease. The variance in qPCR data was much lower compared to NGS data, and showed much higher statistical power for clinical utility. The qPCR analysis of target common bacterial species can be a powerful, cost and time efficient approach for monitoring disease status and identify better personalized treatment options for IBD patients.

# Quantitative real-time PCR analysis of bacterial biomarkers enable fast and accurate monitoring in Inflammatory Bowel Disease

Efe SEZGİN[1,2*], Gamze TERLEMEZ[1], Berkay BOZKURT[3], Göksel BENGİ[4], Hale AKPINAR[4], İlker BÜYÜKTORUN[4]

[1] Biotechnology Program, Izmir Institute of Technology, Urla, Izmir, Turkey
[2] Department of Food Engineering, Izmir Institute of Technology, Urla, Izmir, Turkey
[3] Bioengineering Program, Izmir Institute of Technology, Urla, Izmir, Turkey
[4] Division of Gastroenterology, Department of Internal Medicine, Dokuz Eylül University School of Medicine, İzmir, Turkey

**Running Title:** Microbial biomarkers for monitoring IBD disease

***Correspondence to:**

Efe Sezgin (efesezgin@iyte.edu.tr)
Department of Food Engineering,
Izmir Institute of Technology,
Urla, Izmir, Turkey.

**Abstract**

Inflammatory bowel diseases (IBD) affect millions of people worldwide with increasing incidence. Ulcerative colitis (UC) and Crohn's disease (CD) are the two most common IBDs. There is no definite cure for IBD, and response to treatment greatly vary among patients. Therefore, there is urgent need for biomarkers to monitor therapy efficacy, and disease prognosis. We aimed to test whether qPCR analysis of common candidate bacteria identified from a patient's individual fecal microbiome can be used as a fast and reliable personalized microbial biomarker for efficient monitoring of disease course in IBD. Next generation sequencing (NGS) of 16S rRNA gene region identified species level microbiota profiles for a subset of UC, CD, and control samples. Common high abundance bacterial species observed in all three groups, and reported to be associated with IBD are chosen as candidate marker species. These species, and total bacteria amount are quantified in all samples with qPCR. Relative abundance of anti-inflammatory, beneficial *Faecalibacterium prausnitzii, Akkermansia muciniphila*, and *Streptococcus thermophilus* was significantly lower in IBD compared to control samples. Moreover, the relative abundance of the examined common species was correlated with the severity of IBD disease. The variance in qPCR data was much lower compared to NGS data, and showed much higher statistical power for clinical utility. The qPCR analysis of target common bacterial species can be a powerful, cost and time efficient approach for monitoring disease status and identify better personalized treatment options for IBD patients.

**Key Words:** Molecular biomarker, Crohn's disease, Ulcerative colitis, Inflammatory bowel disease, quantitative real-time PCR

## 1. Introduction

Inflammatory bowel diseases (IBD) are complex, heterogeneous diseases arising from chronic and uncontrolled inflammation of the gastrointestinal (GI) tract (Podolsky, 2002; Tontini *et al*., 2015). Microbiota, genetics, and environmental factors are suggested to be underlying factors for susceptibility to IBD (Albenberg *et al*., 2012). Ulcerative colitis (UC) and Crohn's disease (CD) are the two most common diseases categorized under IBD. Accurate IBD diagnosis requires examination of clinical, endoscopic, and histopathological characteristics, but none of the findings are definitive, and even some patients' differential diagnoses cannot be made. IBD has become a worldwide disease affecting millions of patients (Alatab *et al*., 2020). The biggest incidence ratios have been reported in Northern Europe and North America for CD and UC (Burisch and Munkholm, 2015).

IBD has important social, psychological and financial implications as well as the deterioration of health-related quality of life. IBD impresses personal life and imposes significant economic burden not only on the patient but also on the health care system, such as treatment costs, time lost from work, and reduced productivity at work (Mehta, 2016; Walter *et al*., 2020). The financial burden can be even higher as IBD also affects individuals at an early age. Given the big personal and cumulatively population level costs, there is great interest in identifying both useful biomarkers and techniques to assay these markers for IBD progression, therapy response, and control.

The definite cause of IBD is not known, so individual or population level biomarker screens to identify people at risk are not possible yet. Most IBD patients seek medical care at a later, more advanced stages of the disease, and early intervention to prevent disease progression is rare. So,

identification of biomarkers, and techniques to assay these markers to monitor therapy efficacy, and disease prognosis is of great importance.

Although a number of biomarkers are suggested for the diagnosis of IBD, however with questionable sensitivity and specificity for UC and CD (Soubieres and Poullis, 2016; Chen *et al*., 2020; Guo *et. al*. 2021), biomarkers for monitoring disease progression or response to therapy is lacking and development of such biomarkers is an active research area. Serological (Miranda-Garcia *et al*., 2016), metabolomic (Bjerrum *et al*., 2017; Keshteli *et al*., 2018; Notararigo *et al*., 2021), proteomic (Kalla *et al*., 2021), metagenomic (Zhou *et al*., 2018; Serrano-Gomez *et al*., 2021), and transcriptomic (Montero-Melendez *et al*., 2013) approaches have been reported. However, these large data driven 'omics' techniques are research based, rather expensive, time consuming, and their clinical utility is questionable. Therefore, faster, cheaper, more accurate techniques that can utilize already available equipment in hospitals or molecular diagnostic laboratories is necessary. Discovery of bacterial biomarkers by next generation sequencing and further quantification of selected bacterial species by quantitative real time PCR (qPCR) is well documented in literature (Machiels et al., 2014; Lopez-Siles et al., 2020; Mondot et al., 2016; Zhou et al., 2016; Lopez-Siles et al., 2014; Lopez-Siles et al., 2018; Pascal et al., 2017). Therefore, qPCR can be a highly efficient molecular tool for monitoring IBD progression and response to therapy.

Based on the urgent need for such reliable methods with possible clinical utility, we aimed to test whether qPCR analysis of common candidate bacterial species identified from a patient's individual fecal microbiome can be used as a fast and reliable personalized microbial biomarker for efficient monitoring of IBD. We focused on Turkish IBD patients because Turkey is among the countries with highest increase of IBD incidence (Can *et al*., 2019), but microbiota studies from Turkey is rather limited, and the utility of fast and accurate molecular techniques in IBD

monitoring has not been explored in this population. In addition to testing previously reported bacterial markers, we searched for novel bacterial species, and evaluated *S. thermophilus* as a biomarker in IBD.

## 2. Material and Methods

### 2.1. Sample Collection

Fecal and blood samples were collected from 18 IBD patients (six diagnosed with Ulcerative colitis and 12 diagnosed with Crohn's disease) and four healthy (control group) individuals in the Gastroenterology Department of Dokuz Eylul University Hospital (Izmir, Turkey). IBD patients (Ulcerative colitis and Crohn's disease) were diagnosed according to international guidelines based on clinical, endoscopic, histopathological, and radiological examinations (Stange *et al*., 2008; Bemelman *et al*., 2018). The healthy controls were candidates without any history of IBD or mucosal lesions in colonoscopy. Fecal samples gathered in sterile and airtight containers, and blood samples collected into EDTA tubes were transported to laboratory within six hours after collection. All study samples were kept at −80℃ until processing. Ethical approval was obtained from the Ethics Committee of the Dokuz Eylül University (2017/08-03). All participants provided informed consent in the format required by the Dokuz Eylül University ethics committee.

### 2.2. Genomic DNA extraction from blood samples

DNA was isolated from blood samples with the Genomic DNA Mini Kit (Blood/Cultured cell) (Geneaid Biotech Ltd., Taiwan) following the manufacturer's protocol. Quality of extracted DNAs (A260/A280 and A260/A230 ratios) was checked with Nanodrop 8000c Spectrophotometer (Thermo Fisher Scientific, USA).

### 2.3. Primer design for candidate variants in IBD associated candidate genes

We targeted rs2066844 (Arg702Trp), rs2066845 (Gly908Arg), rs2066847 (Leu1007insC) SNPs for the *NOD2* gene; rs11209026 (Arg381Gln) for the *IL-23R* gene, and rs2241880 (Thr300Ala)

for the *ATG16L1* gene. DNA sequences for each gene were obtained from NCBI (https://www.ncbi.nlm.nih.gov/), ENSEMBL (https://www.ensembl.org/index.html), and UCSC Genome Browser (https://genome.ucsc.edu/cgi-bin/hgGateway) databases. These gene sequences were used to design polymerase chain reaction (PCR) primers using the IDT SciTools PCR algorithm (Integrated DNA Technologies) (Supplemental Table 1). Oligonucleotide properties, melting temperature, hairpins, dimers, and mismatches were identified by IDT SciTools OligoAnalyzer 3.1 (https://www.idtdna.com/calc/analyzer) software (Owczarzy et al., 2008), and specificity of primers were confirmed with Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/).

### 2.4. Amplification and sequencing of *IL-23R*, *ATG16L1* and *NOD2* variants

With primers designed for the specified SNPs, PCR analysis was performed using the FastStart High Fidelity PCR System, dNTPack kit (Roche Applied Science, Germany). Reaction mixes were made separately for *IL-23R*, *NOD2*, and *ATG16L1* genes in a final volume 25 µl of using 17.25 µl PCR-grade water, 0.5 µl forward and reverse primers, 0.5 µl PCR Grade Nucleotide Mix, 2.5 µl FastStart High Fidelity Reaction Buffer, 0.5 µl dimethyl sulfoxide (DMSO), 0.25 µl FastStart High Fidelity Enzyme Blend, and 3 µl DNA. The thermal cycling was subjected to the following conditions: denaturation at 94 °C for 10 min followed by 35 cycles of 94 °C for 2 minutes, annealing at 57 °C for 30 seconds, and elongation at 72 °C for 1 minute using SimpliAmp Thermal Cycler (ThermoFisher Scientific, USA).

PCR products were verified by agarose gel electrophoresis. Briefly, 5 µL of PCR product was mixed 1 µL 6X DNA loading buffer, and run on 1.4% agarose gel in 0.5X TBE buffer under a steady voltage of 100 V for 60 min at room temperature.

PCR products were purified using ExoSAP-IT™ PCR Product Cleanup Reagent (Thermo Fisher Scientific, USA). 10 µl sequencing mixture contained 4µl dd$H_2O$, 1 µl 5X ABI Buffer, 1µl of primer (3.2 pmol/µl), 2 µl BigDye Ready Reaction Mix, and 2 µl of the PCR product. Samples were sequenced on ABI Prism 3130xl Genetic Analyzer (Applied Biosystems, USA). Sequencing results (ABI chromatograms) were analyzed in Unipro UGENE v.33 (Okonechnikov et al., 2012) program. Multiple sequence alignments using ClustalW algorithm were performed in Unipro UGENE v.33 program (Okonechnikov et al., 2012).

**2.5. Bacterial DNA isolation from stool samples**

DNA was extracted from stool samples using the QIAamp DNA Stool Mini Kit (Qiagen, Germany) according to the manufacturer's protocol. The concentration (ng/µL) and purity (A260/A280 and A260/A230 ratios) of the DNA samples were determined by Nanodrop 8000c Spectrophotometer (Thermo Fisher Scientific, USA).

**2.6. 16S rRNA gene amplicon sequencing by next generation sequencing (NGS)**

Nine stool samples (three samples from each UC, CD patients, and healthy volunteers) were chosen for amplicon analysis. The variable V3-V4 region of 16S rRNA gene was targeted and amplified with the following PCR primers: 341F (5'-CCTAYGGGRBGCASCAG-3') and 806R (5'-GGACTACNNGGGTATCTAAT-3'). After the purification of PCR products, sequencing libraries were generated with Nextera XT DNA Library Preparation Kit (Illumina, USA). The concentration of sequencing libraries are standardized to 4nM each. Normalized samples were pooled and sequenced by Illumina NovaSeq 6000 as paired-end (2×250 bp) using the manufacturer's standard procedure. Raw data quality control check was performed by FastQC, and quality control of the reads was checked by QIIME2 (Caporaso et al., 2010). Effective tags were

obtained after removing primer and barcode sequences, chimeric reads, and reads with Phred Score less than 20 by DADA2 (Callahan et al., 2016). By utilizing the effective tags, representative sequence for each Operational Taxonomic Units (OTUs) were acquired with ≥97% similarity against the Greengenes and SILVA databases. QIIME2 was used for taxonomic determination of each OTU. Rarefaction curves plotting sequencing depth vs. number of taxa identified were used to judge the appropriateness of sequencing depth for each sample (Pereira-Marques et al., 2019; Zaheer et al., 2018).

Species diversity within samples were assessed by five different Alpha diversity estimates including Observed-species, Chao1, Shannon, Simpson, and ACE indices. Alpha diversity indices were visualized via boxplots. All downstream analyzes were performed with "phyloseq" (McMurdie and Holmes, 2013) and "ggplot2" packages in R software (Version 4.0.5) (https://www.r-project.org).

### 2.7. Quantification of selected bacterial species levels using Real-Time quantitative PCR (qPCR) analysis

Six bacterial species (*Faecalibacterium prausnitzii, Clostridioides difficile, Akkermansia muciniphila, Bacteroides vulgatus, Streptococcus thermophilus*, and Shiga toxin-producing *Escherichia coli* (*stx1* gene positive) were chosen as bacterial biomarkers for further quantification via qPCR in all samples. Primers were designed using the IDT SciTools® platform (Owczarzy et al., 2008). Primer sequences for each species and total bacterial quantification are shown in Table 1. In each run both positive and negative controls were added to qPCR-plates. For positive control *A. muciniphila* (ATCC BAA-835), *B. vulgatus* (ATCC 8482), *S. thermophilus* (ATCC 19258), *F. prausnitzii* (ATCC 27766), *E. coli* (ATCC 43890), and *C. difficile* (ATCC 9689) were utilized. 2.5 µl of distilled water was used as a negative control in each run.

The reactions were conducted with 2.5 µl DNA, 1.9µl PCR-grade water, 0.3 µl primer, and 5 µl LightCycler® 480 SYBR Green I Master enzyme (Roche Applied Science, Germany) in a total volume of 10µl. All reactions were carried out with triple replicates on a LightCycler® 480 II (Roche Applied Science, Germany) qPCR machine. Each sample and species is assayed at least twice. The qPCR reaction was 95 °C for 10 min with initial denaturation followed by 50 cycles of 95 °C for 10 s, 58 °C for 15 s, and 72 °C for 15 s. Melting curve analysis for the qPCR products was performed under the following conditions: 95 °C for 5 s, 63 °C for 1 min and a denaturing temperature ramp from 63 to 97 °C with a rate of 0.11 °C/s. Amplification and melting curves for each sample were obtained using Absolute Quantitation/Second Derivative and Tm Calling analysis modes in the LightCycler® 480 II Software v.1.5.

SYBR Green dye fluorescence intensity was used for quantification. The target bacterial DNA concentration correlated with the threshold cycle number (Ct), the cycle number at which fluorescence signal was first detected. Roche LightCycler® 480 System melting curve program analysis is used for the confirmation of success of qPCR reactions (Simenc and Potocnik, 2011). Data analysis was conducted using both the $2^{-\Delta Ct}$ , and $2^{-\Delta\Delta Ct}$ methods (Livak and Schmittgen, 2001). Target microorganisms were considered as a target while total bacteria measurement was used as a reference (Navidshad *et al*., 2012). Relative abundance of bacteria was expressed as log2 transformed fold change values, and calculated according to the following formulas.

Relative abundance of target bacteria species with respect to abundance of total bacteria:

$$2^{-\Delta Ct} = 2^{-(\text{Ct of target bacteria} - \text{Ct of total bacteria})}$$

Fold change of relative abundance of target bacteria in IBD patients compared to healthy controls

$$2^{-\Delta\Delta Ct} = 2^{-[(\text{Ct of target bacteria} - \text{Ct of total bacteria})\ \text{patient} - (\text{Ct of target bacteria} - \text{Ct of total bacteria})\ \text{control}]}$$

To verify PCR efficiency, standard curves were generated by 10-fold dilutions of bacterial DNA for all primer sets. In all sets, qPCR efficiency was >90% and calculated by $E= 10^{(-1/slope)-1}$ equation. According to the serial dilutions, the limit of detection of qPCR assays was 10-100 copies.

**2.8. Statistical analysis**

Distribution of candidate gene's genotypes among UC, CD, and control groups was compared with Chi-square categorical analyses. Shapiro–Wilk test was used to test the normality assumption of numeric variables. All microbial diversity and quantity parameter estimates violated normality assumption, so groups were compared using the non-parametric Kruskal-Wallis and pairwise Wilcoxon rank-sum tests. P-value less than 0.05 was assessed as statistically significant. All statistical analyzes were performed with R software (Version 4.0.5) (https://www.r-project.org). We also estimated the statistical power to detect a 10% change with 95% confidence in the numeric abundances of *A. muciniphila, B. vulgatus, S. thermophilus*, and *F. prausnitzii* in UC and CD groups based on qPCR relative abundance estimates. Statistical power calculations followed formulations in (Cohen, J. 1988) as implemented in the *pwr* package (https://cran.r-project.org/web/packages/pwr/index.html) of R software. Experimentally determined relative abundance means and standard errors of *A. muciniphila, B. vulgatus, S. thermophilus*, and *F. prausnitzii* are used for statistical power calculations, where Type I error probability (alpha) is set to 0.05.

## 3. Results

### 3.1. Patient characteristics

UC and CD patients consisted of similar age and sex groups, however, included a variety of disease locations and disease activity phenotypes (Table 2, Supplementary Table 2). Distribution of genotypic frequencies of *ATG16L1*, *IL23R* and *NOD2* variants in UC and CD patients was similar (Supplementary Table 3), suggesting similar IBD genetic risk profile in these patient groups.

### 3.2. Identification of fecal microbiota profile in the IBD patients and controls

An average of 148545 reads per sample (range 125105-172435) were generated for the 9 fecal samples (three samples from each of CD, UC, and control groups). Eight phyla, 15 classes, 23 orders, 43 families, 96 genus, and 233 species were represented in all sequences based on a 97% similarity level. The most abundant phyla in CD samples were Firmicutes (36.22%), Proteobacteria (29.22%), Verrucomicrobia (25.79%), Bacteroidetes (8.68%), Actinobacteria (0.07%), Fusobacteria (0.01%), and Synergistetes (0.01%). Similarly in UC samples, Proteobacteria (45.63%), Firmicutes (29.21%), Bacteroidetes (10.14%), Fusobacteria (9.98), Actinobacteria (2.80%), and Verrucomicrobia (2.25%) were observed (Figure 1A). The most common phyla in the control samples were Bacteroidetes (62.63%), followed by Firmicutes (31.47%), Proteobacteria (4.01%), Actinobacteria (1.81%), Fusobacteria (0.04%), Lentisphaerae (0.04%), and Verrucomicrobia (0.01%) (Figure 1A). Despite small sample size Kruskal-Wallis tests showed abundance difference for Proteobacteria ($p=0.03$), and Firmicutes ($p=0.07$) between IBD and control samples. At the family level, significant abundance differences for Bacteroidaceae ($p=0.006$), Rikenellaceae ($p=0.03$), Acidaminococcaceae ($p=0.05$), Victivallaceae ($p=0.03$), and Enterobacteriaceae ($p=0.03$) were observed between IBD and control samples (Figure 1B).

Reduction in alpha diversity estimates in the IBD samples (UC and CD groups) compared to control samples was observed (Supplementary Figure 1), however only the Shannon diversity estimate was significantly lower in the CD group compared to the control group (p=0.04). Overall microbial dysbiosis commonly reported in IBD, and greater dysbiosis in CD is captured in the study.

### 3.3. Identification and quantification of biomarker bacterial species

Fecal microbiota analyses identified not only the taxa unique for UC, CD, and control groups, but also common bacterial species among the groups (Supplementary Table 4). In fact, 29 bacterial species that were found in all three groups cumulatively constituted nearly 60% of all taxon assigned reads identified in the fecal microbiota (Figure 1C). Among the 29 bacterial species found in all three groups, *Faecalibacterium prausnitzii*, *Akkermansia muciniphila*, *Bacteroides vulgatus*, *Streptococcus thermophilus*, and *Escherichia coli* were the most common ones making up 84% of the reads assigned to these 29 bacterial species.

Shiga toxin-producing *Escherichia coli* is a proinflammatory bacteria, reported to show increasing abundance in IBD patients, whereas harmful or beneficial association of *Bacteroides vulgatus* in IBD is less certain (da Silva Santos et al., 2015; Palmela et al., 2018; Zafar and Saier, 2021). *Faecalibacterium prausnitzii*, *Streptococcus thermophilus*, and *Akkermansia muciniphila* are beneficial bacteria, reported to be reduced in IBD patients (Sokol et al., 2008; Prosberg et al., 2016; Zafar and Saier, 2021). These five species are chosen as potential biomarker bacteria for further quantification in all UC and CD samples. Because higher abundance of pathogenic *Clostridioides difficile* is also reported to be associated with IBD (Issa et al., 2007), *C. difficile* is also chosen to be further quantified in the IBD samples.

Quantitative real-time PCR analyses were conducted with primers specific to each target species (Table 1). Specificity of primers and success of the qPCR reactions were checked by melting curve analyses (Supplementary Figure 2). After quantification of six selected species in all twenty two samples by qPCR, the relative abundance of candidate species compared to total bacteria amount is calculated by the $2^{-\Delta Ct}$ method for each sample. Reduction in the relative abundance of beneficial species *F. prausnitzii* and *A. muciniphila*, in UC and CD samples compared to total bacteria amount was evident (Figure 2). The relative abundance of *B. vulgatus* was higher in the control samples (Figure 2). Interestingly, the relative abundance of *S. thermophilus* was highest in the UC samples (Figure 2). Shiga toxin-producing *E. coli* and *C. difficile* was observed in only three of the CD patients. The next analyses compared the fold change of *A. muciniphila*, *B. vulgatus*, *S. thermophilus*, and *F. prausnitzii* in IBD samples with respect to control samples calculated by the $2^{-\Delta\Delta Ct}$ method. Significant reduction in *A. muciniphila*, *S. thermophilus*, and *F. prausnitzii* in combined IBD samples compared to controls was observed (Figure 3). However, fold change values of these four species were similar in the UC and CD samples (Figure 3).

Effect of disease course and treatment on common bacterial species is also examined. Patients on biologics treatment presented with higher reduction in the relative abundance of beneficial *F. prausnitzii* compared to patients who are not on biologics (mean log reduction 5.8 vs. 2.1, p=0.04). Within the CD group, patients who had surgery showed higher reduction in the relative abundance of *A. muciniphila* compared to patients who did not have surgery (mean log reduction 21 vs. 15, p=0.02). Moreover, patients with penetrating perianal CD (more severe version of CD) again showed higher reduction in the relative abundance of *A. muciniphila* (mean log reduction 21 vs. 16, p=0.05), and increase in the relative abundance of *B. vulgatus* (mean log increase 10 vs. 5, p=0.05). Within the UC group, patients with CRP levels higher than 5 showed higher reduction in

the relative abundance of *B. vulgatus* compared to patients with CRP levels less than 5 (mean log reduction 14 vs. 4, p=0.05).

### 3.4. Utility of NGS (Next Generation Sequencing) and qPCR results

None of the correlations between the NGS read and the qPCR results of either individual selected bacteria or total bacteria was statistically significant. Highest correlation was observed for total bacteria results (Adjusted $R^2$=0.14) but it was not significant (p=0.18). Lack of congruency between the two techniques was in part due to great variation in NGS read results. The standard deviation of NGS reads ranged from a lowest of 724 for *A. muciniphila* to 4352 for *S. thermophilus*. Contrary, the standard deviation of qPCR results ranged from a lowest of 6.3 for *A. muciniphila* to 14.4 for *B. vulgatus*, showing orders of magnitude smaller variance in the qPCR results compared to NGS results. However, the lower variation in qPCR assays is also due to the concomitant characteristics of the qPCR technique and range of the data generated. Moreover, we did not incorporate UMI (Unique Molecular Identifier) analysis in NGS data that enables identification of PCR duplicates into sequencing libraries, which can introduce noise in the read counts.

The much lower variation inherent in qPCR results suggests higher accuracy, precision, and higher clinical utility for the qPCR technique compared to NGS based amplicon sequencing approach especially with smaller sample sizes. Statistical power calculations based on mean and standard deviation estimates from our experimental data showed that qPCR results have over 80% statistical power to detect a minimum 10% difference in candidate species abundance with less than 200 samples (Figure 4). Such high statistical power is nearly impossible to achieve with NGS methods. Given that IBD clinics have hundreds of patients, and microbiota alterations as a response to therapy/interventions are usually much higher than 10%, qPCR surveillance of candidate bacteria has good clinical potential to monitor microbiota response through disease and therapy course.

334

335

336

337

338 **4. Discussion**

339 We aimed to test whether qPCR analysis of candidate common bacterial species identified from
340 a patient's individual fecal microbiome can be used as a fast and reliable personalized microbial
341 biomarker for efficient monitoring of IBD. NGS based fecal microbiota analyses followed by
342 targeted qPCR analyses of candidate common bacterial species showed to be an efficient and
343 reliable method for monitoring of disease status in IBD patients.
344 Firstly, thorough microbiota analyses identified bacterial taxa in UC, CD, and controls at the
345 species level resolution. Microbiota profiles obtained in this study was similar to the reported
346 profiles in the literature agreeing with dysbiosis in CD and UC patients (Kostic et al., 2014;
347 Pascal et. al. 2017). Based on the findings of microbiota analyses, bacterial species reported to be
348 positively and negatively associated with IBD are chosen as candidate biomarker species.
349 However, rather than targeting rare species that are observed in UC or CD, we primarily focused
350 on high abundance bacterial species that are commonly observed not only in IBD but also in
351 healthy controls. *Faecalibacterium prausnitzii*, *Akkermansia muciniphila*, and *Streptococcus*
352 *thermophilus* are gut bacteria with anti-inflammatory properties suggested to be important in gut
353 homeostasis. Their abundance is reported to be reduced in IBD patients (Prosberg et al., 2016;
354 Pascal et. al. 2017; Zafar and Saier, 2021). Our qPCR analyses also showed significant reduction
355 in abundance of these beneficial bacteria in the IBD samples compared to healthy control
356 samples. Moreover, the reduction in the relative abundance of these bacteria was greater in
357 patients with worse disease progression such as in patients with penetrating CD, higher CRP
358 levels (higher inflammation), and patients who require biologics treatment. *Clostridium difficile*,
359 and Shiga toxin-producing *Escherichia coli*, on the other hand, are harmful bacteria with mucus
360 degrading, invasive, pro-inflammatory properties reported to be in high abundance in IBD

patients (Issa et al., 2007; da Silva Santos *et al*., 2015; Prosberg *et al*., 2016; Palmela *et al*., 2018). In our qPCR analyses *C. difficile* and Shiga toxin-producing *E. coli* were only observed in CD samples. These results show that qPCR results are specific to targeted species, are in agreement with literature reports, and therefore are reliable.

IBD microbiota studies are advancing from just reporting descriptive microbiota changes to examining correlations between microbiota profiles, and IBD disease activity, course, and treatment response. Recently, certain bacterial taxa (such as *Clostridiales*, *Eubacteria*, *Bifidobacteria*) are suggested to be associated with treatment response, relapse, and disease progression (Rajca *et al*., 2014; Kolho *et al*., 2015; Zhou *et al*., 2018). However, in these studies, the taxonomic resolution is course and not at the species level. With appropriate species specific primers, qPCR analysis can be highly sensitive and accurate identifying the altered species in IBD. Species level information can be used for better association tests and predictions with respect to clinical and treatment phenotypes. In addition, statistical modelling and (clinical) interpretation of multivariate microbiota (microbiome) data with respect to a (clinical) phenotype is much harder compared to univariate species specific statistical association analysis, limiting the clinical usefulness of multivariate microbiota data.

Although NGS approach is proposed to identify rare taxa that can be unique to UC or CD, the clinical utility of microbiota data generated by NGS based 16S rRNA gene amplicon sequencing is still debated. Firstly, NGS based methods are more expensive, time consuming, and require bioinformatics infrastructure and expertise. In addition, methodological issues (due to PCR artifacts, sequencing platform, DNA isolation and contamination, etc.), huge variation in the number of sequence reads, different microbiota results generated even analyzing the same sample (Hiergeist *et al*., 2016; Boers *et al*., 2019) hinder usage of NGS based microbiota results

384 in monitoring disease status and course in IBD patients. In this study, the variance associated
385 with 16S rRNA gene NGS reads was also much higher compared to the relative abundance
386 variances estimated from qPCR data, making 16S rRNA gene NGS data more noisy for
387 statistical comparisons.
388 Some bacterial species and strains have multiple 16S rRNA gene copies in their genomes making
389 the16S rRNA gene NGS based estimates of relative abundance and representation of these taxa
390 in the microbiome erroneous (Vetrovsky and Baldrian, 2013; Louca *et al*., 2018). Because all
391 taxon identification databases are based on 16S rRNA gene sequence, NGS based methods do
392 not have the alternative of targeting other genome regions. A possible solution is adapting a
393 metagenomics approach, and sequencing whole genomes. However, metagenomics is even
394 harder, more problematic, time consuming, and much more expensive than 16S rRNA gene
395 amplicon sequencing. On the other hand, in a qPCR approach one can easily target genome
396 regions other than the 16S rRNA gene, and alleviate possible distorted relative abundance
397 estimates due to multiple 16S rRNA gene copies.
398 There are several limitations of the study. The sample size is small, and longitudinal sampling of
399 microbiota is not available. Although the sample size is small, patients with diverse disease
400 location and activity phenotype characteristics are involved in the study. So, the results of the
401 qPCR approach are not just specific to a subgroups of UC or CD patients, but can be
402 generalizable to broader IBD patients. There are several other pathological *E. coli* associated
403 with IBD, however we only focused on Shiga toxin-producing *E. coli*, quantification of *E. coli* in
404 general could be more informative. Moreover, IBD has a genetic component, and genetic
405 variants in *NOD2*, *ATG16L1*, and *IL23R* are reported to be associated with highest IBD risk
406 (McGovern *et al*., 2015). We tested whether genetics can be a confounder of the results, but IBD

genetic risk profiles of UC and CD groups were similar. There was no single dominant IBD risk genotype in the UC and CD groups. So a genetic stratification confounding the results is unlikely. We acknowledge that there can be other genetic factors that can influence the course of IBD and treatment response. These additional genetic factors can be considered in the future studies, if deemed necessary by the medical community.

## 5. Conclusions

In conclusion, qPCR analysis of common candidate bacterial species identified from a patient's individual fecal microbiome can be used as a fast and reliable personalized microbial biomarker for efficient monitoring of IBD. Moreover, the relative abundance of these common bacterial species showed association with worse disease progression in IBD.

Our results should stimulate further studies adopting personalized microbiota based qPCR analysis of targeted bacterial species longitudinally sampled from larger sized cohorts of IBD patients.

## References


Alatab S, Sepanlou SG, Ikuta K, Vahedi H, Bisignano C et al. (2020). The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990-2013;2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet Gastroenterology & Hepatology 5 (1):17-30. doi:10.1016/S2468-1253(19)30333-4

Albenberg LG, Lewis JD, Wu GD (2012). Food and the gut microbiota in inflammatory bowel diseases: a critical connection. Curr Opin Gastroenterol 28 (4):314-320. doi:10.1097/MOG.0b013e328354586f

Balamurugan R, Balaji V, Ramakrishna BS (2008). Estimation of faecal carriage of Clostridium difficile in patients with ulcerative colitis using real time polymerase chain reaction. Indian Journal of Medical Research 127 (5):472-477

Bemelman WA, Warusavitarne J, Sampietro GM, Serclova Z, Zmora O et al. (2018). ECCO-ESCP Consensus on Surgery for Crohn's Disease. J Crohns Colitis 12 (1):1-16. doi:10.1093/ecco-jcc/jjx061

Bjerrum JT, Steenholdt C, Ainsworth M, Nielsen OH, Reed MA et al. (2017). Metabonomics uncovers a reversible proatherogenic lipid profile during infliximab therapy of inflammatory bowel disease. BMC Medicine 15 (1):184. doi:10.1186/s12916-017-0949-7

Boers SA, Jansen R, Hays JP (2019). Understanding and overcoming the pitfalls and biases of next-generation sequencing (NGS) methods for use in the routine clinical microbiological diagnostic laboratory. European Journal of Clinical Microbiology & Infectious Diseases 38 (6):1059-1070. doi:10.1007/s10096-019-03520-3

Burisch J, Munkholm P (2015). The epidemiology of inflammatory bowel disease. Scandinavian Journal of Gastroenterology 50 (8):942-951. doi:10.3109/00365521.2015.1014407

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., et al. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem,* 55**,** 611-22.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ et al. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13 (7):581-583. doi:10.1038/nmeth.3869

Can G, Posul E, Yilmaz B, Can H, Korkmaz U et al. (2019). Epidemiologic features of inflammatory bowel disease in Western Blacksea region of Turkey for the last 10 years: retrospective cohort study. Korean Journal of Internal Medicine 34 (3):519-529. doi:10.3904/kjim.2015.310

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7 (5):335-336. doi:10.1038/nmeth.f.303

Chen P, Zhou G, Lin J, Li L, Zeng Z et al. (2020). Serum Biomarkers for Inflammatory Bowel Disease. Front Med (Lausanne) 7:123. doi:10.3389/fmed.2020.00123

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences,* Hillsdale, N.J., L. Erlbaum Associates.

da Silva Santos AC, Gomes Romeiro F, Yukie Sassaki L, Rodrigues J (2015). Escherichia coli from Crohn's disease patient displays virulence features of enteroinvasive (EIEC),

enterohemorragic (EHEC), and enteroaggregative (EAEC) pathotypes. Gut Pathogens 7 (1):2. doi:10.1186/s13099-015-0050-8

Falentin H, Henaff N, Le Bivic P, Deutsch SM, Parayre S et al. (2012). Reverse transcription quantitative PCR revealed persistency of thermophilic lactic acid bacteria metabolic activity until the end of the ripening of Emmental cheese. Food Microbiology 29 (1):132-140. doi:10.1016/j.fm.2011.09.009

Fernandez S, Fraga M, Castells M, Colina R, Zunino P (2020). Effect of the administration of Lactobacillus spp. strains on neonatal diarrhoea, immune parameters and pathogen abundance in pre-weaned calves. Benef Microbes 11 (5):477-488. doi:10.3920/BM2019.0167

Fujimoto T, Imaeda H, Takahashi K, Kasumi E, Bamba S et al. (2013). Decreased abundance of Faecalibacterium prausnitzii in the gut microbiota of Crohn's disease. Journal of Gastroenterology and Hepatology 28 (4):613-619. doi:10.1111/jgh.12073

Guo X, Huang C, Xu J, Xu H, Liu L et al. (2021). Gut Microbiota Is a Potential Biomarker in Inflammatory Bowel Disease. Front Nutr 8:818902. doi:10.3389/fnut.2021.818902

Hiergeist A, Reischl U, Priority Program Intestinal Microbiota Consortium/ quality assessment p, Gessner A (2016). Multicenter quality assessment of 16S ribosomal DNA-sequencing for microbiome analyses reveals high inter-center variability. International Journal of Medical Microbiology 306 (5):334-342. doi:10.1016/j.ijmm.2016.03.005

Ishaq HM, Mohammad IS, Sher Muhammad K, Li H, Abbas RZ et al. (2021). Gut microbial dysbiosis and its association with esophageal cancer. J Appl Biomed 19 (1):1-13. doi:10.32725/jab.2021.005

Issa M, Vijayapal A, Graham MB, Beaulieu DB, Otterson MF et al. (2007). Impact of Clostridium difficile on inflammatory bowel disease. Clinical Gastroenterology and Hepatology 5 (3):345-351. doi:10.1016/j.cgh.2006.12.028

Kalla R, Adams AT, Bergemalm D, Vatn S, Kennedy NA et al. (2021). Serum proteomic profiling at diagnosis predicts clinical course, and need for intensification of treatment in inflammatory bowel disease. J Crohns Colitis 15 (5):699-708. doi:10.1093/ecco-jcc/jjaa230

Keshteli AH, Tso R, Dieleman LA, Park H, Kroeker KI et al. (2018). A Distinctive Urinary Metabolomic Fingerprint Is Linked With Endoscopic Postoperative Disease Recurrence in Crohn's Disease Patients. Inflammatory Bowel Diseases 24 (4):861-870. doi:10.1093/ibd/izx070

Kolho KL, Korpela K, Jaakkola T, Pichai MV, Zoetendal EG et al. (2015). Fecal Microbiota in Pediatric Inflammatory Bowel Disease and Its Relation to Inflammation. American Journal of Gastroenterology 110 (6):921-930. doi:10.1038/ajg.2015.149

Kostic AD, Xavier RJ, Gevers D (2014). The microbiome in inflammatory bowel disease: current status and the future ahead. Gastroenterology 146 (6):1489-1499. doi:10.1053/j.gastro.2014.02.009

Livak KJ, Schmittgen TD (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25 (4):402-408. doi:10.1006/meth.2001.1262

Lopez-Siles, M., Aldeguer, X., Sabat-Mir, M., Serra-Pages, M., Duncan, S. H., Flint, H. J., et al. (2020). Evaluation of bacterial biomarkers to aid in challenging inflammatory bowel diseases diagnostics and subtype classification. *World J Gastrointest Pathophysiol,* 11**,** 64-77.

Lopez-Siles, M., Enrich-Capo, N., Aldeguer, X., Sabat-Mir, M., Duncan, S. H., Garcia-Gil, L. J., et al. (2018). Alterations in the Abundance and Co-occurrence of Akkermansia muciniphila and Faecalibacterium prausnitzii in the Colonic Mucosa of Inflammatory Bowel Disease Subjects. *Front Cell Infect Microbiol,* 8**,** 281.

Lopez-Siles, M., Martinez-Medina, M., Busquets, D., Sabat-Mir, M., Duncan, S. H., Flint, H. J., et al. (2014). Mucosa-associated Faecalibacterium prausnitzii and Escherichia coli co-abundance can distinguish Irritable Bowel Syndrome and Inflammatory Bowel Disease phenotypes. *Int J Med Microbiol,* 304**,** 464-75.

Louca S, Doebeli M, Parfrey LW (2018). Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. Microbiome 6 (1):41. doi:10.1186/s40168-018-0420-9

Machiels, K., Joossens, M., Sabino, J., De Preter, V., Arijs, I., Eeckhaut, V., et al. (2014). A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. *Gut,* 63**,** 1275-83.

McGovern DP, Kugathasan S, Cho JH (2015). Genetics of Inflammatory Bowel Diseases. Gastroenterology 149 (5):1163-1176 e1162. doi:10.1053/j.gastro.2015.08.001

McMurdie PJ, Holmes S (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PloS One 8 (4):e61217. doi:10.1371/journal.pone.0061217

Mehta F (2016). Report: economic implications of inflammatory bowel disease and its management. American Journal of Managed Care 22 (3 Suppl):s51-60

Miranda-Garcia P, Chaparro M, Gisbert JP (2016). Correlation between serological biomarkers and endoscopic activity in patients with inflammatory bowel disease. Gastroenterología y Hepatologia 39 (8):508-515. doi:10.1016/j.gastrohep.2016.01.015

Mondot, S., Lepage, P., Seksik, P., Allez, M., Treton, X., Bouhnik, Y., et al. (2016). Structural robustness of the gut mucosal microbiota is associated with Crohn's disease remission after surgery. *Gut,* 65**,** 954-62.

Montero-Melendez T, Llor X, Garcia-Planella E, Perretti M, Suarez A (2013). Identification of novel predictor classifiers for inflammatory bowel disease by gene expression profiling. PloS One 8 (10):e76235. doi:10.1371/journal.pone.0076235

Navidshad B, Liang JB, Jahromi MF (2012). Correlation coefficients between different methods of expressing bacterial quantification using real time PCR. International Journal of Molecular Sciences 13 (2):2119-2132. doi:10.3390/ijms13022119

Notararigo S, Martin-Pastor M, Vinuela-Roldan JE, Quiroga A, Dominguez-Munoz JE et al. (2021). Targeted (1)H NMR metabolomics and immunological phenotyping of human fresh blood and serum samples discriminate between healthy individuals and inflammatory bowel disease patients treated with anti-TNF. Journal of Molecular Medicine (Berlin, Germany) 99 (9):1251-1264. doi:10.1007/s00109-021-02094-y

Okonechnikov K, Golosova O, Fursov M, team U (2012). Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics 28 (8):1166-1167. doi:10.1093/bioinformatics/bts091

Osman MA, Neoh HM, Ab Mutalib NS, Chin SF, Mazlan L et al. (2021). Parvimonas micra, Peptostreptococcus stomatis, Fusobacterium nucleatum and Akkermansia muciniphila as a four-bacteria biomarker panel of colorectal cancer. Scientific Reports 11 (1):2925. doi:10.1038/s41598-021-82465-0

Owczarzy R, Tataurov AV, Wu Y, Manthey JA, McQuisten KA et al. (2008). IDT SciTools: a suite for analysis and design of nucleic acid oligomers. Nucleic Acids Research 36 (Web Server issue):W163-169. doi:10.1093/nar/gkn198

Palmela C, Chevarin C, Xu Z, Torres J, Sevrin G et al. (2018). Adherent-invasive Escherichia coli in inflammatory bowel disease. Gut 67 (3):574-587. doi:10.1136/gutjnl-2017-314903

Pascal, V., Pozuelo, M., Borruel, N., Casellas, F., Campos, D., Santiago, A., et al. (2017). A microbial signature for Crohn's disease. *Gut,* 66**,** 813-822.

Pereira-Marques, J., Hout, A., Ferreira, R. M., Weber, M., Pinto-Ribeiro, I., Van Doorn, L. J., et al. (2019). Impact of Host DNA and Sequencing Depth on the Taxonomic Resolution of Whole Metagenome Sequencing for Microbiome Analysis. *Front Microbiol,* 10**,** 1277.

Podolsky DK (2002). Inflammatory bowel disease. New England Journal of Medicine 347 (6):417-429. doi:10.1056/NEJMra020831

Prosberg M, Bendtsen F, Vind I, Petersen AM, Gluud LL (2016). The association between the gut microbiota and the inflammatory bowel disease activity: a systematic review and meta-analysis. Scandinavian Journal of Gastroenterology 51 (12):1407-1415. doi:10.1080/00365521.2016.1216587

Rajca S, Grondin V, Louis E, Vernier-Massouille G, Grimaud JC et al. (2014). Alterations in the intestinal microbiome (dysbiosis) as a predictor of relapse after infliximab withdrawal in Crohn's disease. Inflammatory Bowel Diseases 20 (6):978-986. doi:10.1097/MIB.0000000000000036

Serrano-Gomez G, Mayorga L, Oyarzun I, Roca J, Borruel N et al. (2021). Dysbiosis and relapse-related microbiome in inflammatory bowel disease: A shotgun metagenomic approach. Comput Struct Biotechnol J 19:6481-6489. doi:10.1016/j.csbj.2021.11.037

Simenc J, Potocnik U (2011). Rapid differentiation of bacterial species by high resolution melting curve analysis. Prikladnaia Biokhimiia i Mikrobiologiia 47 (3):283-290

Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermudez-Humaran LG et al. (2008). Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. Proceedings of the National Academy of Sciences of the United States of America 105 (43):16731-16736. doi:10.1073/pnas.0804812105

Soubieres AA, Poullis A (2016). Emerging role of novel biomarkers in the diagnosis of inflammatory bowel disease. World Journal of Gastrointestinal Pharmacology and Therapeutics 7 (1):41-50. doi:10.4292/wjgpt.v7.i1.41

Stange EF, Travis SP, Vermeire S, Reinisch W, Geboes K et al. (2008). European evidence-based Consensus on the diagnosis and management of ulcerative colitis: Definitions and diagnosis. J Crohns Colitis 2 (1):1-23. doi:10.1016/j.crohns.2007.11.001

Tontini GE, Vecchi M, Pastorelli L, Neurath MF, Neumann H (2015). Differential diagnosis in inflammatory bowel disease colitis: state of the art and future perspectives. World Journal of Gastroenterology 21 (1):21-46. doi:10.3748/wjg.v21.i1.21

Vetrovsky T, Baldrian P (2013). The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. PloS One 8 (2):e57923. doi:10.1371/journal.pone.0057923

Walter E, Hausberger SC, Gross E, Siebert U (2020). Health-related quality of life, work productivity and costs related to patients with inflammatory bowel disease in Austria. Journal of Medical Economics 23 (10):1061-1071. doi:10.1080/13696998.2020.1801187

612 Zafar H, Saier MH, Jr. (2021). Gut Bacteroides species in health and disease. Gut Microbes 13
613 (1):1-20. doi:10.1080/19490976.2020.1848158
614 Zaheer, R., Noyes, N., Ortega Polo, R., Cook, S. R., Marinier, E., Van Domselaar, G., et al.
615 (2018). Impact of sequencing depth on the characterization of the microbiome and
616 resistome. *Sci Rep,* 8**,** 5890

617 Zhou, Y., Chen, H., He, H., Du, Y., Hu, J., Li, Y., et al. (2016). Increased Enterococcus faecalis
618 infection is associated with clinically active Crohn disease. *Medicine (Baltimore),* 95**,**
619 e5019.

620 Zhou Y, Xu ZZ, He Y, Yang Y, Liu L et al. (2018). Gut Microbiota Offers Universal Biomarkers
621 across Ethnicity in Inflammatory Bowel Disease Diagnosis and Infliximab Response
622 Prediction. mSystems 3 (1). doi:10.1128/mSystems.00188-17

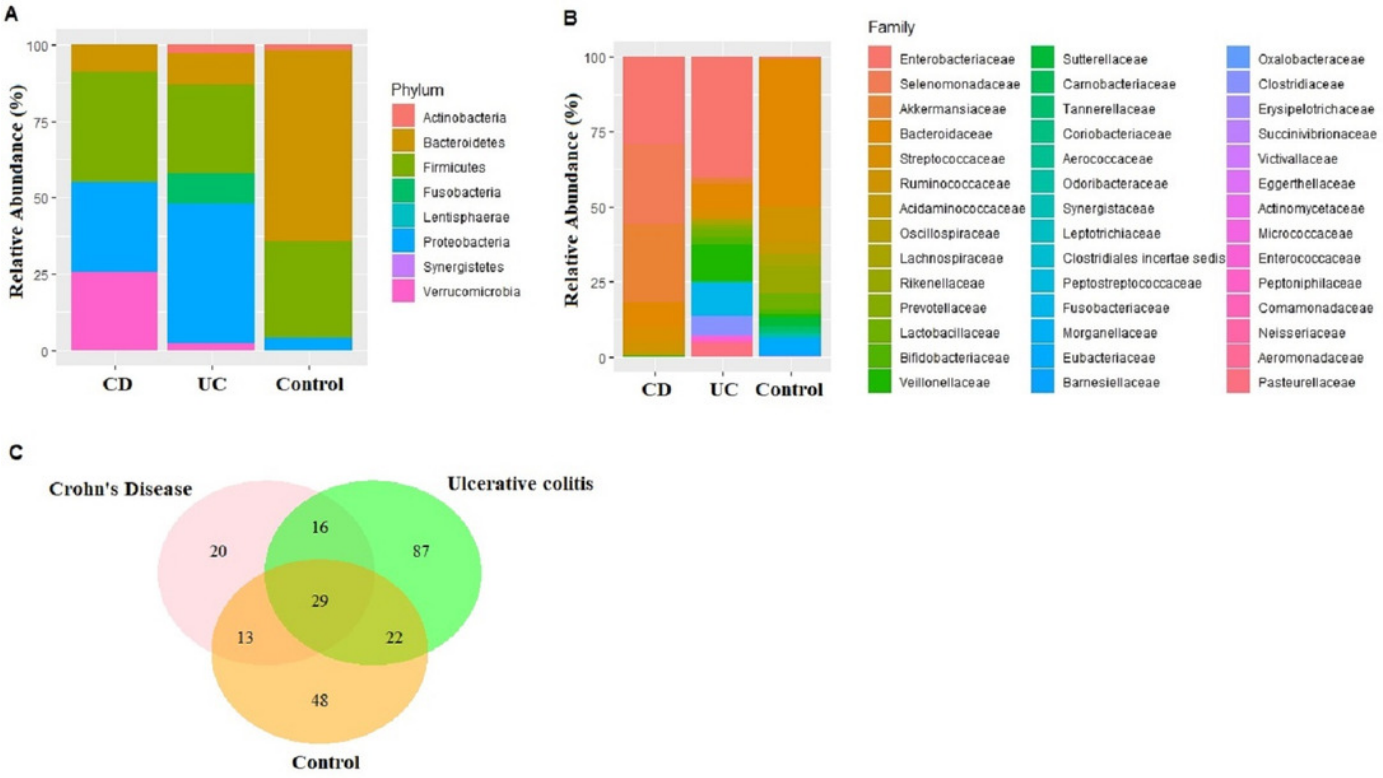
623

624

# Figure 1

Microbiota composition comparisons

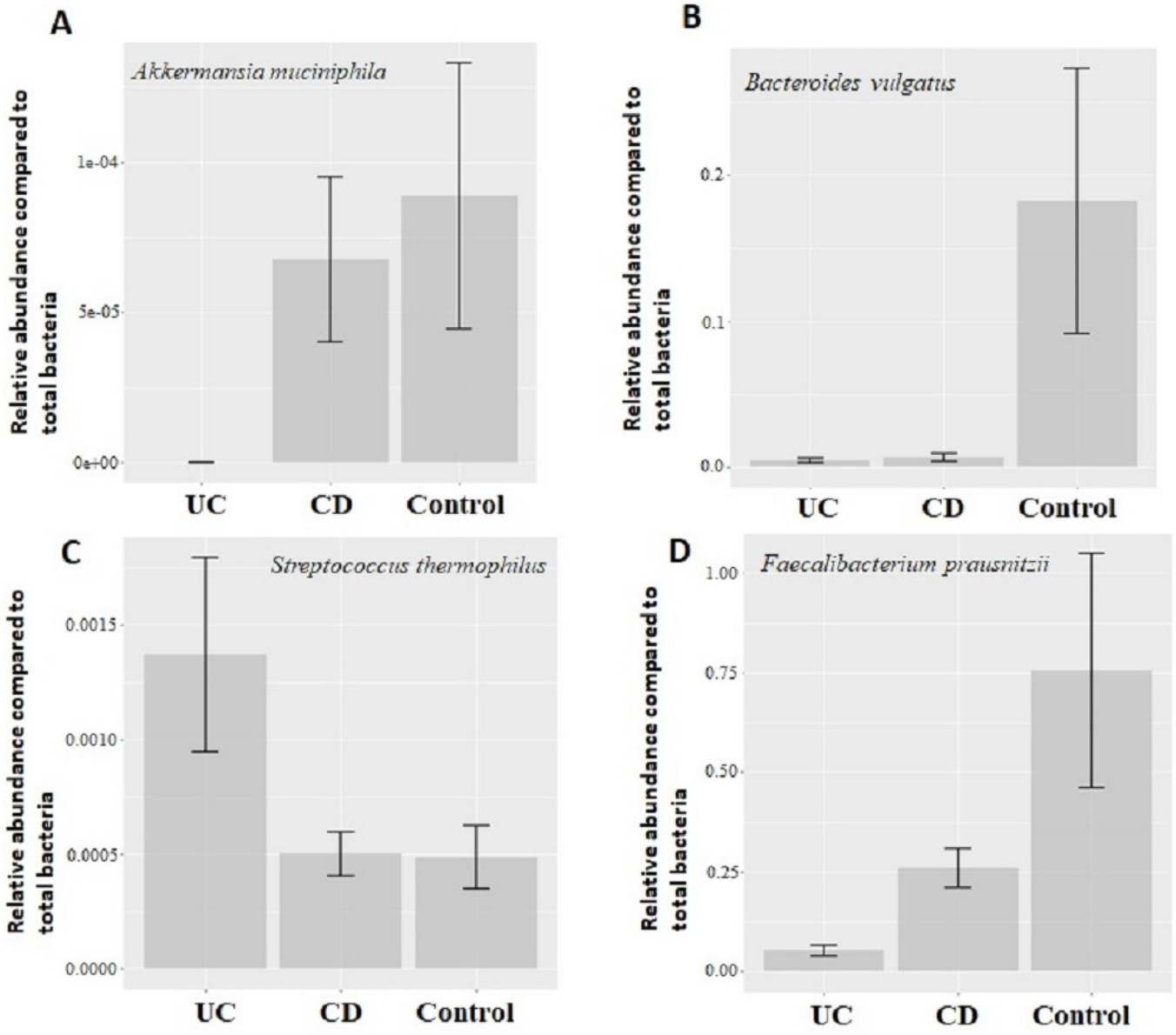
Phylum level **(A),** and family level **(B)** taxon relative abundance comparisons among Crohn (CD), Ulcerative colitis (UC), and control groups. **(C)** Venn diagram illustrating the number of common shared and unique bacterial species observed in Crohn disease, Ulcerative colitis, and control groups.



# Figure 2

Comparison of relative abundance of selected common bacteria

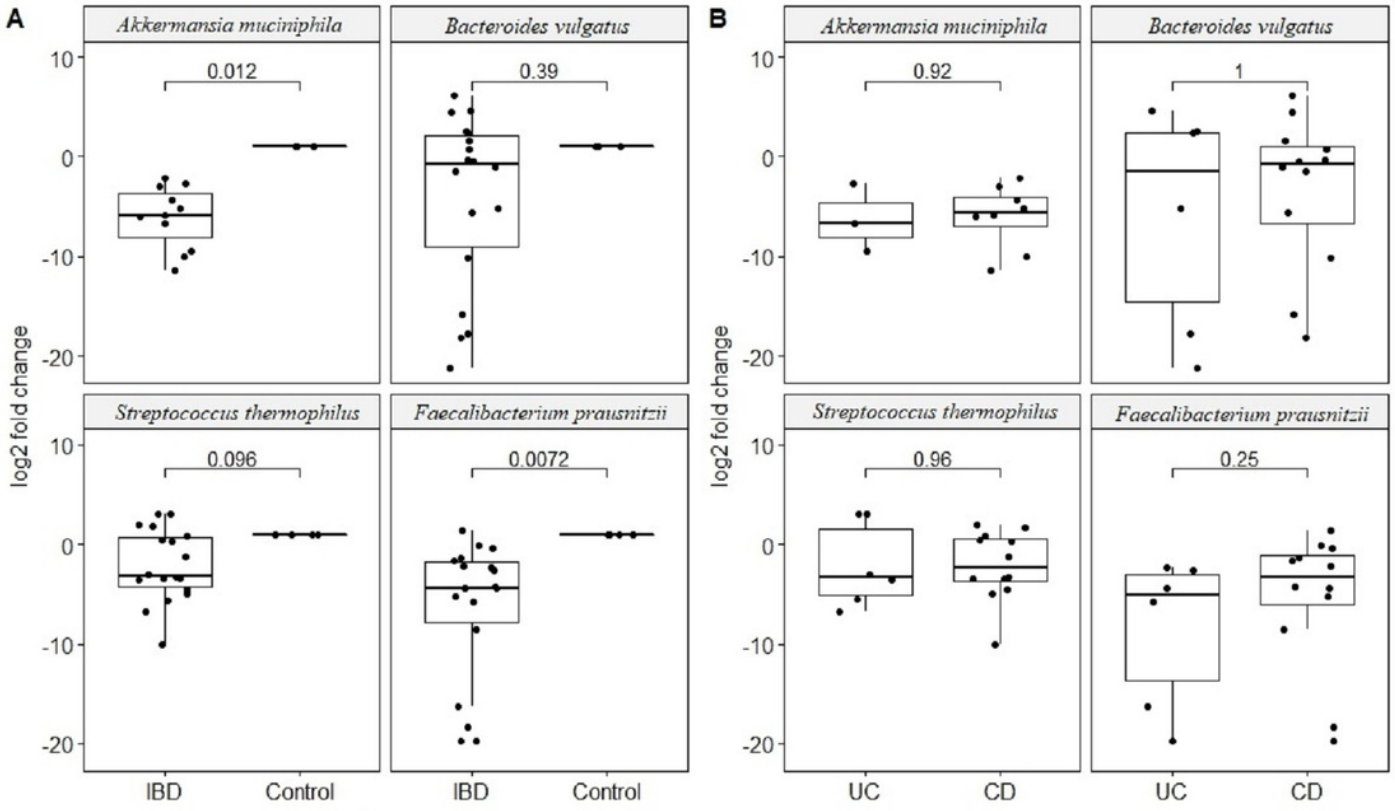
Relative abundance of (A) *Akkermansia muciniphila* , (B) *Bacteroides vulgatus* , (C) *Streptococcus thermophiles*, (D) *Faecalibacterium prausnitzii* with respect to total bacteria quantified by qPCR within each group. Y-axis values are the ratio of target bacteria abundance to total bacteria abundance. UC: Ulcerative colitis, CD: Crohn.



# Figure 3

Comparison of relative abundance change of *Akkermansia muciniphila*, *Bacteroides vulgatus*, *Streptococcus thermophiles*, and *Faecalibacterium prausnitzii*

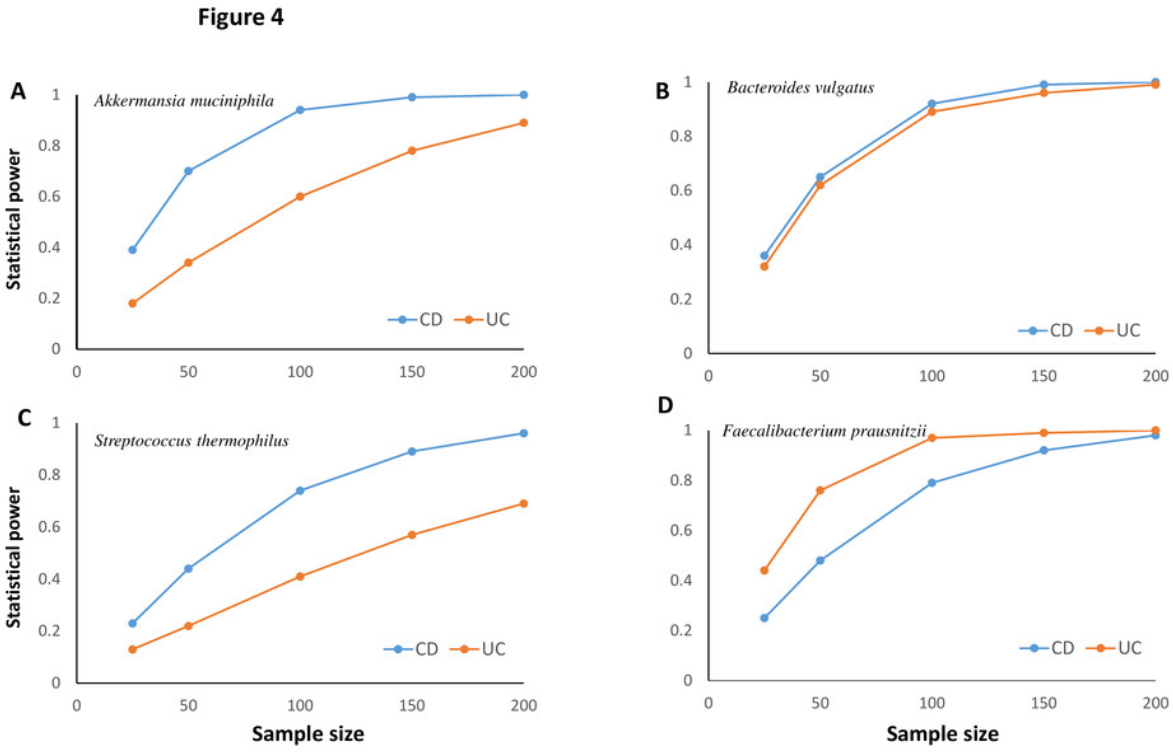
Comparison of log2 fold change of relative abundance of *Akkermansia muciniphila*, *Bacteroides vulgatus*, *Streptococcus thermophiles*, and *Faecalibacterium prausnitzii* with respect to control samples in (A) IBD (UC and CD groups together), (B) in Ulcerative colitis (UC) and Crohn (CD) samples. P values calculated from non-parametric Wilcoxon rank-sum tests.



# Figure 4

Relationship between sample size and statistical power

Relationship between sample size and statistical power to detect a 10% change with 95% confidence in relative abundance of (A) *Akkermansia muciniphila* , (B) *Bacteroides vulgatus* , (C) *Streptococcus thermophiles*, (D) *Faecalibacterium prausnitzii* in Ulcerative colitis (UC) and Crohn (CD) samples. Power calculations are based on the mean and standard deviation estimates from experimental results for each bacterial species. For *B. vulgatus* power to detect 20% change numbers are presented.



## Table 1(on next page)

Primers used for detection and quantification of selected bacterial species

**Table 1.** Primers used for detection and quantification of selected bacterial species

| Target Bacteria | Genome region | DNA sequences of the primers (5'-3') | | Product size (bp) | References |
|---|---|---|---|---|---|
| *F. prausnitzii* | 16S rRNA | Forward | GGAGGAAGAAGGTCTTCGG | 248 | Fujimoto et al., 2013 |
| | | Reverse | AATTCCGCCTACCTCTGCACT | | |
| *C. difficile* | 16S rRNA | Forward | TTGAGCGATTTACTTCGGTAAAGA | 157 | Balamurugan et al., 2008 |
| | | Reverse | CCATCCTGTACTGGCTCACCT | | |
| *A. muciniphila* | Genomic location | Forward | GAAGACGGAGGACGGAACT | 126 | Osman et al., 2021 |
| | | Reverse | GCGGATTGCTGACGAAGG | | |
| *B. vulgatus* | 16S rRNA | Forward | GCATCATGAGTCCGCATGTTC | 287 | Ishaq et al., 2021 |
| | | Reverse | TCCATACCCGACTTTATTCCTT | | |
| *S. thermophilus* | *groL* | Forward | GCTGTGGAAGAGCTTAAAGTC | 138 | Falentin et al., 2012 |
| | | Reverse | ACCATCATTACCAACGCGT | | |
| Shiga toxin-producing *E. coli* | *stx1* | Forward | GTCACAGTAACAAACCGTAACA | 95 | Fernández et al., 2020 |
| | | Reverse | TCGTTGACTACTTCTTATCTGGA | | |
| Total bacteria | 16S rRNA | Forward | TCCTACGGGAGGCAGCAGT | 466 | Balamurugan et al., 2008 |
| | | Reverse | GGACTACCAGGGTATCTAATCCTGTT | | |

## Table 2(on next page)

Patient and disease characteristics

**Table 2.** Control, patient group, and disease characteristics

| | Crohn's disease (%) N=12 | Ulcerative colitis (%) N=6 | Control (%) N=4 |
|---|---|---|---|
| **Median Age, years (25%,75%)** | 51 (48,56) | 47 (41,50) | 24 (22,24) |
| **Sex (%)** | | | |
| Female | 6 (50) | 4 (67) | 3 (75) |
| Male | 6 (50) | 2 (33) | 1 (25) |
| | | | - |
| **Median disease duration, years (25%,75%)** | 7 (2,9) | 6 (1,14) | - |
| **Smoking History** | | | |
| Yes (%) | 2 (17) | 2 (33) | 1 (25) |
| **Disease localization (%)** | | | |
| Ileal | 1 (8) | - | - |
| Colonic | 4 (33) | - | - |
| Ileocolonic | 3 (25) | - | - |
| Surgery | 3 (25) | - | - |
| Penetrating perianal disease | 3 (25) | - | - |
| Distal colitis | - | 1 (17) | - |
| Left colitis | - | 2 (33) | - |
| Pancolitis | - | 3 (50) | - |
| **Treatment (%)** | | | - |
| Biologics | 9 (75) | 4 (67) | - |
| Non-biologics | 3 (25) | 2 (33) | - |
| **Median Mayo Score (25%,75%)** | - | 5 (4,6) | - |
| **Harvey-Bradshaw Index (25%,75%)** | 6.5 (3,7.5) | - | - |
| **C-reactive protein (CRP mg/dL)** | | | |
| <5 (%) | 6 (50) | 4 (67) | - |
| >5 (%) | 6 (50) | 2 (33) | - |

Note: CRP level at fecal sample collection time indicating severity of disease activity