Tissue-Associated Bacterial Alterations in Rectal Carcinoma Patients Revealed by 16S rRNA Community Profiling

Andrew M. Thomas [1,2,3], Eliane C. Jesus [1,4], Ademar Lopes [4], Samuel Aguiar Jr. [4], Maria D. Begnami [5], Rafael M. Rocha [6], Paola Avelar Carpinetti [7], Anamaria A. Camargo [7], Christian Hoffmann [8], Helano C. Freitas [1,9], Israel T. Silva [10], Diana N. Nunes [1], João C. Setubal [2,11] and Emmanuel Dias-Neto [1,12]*

[1] Medical Genomics Laboratory, CIPE/A.C. Camargo Cancer Center, São Paulo, Brazil, [2] Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil, [3] Curso de Pós-Graduação em Bioinformática, Universidade de São Paulo, São Paulo, Brazil, [4] Department of Pelvic Surgery, A.C. Camargo Cancer Center, São Paulo, Brazil, [5] Department of Pathology, A.C. Camargo Cancer Center, São Paulo, Brazil, [6] Laboratory of Molecular Gynecology, Department of Gynecology, Medicine College, Federal University of São Paulo, São Paulo, Brazil, [7] Centro de Oncologia Molecular, Hospital Sirio-Libanês, São Paulo, Brazil, [8] Departamento de Alimentos e Nutrição Experimental, Faculdade de Ciências Farmacêuticas, Food Research Center (FoRC), Universidade de São Paulo, São Paulo, Brazil, [9] Department of Clinical Oncology, A.C. Camargo Cancer Center, São Paulo, Brazil, [10] Laboratory of Computational Biology and Bioinformatics, A.C. Camargo Cancer Center, São Paulo, Brazil, [11] Biocomplexity Institute, Virginia Tech, Blacksburg, VA, USA, [12] Laboratory of Neurosciences (LIM-27) Alzira Denise Hertzog Silva, Institute of Psychiatry, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil

Sporadic and inflammatory forms of colorectal cancer (CRC) account for more than 80% of cases. Recent publications have shown mechanistic evidence for the involvement of gut bacteria in the development of both CRC-forms. Whereas, colon and rectal cancer have been routinely studied together as CRC, increasing evidence show these to be distinct diseases. Also, the common use of fecal samples to study microbial communities may reflect disease state but possibly not the tumor microenvironment. We performed this study to evaluate differences in bacterial communities found in tissue samples of 18 rectal-cancer subjects when compared to 18 non-cancer controls. Samples were collected during exploratory colonoscopy (non-cancer group) or during surgery for tumor excision (rectal-cancer group). High throughput 16S rRNA amplicon sequencing of the V4–V5 region was conducted on the Ion PGM platform, reads were filtered using *Qiime* and clustered using *UPARSE*. We observed significant increases in species richness and diversity in rectal cancer samples, evidenced by the total number of OTUs and the Shannon and Simpson indexes. Enterotyping analysis divided our cohort into two groups, with the majority of rectal cancer samples clustering into one enterotype, characterized by a greater abundance of *Bacteroides* and *Dorea*. At the phylum level, rectal-cancer samples had increased abundance of candidate phylum *OD1* (also known as *Parcubacteria*) whilst non-cancer samples had increased abundance of *Planctomycetes*. At the genera level, rectal-cancer samples had higher abundances of *Bacteroides*, *Phascolarctobacterium*, *Parabacteroides*, *Desulfovibrio,* and *Odoribacter* whereas non-cancer samples had higher abundances of *Pseudomonas*, *Escherichia*, *Acinetobacter*, *Lactobacillus,* and *Bacillus*. Two *Bacteroides fragilis* OTUs were more

abundant among rectal-cancer patients seen through 16S rRNA amplicon sequencing, whose presence was confirmed by immunohistochemistry and enrichment verified by digital droplet PCR. Our findings point to increased bacterial richness and diversity in rectal cancer, along with several differences in microbial community composition. Our work is the first to present evidence for a possible role of bacteria such as *B. fragilis* and the phylum *Parcubacteria* in rectal cancer, emphasizing the need to study tissue-associated bacteria and specific regions of the gastrointestinal tract in order to better understand the possible links between the microbiota and rectal cancer.

INTRODUCTION

The gut microbiota is a vast and diverse ensemble of bacteria and other microorganisms that work together to help digestion, produce vitamins, fatty acids, amino acids and other bioactive compounds, and participate in the regulation of our immune, metabolic, and neurological systems (Shapiro et al., 2014; Boulangé et al., 2016). The understanding of our microbiota, together with the determination of its composition when contrasting healthy vs. diseased states allows the identification of microorganism disturbances that are possibly related to disease development and, therefore, offers a new approach for diagnosis as well as preventive and therapeutic interventions.

Specific dietary components, tobacco and alcohol consumption, which have been linked to the development of a number of pathological states (such as obesity, allergy, diabetes, Crohn's disease, irritable colon syndrome, and cancer) are known to drive microbiome alterations and lead to dysbiosis (Turnbaugh et al., 2009; Leclercq et al., 2014; Allais et al., 2016). The direct action of these elements or of the dysbiosis they cause, appears to be instrumental in the pathogenesis of many diseases and, under certain circumstances, it is possible that dysbiosis may, *per se,* have a direct link with disease development (Duboc et al., 2013). In oncology, studies have been conducted in different neoplastic conditions, identifying roles for specific bacteria in carcinogenesis (Kostic et al., 2012; Riley et al., 2013; Rubinstein et al., 2013), immune evasion (Gur et al., 2015), modulation of the tumor microenvironment (Kostic et al., 2013), and interference with anti-cancer immune responses and immune-surveillance that facilitate chemotherapy activity (Zitvogel et al., 2013; Galluzzi et al., 2015; Vétizou et al., 2015). As a consequence, the emerging concept that cancer needs to be studied considering the complex tumor microenvironment, which includes components such as tumor cells, the surrounding microenviroment and the microbiome, may aid in the development and improvement of cancer treatment, including immunotherapy (Pitt et al., 2016).

Tumors of the lower digestive tract, which include colon and rectal cancer, are among the most prevalent neoplasias worldwide, as well as one of the most fatal. Colorectal cancer (CRC) is the third most commonly diagnosed cancer with 1.4 million people diagnosed annually (Torre et al., 2015). The World Health Organization estimates an increase of 77% in the number of newly diagnosed CRC cases and an increase of 80% in deaths from CRC by 2030 (Binefa et al., 2014). Whereas, colon and rectal cancers have been routinely studied together as CRC, evidences indicate these to be distinct nosological entities. Differences in embryological origin, anatomy, treatment, metastatic potential, and outcome between colon cancer and rectal cancer have led to discussions as to whether neoplastic lesions of these two anatomical sites should be considered as different diseases, with further dichotomization of colon cancers into distal and proximal (Tamas et al., 2015).

The mechanisms involved in sporadic CRC predisposition or development are still poorly understood and the long list of cancer risk factors is continuously expanding and includes age, tobacco, and alcohol consumption, lack of physical activity, increased body weight and, most importantly, diet (Moore and Moore, 1995; Bingham, 2000). Of particular importance is the fact that all these risk factors can directly or indirectly modify the microbiota, making the precise definition of their roles a very complex task. Fecal microbiota studies have contributed greatly in our understanding of the general gut microbiota composition and its dysbiosis in different scenarios (Wu et al., 2013; Sabino et al., 2016). However, possibly due to practical issues related to obtaining the required biopsy samples—from patients and controls—there are still very few available studies focused on the analysis of microbial community compositions of more specific regions of the lower digestive tract, such as the proximal and distal colon, and the rectal tissue. Furthermore, few studies contemplate the fact that fecal- and tissue-associated microbiota are significantly different (Durbán et al., 2011; Hong et al., 2011; Mira-Pascual et al., 2015; Flemer et al., 2016). This fact is particularly relevant for CRC as the intimate crosstalk between the host's epithelium layer and the gut microbial community is a key factor for cell proliferation and development, as well as the regulation of inflammation, a major driver of rectal carcinogenesis (Arthur et al., 2014). Such differences lead to a lack of representativeness with respect to the bacterial biofilm of the rectal mucosa (Durbán et al., 2011; Gevers et al., 2014) and may reflect the disease state but possibly not the tumor microenvironment, which is of great importance to study possible microbiota:disease links.

Here we have addressed such shortcomings by studying the tissue-associated microbiota of 36 subjects, 18 with and 18 without rectal adenocarcinoma. The use of 16S rDNA deep

sequencing allowed us to compare non-cancer x cancer mucosa, pointing to specific OTUs and bacterial genera potentially associated with rectal adenocarcinoma.

MATERIALS AND METHODS

Cohort

A total of 36 subjects were included after approval by AC Camargo Cancer Center's review board (ACCCC - 1614/11, January 30th, 2012). Tissue biopsies were collected from subjects belonging to one of the following groups:

Non-cancer Subjects

(Non-Cancer, NC, $n = 18$): All subjects had medical indication of exploratory colonoscopy due to complaints, such as bleeding, abdominal pain, constipation, and chronic diarrhea. No subjects had personal or familial history of colorectal cancer or colitis (either ulcerative, Crohn's, radiation or infectious colitis, chronic inflammatory illnesses), previous colonic or small bowel resection, nor previous colon adenomas or familial polyposis syndrome. Only individuals with complete colonoscopies that allowed the full visualization of the entire colon and showed no significant clinical alterations were included.

Colonoscopy and Biopsy Procedures for the NC Subjects

All patients received standard instructions for preparation for colonoscopy that included consumption of 500 ml of mannitol for bowel cleansing, luftal, and bisacodyl. Eligible subjects gave written informed consent to provide colorectal biopsies, had their anthropometric measures taken and answered questions about diet, consumption of alcohol, and tobacco. Colonoscopy was performed using a Pentax videoscope model FC38LX. During biopsy procurement, the rectum was inflated with air and care was taken not to use any suction during advancement of the scope to 7–8 cm from the anal verge. Sterile biopsy forceps were not taken out of the channel of the scope until an area that was completely clear of stool was seen with clear pink mucosa. Biopsies were taken with 2.2 mm sterile standard forceps.

Patients Diagnosed with Rectal Adenocarcinomas

(Rectal-Cancer, RC, $n = 18$): Tumor specimens, located in the higher ($N = 15$), mid ($N = 2$), and lower rectum ($N = 1$), were obtained from surgeries to remove the tumor mass. All subjects belonging to this group were recruited at AC Camargo Cancer Center's Pelvic Surgery Department, in São Paulo, Brazil. We included patients that were diagnosed with rectal adenocarcinoma (tumors of stage pT1 or pT2 low- or mid-straight, pT1 or pT2 or pT3 high-straight), that had not undergone any neoadjuvant therapy and had their tumors surgically resected at the Pelvic Surgery Department, AC Camargo Cancer Center, with diagnosis confirmed by the Pathology Department of the same institution. After the histopathologic confirmation of rectal adenocarcinoma diagnosis, surplus samples were macrodissected by an experienced pathologist and used for DNA extraction and bacterial community profiling. *Exclusion*

criteria were: Patients subjected to neoadjuvant therapy prior to tissue collection; patients reporting inflammatory bowel diseases or with hereditary cancer syndromes. We also excluded all subjects (cases and controls) who reported the use of antibiotics for at least 4 weeks prior to sample-collection.

DNA Extraction

DNA extraction started after incubating the samples for 18 h in 600 μl of a lysis buffer (Qiagen) and 15 μl of proteinase K (20 μg/μl) at 55°C. After this period, DNA samples were extracted using a standard phenol chloroform protocol, followed by ethanol precipitation, quantification using a spectrophotometer (Nanodrop—Thermo Scientific), and visualized on 2% agarose gels to inspect DNA integrity.

PCR Amplification and Sequencing of the V4–V5 Region of 16S rRNA Gene

Oligonucleotide Primer Selection and Coverage Analysis

The V4–V5 region was amplified using a primer set designed to generate amplicons compatible with the chemistry available for the Ion Torrent PGM platform, that allowed ~400 nt of high quality sequences (Ion PGM Sequencing 400 Kit). Coverage of the primer set was evaluated using the Ribosomal Database Project's (RDP—Release 11.2) ProbeMatch (Cole et al., 2014) and the ARB Silva's (Release 115) TestPrime (Klindworth et al., 2013). The forward primer (5′-AYTGGGYDTAAAGNG-3′) and reverse primer (5′-CCGTCAATTCNTTTRAGTTT-3′) corresponded to positions 562 and 906, respectively, of the *Escherichia coli* 16S rRNA gene.

PCR Amplification and Amplicon Sequencing

Three 50 μl amplification replicate reactions were performed per sample, each containing: 2.5 μM of each primer; 25 μl of Kapa Hotstart High Fidelity Master Mix (Kapa Technologies) and 25 ng of genomic DNA (gDNA). Thermocycling conditions were: 95°C, 3 min; 98°C, 15 s, and 40°C, 30 s for 35 cycles; followed by a last extension step at 72°C for 5 min. Amplicons of the three reactions from each subject were pooled and purified using a MinElute PCR Purification Kit (Qiagen). The purified products were run on 1.5% agarose gels and gel bands within the expected amplicon range were excised using sterile and disposable scalpels and purified using the Qiaquick gel extraction kit (Qiagen) to remove artifacts, primer-dimers and non-specific bands. Amplicons were end-repaired and Ion Torrent adaptors with barcodes were ligated. Equimolar amounts of amplicons from each sample were pooled, using the Ion Torrent qPCR quantitation kit (Thermo Scientific, Carlsbad, USA), and used for emulsion PCR. All samples were sequenced on the Ion torrent PGM platform (Thermo Scientific, Carlsbad, USA) using two 318 v2 chips. Samples from both groups were processed simultaneously, to avoid possible batch effects.

Sequence Analysis

Sequence Filtering

Sequences processed by the latest version of the Ion Torrent server (v3.6.2) were used as input into the *Qiime* (*Quantitative insights into microbial ecology*) software package (Version 1.6.0) (Caporaso et al., 2010a). We first removed sequences with an average quality score <20 using a 50 nt sliding window. Then, we identified barcodes used for subject-assignment, allowing a maximum of 2 mismatches, and discarded sequences with no barcodes, and <200 nt or >500 nt after barcode removal. PCR primers identified at the start or at the end of the reads, allowing a maximum of 4 nt mismatches, were trimmed and sequences with no identifiable primers were discarded. After primer trimming we removed all sequences below 200 nt and the remaining sequences were used as input for downstream analysis.

Sequence Clustering and OTU Filtering

Filtered sequences were clustered with 97% identity using UPARSE (implemented in USEARCH v7) (Edgar, 2013) and the seed sequence of each cluster was picked as a representative. Chimeric sequences (and clusters) were identified using UCHIME (Edgar et al., 2011) and the Broad Institute's chimera slayer database (version microbiomeutil-r20110519) and excluded from further analysis. The RDP classifier (Wang et al., 2007), as implemented within the *Qiime* interface (default parameters), was used to assign taxonomic ranks using a minimum confidence value of 80% and, subsequently, to each operational taxonomic unit (OTU). Unless otherwise stated, OTUs that occurred in less than 25% of all samples and with less than 3 reads were not considered.

Alpha and Beta Diversity Analysis

We rarefied the OTU table to 17,414 sequences per sample in order to calculate species diversity, using the Shannon-Weaver index (Shannon, 1948) and the Simpson index (Simpson, 1949), and richness (by using the observed species) implemented in the R Phyloseq package (McMurdie and Holmes, 2013).

For beta diversity analysis, OTU-representative sequences were aligned using PyNAST (Caporaso et al., 2010b) against the aligned green genes core set (DeSantis et al., 2006) with *Qiime* default parameters, and the alignments were lanemask-filtered (Lane, 1991). A phylogenetic tree was built using FastTree (Price et al., 2009), weighted and unweighted UniFrac (Lozupone and Knight, 2005) distances were calculated and a distance matrix was generated. Using the R phyloseq package, distance matrices were used to calculate coordinates for principal coordinate analysis (PCoA).

Enterotypes

Community types of each sample were analyzed by the Dirichlet multinomial mixture model-based method (Holmes et al., 2012) using rarefied genera level counts of 16S rRNA sequencing reads. Partioning around medoids (PAM) enterotyping was performed in R using genera level relative abundances and the "cluster" package (Maechler et al., 2015). We applied 4 distance metrics: Weighted UniFrac, Unweighted UniFrac, root Jensen-Shannon divergence, and Bray-Curtis and assessed the quality of

the clusters using prediction strength (Tibshirani and Walther, 2005), silhouette index (Rousseeuw, 1987), and the Caliński-Harabasz statistic (Calinski and Harabasz, 1974) using the "fpc" R package (Hennig, 2015).

Differential Abundance Analysis

To investigate differences in OTU, phyla and genera abundances between both groups, raw counts were normalized then log transformed using the normalization method below, as performed by a previous study (Sanapareddy et al., 2012):

$$Normalized\ count = \log_{10}((\frac{raw\ count}{number\ of\ sequences\ in\ that\ sample})$$
$$\times\ average\ number\ of\ sequences\ per\ sample + 1)$$

We also evaluated high-level phenotypical differences in microbial composition between both groups. Quality control passed sequences were closed-reference picked at 97% identity using UCLUST_Ref (Edgar, 2010) and the green genes core set (Version 13.5). The resulting OTU table was rarefied to 13,944 sequences and submitted to BugBase (http://github.com/danknights/bugbase) in order to calculate differences between both groups in terms of microbial phenotypes.

Data Validation

Digital Droplet PCR of Bacteroides Fragilis 16S rRNA

We detected and quantified the absolute number of 16S rRNA *B. fragilis* copies in our samples using the *QX200^{TM} Droplet Digital^{TM} PCR System* (Bio-Rad). The primers used to amplify the *B. fragilis* 16S rRNA gene were: BF-fwd 5′-TCRGGAAGAAAGCTTGCT-3′ and BF-rev 5′-CATCCTTTACCGGAATCCT-3′(Tong et al., 2011) and to ensure further specificity, a labeled probe BF-p 5′(FAM)-ACACGTATCCAACCTGCCCTTTACTCG-3′ (BHQ1) (Tong et al., 2011) was included in the reaction. We used a commercial RNAseP *Copy Number Reference Assay* (Thermo-Fisher) to detect and quantify human DNA. Microdroplets (≈20.000/reaction) were generated on the Bio-Rad *QX-100* following the manufacturer's instructions. RNAse P and *B. fragilis* ddPCR were performed in 96 well-plates, in a final volume of 20 μl, containing: 15 ng of total DNA, 10 ul of ddPCR supermix for probes (Bio-Rad), 8 pmol of each PCR BF-primer and 2 pmol of the BF-probe, or 1 μl of RNAse P assay. PCR conditions were: 50°C- 2 min; 95°C- 10 min; 95°C- 15 s and 60°C- 1 min for 40 cycles. After cycling, the 96-well plate was immediately transferred on a *QX200 Droplet Reader* (Bio-Rad), where flow cytometric analysis determined the fraction of PCR-positive droplets vs. the number of PCR-negative droplets in the original sample. Data acquisition and quantification was carried out using *QuantaSoft Software* (Bio-Rad). To ensure the accuracy of the results, a minimum of 10,000 acceptable droplets per reaction were required for quantification using the *QuantaSoft software*. Samples yielding a minimum of 3 positive droplets from 10–15,000 droplets analyzed were scored as positive.

Immunohistochemistry

Immunohistochemistry was performed in an automated Benchmark platform (Ventana Medical Systems, USA) for Anti-*B. fragilis* LPS antibody (mouse monoclonal—Abcam 1265/30) in whole slide tissues. Alkaline phosphatase conjugated to secondary polymeric system was used for IHC visualization. The selection of positive and negative samples was guided by the high-throughput sequencing (HTS) data and used to confirm the presence of *B. fragilis* in the sample set. The primary antibody was omitted to evaluate background staining.

Statistical Analysis

Wilcoxon tests were used to compare mean differences between tumor and biopsy samples for phyla, genera and OTU log-abundances. Considering t = total number of taxa tested, p = raw p-value and R = sorted rank of the taxon, *P*-values were corrected for multiple testing (Sanapareddy et al., 2012) using:

$$Adjusted\ p\ value\ =\ \frac{t\ \times\ p}{R}$$

Fold changes for each genera/OTU were calculated using:

$$Log2\,FC\ =\ \log_2(RC\ average\ +\ 1)\ -\ \log_2(NC\ average\ +\ 1)$$

Chi-Square tests were performed on subject's categorical data such as gender, alcohol and tobacco use and vital status. Student *t*-tests were performed to compare differences in the means between both groups for age, height, weight, BMI, and alpha diversity. We used ANOSIM and ADONIS (Oksanen et al., 2016) to compare differences in beta-diversity between groups using 3 distance metrics weighted UniFrac, unweighted UniFrac and Bray-Curtis for categorical, and numerical variables, respectively. Linear models were built using normalized counts at the genera and OTU level to investigate associations with clinical-pathological characteristics of rectal-cancer samples, such as lymph node and perineural neoplastic invasion status. Unless otherwise stated, values were reported as mean ± *SD* (standard deviation) and *P*-values <0.05 were considered statistically significant. All calculations were performed within the R statistical computing environment (R Foundation, 2011) unless otherwise stated.

RESULTS

Subjects and Tissue Sample Characteristics

We analyzed tissue-associated bacteria from mucosal biopsies of 18 non-cancer controls and 18 rectal adenocarcinoma tumors using 16S rRNA high throughput amplicon sequencing. We found no significant differences between rectal-cancer and non-cancer subjects regarding age and gender distribution, tobacco, and alcohol use and other risk factors (**Table 1**). All samples consisted of rectal-biopsies. The biopsies of individuals with no tumor lesions derived from the mid rectum and were distributed along the ~12 cm-long human rectum, with most samples deriving from the higher-mid rectum (94%).

TABLE 1 | Subject and sample data.

Demographic	Non-cancer (*n* = 18)	Rectal-cancer (*n* = 18)	*P*-value
Age	55.2 ± 15.7	59.3 ± 8.8	0.348
GENDER (%)			
Female	9 (50)	8 (44)	1
Male	9 (50)	10 (56)	
Height	1.65 ± 0.08	1.70 ± 0.09	0.1
Weight	73 ± 14.1	73.8 ± 13.5	0.87
BMI	26.6 ± 3.7	25.3 ± 3.6	0.29
ALCOHOL USE (%)			
Yes	8 (44)	5 (28)	0.568
No	10 (56)	12 (67)	
Undetermined	0 (0)	1 (5)	
TOBACCO USE (%)			
Yes	12 (67)	6 (28)	0.129
No	6 (33)	11 (62)	
Undetermined	0 (0)	1 (5)	
Pathological tumor size staging (%)			
pT2	N.A.	5 (28)	N.A.
pT3		13 (72)	
Pathological lymph node metastasis staging (%)			
pN0	N.A.	11 (62)	N.A.
pN1		3 (16)	
pN2		4 (22)	
Distant metastasis staging (%)			
M0	N.A.	18 (100)	N.A.
Invasion (%)			
Perineural	N.A.	4 (22)	N.A.
Angiolymphatic		14 (78)	
Vital status (%)			
Alive	18 (100)	17 (95)	1
Deceased	0 (0)	1 (5)	

N.A., Not applicable.

Primer Coverage

Our analyses indicate that the PCR primers used here (V4–V5 region of the 16S rRNA gene) cover 84.4 and 52.1% of all eubacterial sequences present in the ARB SILVA database and the Ribosomal Database Project, respectively (Supplementary Table 1). Coverage rates were evenly distributed among most bacterial phyla, except for *Verrucomicrobia*, where coverage rates were 21 and 10.9%, dropping below the 75 and 48% averages of taxa present in the SILVA and RDP databases, respectively.

Sequence Analysis
Sequence Generation and Filtering

A total of 12,078,140 sequence reads were generated, with a mean sequence length of 304.5 ± 97.34 nt (standard deviation—std). After quality filtering and primer trimming,

5,593,020 (46.3%) sequences remained, with an average of 155,361 sequences/sample and a mean sequence length of 315 ± 30 nt.

Sequence Clustering and OTU Filtering

When all individuals were considered, a total of 3222 OTUs were obtained. Thirty-one (0.7%) OTUs were identified as chimeras by UCHIME and 209 (4.7%) could not be assigned to a taxonomic rank. After filtering OTUs with less than three sequences and not present in at least 25% of all samples (NC and RC combined), 1492 OTUs remained.

Alpha and Beta Diversity

Species Richness and Diversity

We observed significantly higher species richness and species diversity in rectal cancer samples compared to controls. This was observed for the number of OTUs, the Shannon index and the Simpson Index (P-values = 0.002, <0.001, and <0.001, respectively) (**Figures 1A,B**). When we stratified rectal-cancer samples into smaller (pT2) and larger tumors (pT3), we observed an increase in species richness, with an average of 280 and 366 OTUs, respectively, compared to 236 OTUs in NC; however this effect reached no statistical significance between pT2 and pT3, maybe because of the reduced number of pT2 samples ($N = 5$, compared to $N = 13$ for pT3) (**Figure 1B** and data not shown).

Beta Diversity

Using three distance metrics we observed consistent and statistically significant differences between the sample groups when considering cancer status (Bray-Curtis, Unweighted and Weighted UniFrac; p-value: 0.001; ANOSIM using 999 permutations), but not for any other categorical or numerical variable, which included amplicon library construction, age, gender, BMI, alcohol, and tobacco use (**Figure 1D**; Supplementary Table 2).

Enterotypes

Enterotyping analysis using a Dirichlet multinomial mixture model divided our cohort in two clusters (**Figures 2A–C**). Enterotype I was significantly enriched for rectal-cancer samples, whilst enterotype II was composed mostly of non-cancer samples (p-value: 0.0001, Fisher's exact test). Enterotype I had higher abundances of *Bacteroides, Clostridiales, Dorea,* and other genera, whilst enterotype II was characterized by elevated amounts of *Pseudomonas* and *Brevundimonas* (**Figure 2D**). When using the PAM based enterotyping method and criterion adopted by a meta-analysis of human enterotypes (Koren et al., 2013), we found two enterotypes with prediction strength above 0.9 (meaning that 90% of the data points fall within the cluster and 10% are outliers) using the Weighted UniFrac distance (Supplementary Figure 1).

Global Signatures of the Microbial Community

Phyla Log Abundances

We observed a significant difference in the log abundances of 6 out of 12 detected phyla between both groups (Supplementary Figure 2). The most abundant phyla identified were (in decreasing order) *Proteobacteria, Firmicutes, Bacteroidetes, Fusobacteria, Actinobacteria,* and *Verrucomicrobia.* In non-cancer samples, we observed higher log abundances of *Actinobacteria, Cyanobacteria, Proteobacteria,* and *Planctomycetes,* whose presence was detected in 9/18 NC samples, with an average log abundance of 0.54 and was absent from all RC individuals (p-value < 0.001). In rectal-cancer we found greater log abundances of *Bacteroidetes* and of the much less known candidate phylum *OD1* (also known as *Parcubacteria),* whose presence was detected in 14/18 RC samples with an average log abundance of 0.71 vs. 1/19 NC samples and an average log abundance of 0.02 (p-value < 0.001).
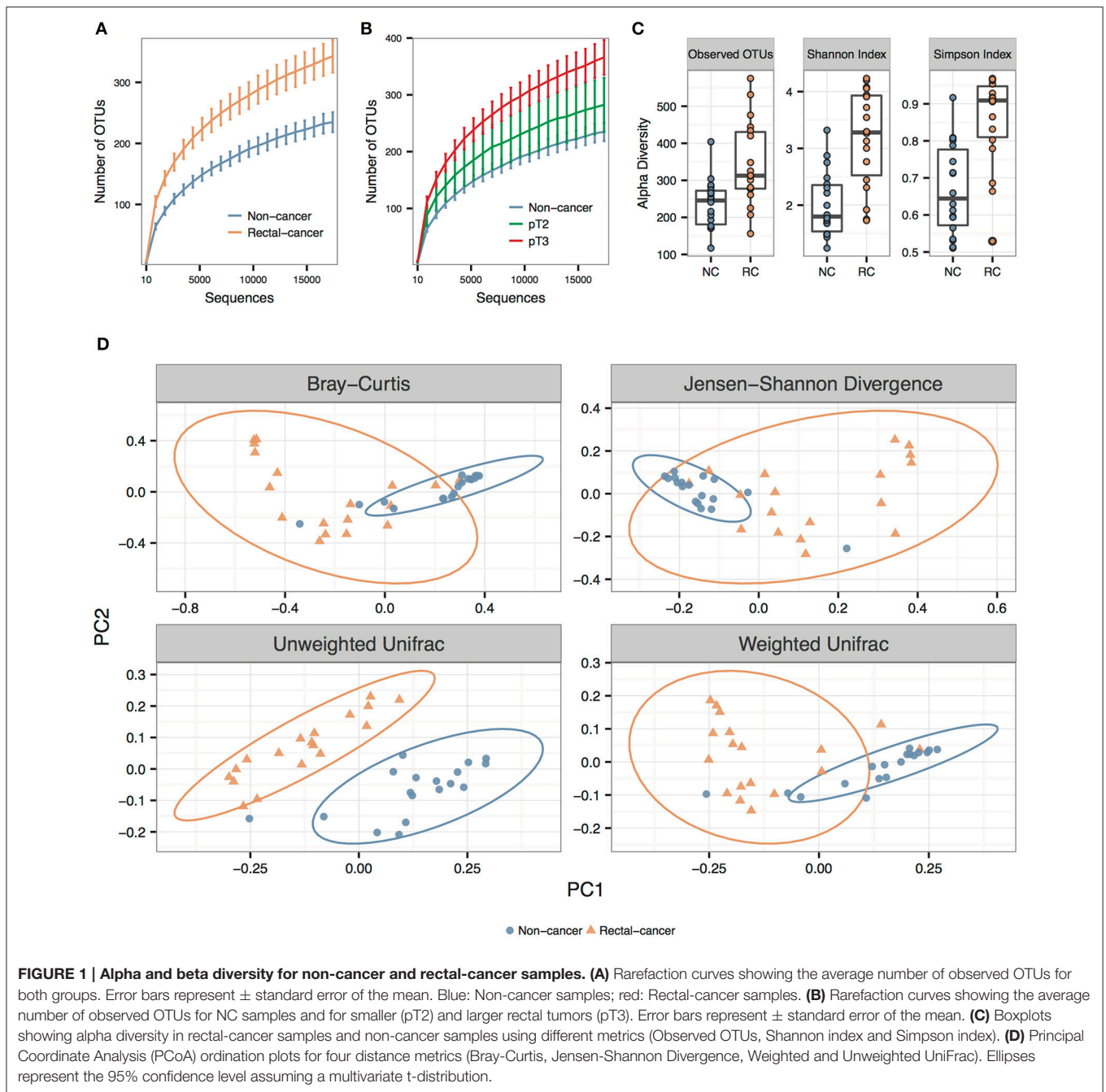
Genera Log Abundances

At the genus level, 86 out of 260 genera (33%) showed significant differential log abundances between both groups (**Figure 3A** and Supplementary Table 3). The top five genera with differential log abundances between the groups were *Bacteroides, Phascolarctobacterium, Odoribacter, Parabacteroides, Desulfobrio* (more abundant in the cancer group), and *Lactobacillus, Pseudomonas, Bacillus, Escherichia, Acinetobacter* (more abundant in the non-cancer set) (**Figure 3B**).

OTU Log Abundances

Of the 1492 OTUs identified, 163 (10.9%) were found to have significant differential log abundances between both groups (**Figure 3C**). Three OTUs assigned to the genus *Bacteroides,* two belonging to *B. fragilis* and one to *B. uniformis,* as well as OTUs assigned to *Bilophila* sp. and *Fusobacterium* sp., were significantly more abundant in rectal-cancer samples (**Figure 3D**). In non-cancer samples, OTUs assigned to *Alcaligenes faecalis, Bacillus cereus, Lactobacillus delbruecki, Prevotella melaninogenica* and *Pseudomonas* ssp had higher log abundances compared to rectal-cancer samples. Four OTUs belonging to the *Bacilli* class were more abundant among non-cancer samples, including *L. delbrueckii* (**Figure 3D**).

When analyzing high-level phenotypical differences, the most striking differences included a higher abundance of anaerobic bacteria and a deficit in biofilm-forming bacteria in rectal-cancer samples (Supplementary Figure 3). In our searches for associations between rectal-cancer samples' clinical data and genera/OTU log abundances using linear regression, we found significant associations between genera/OTUs with regards to the presence of lymph node disease (Supplementary Table 4) and perineural invasion (data not shown). We found a significant increase of *Coprococcus, Dorea, Roseburia,* and *Mogibacterium* in lymph node positive rectal-cancer (Supplementary Figure 4).

FIGURE 1 | Alpha and beta diversity for non-cancer and rectal-cancer samples. (A) Rarefaction curves showing the average number of observed OTUs for both groups. Error bars represent ± standard error of the mean. Blue: Non-cancer samples; red: Rectal-cancer samples. (B) Rarefaction curves showing the average number of observed OTUs for NC samples and for smaller (pT2) and larger rectal tumors (pT3). Error bars represent ± standard error of the mean. (C) Boxplots showing alpha diversity in rectal-cancer samples and non-cancer samples using different metrics (Observed OTUs, Shannon index and Simpson index). (D) Principal Coordinate Analysis (PCoA) ordination plots for four distance metrics (Bray-Curtis, Jensen-Shannon Divergence, Weighted and Unweighted UniFrac). Ellipses represent the 95% confidence level assuming a multivariate t-distribution.
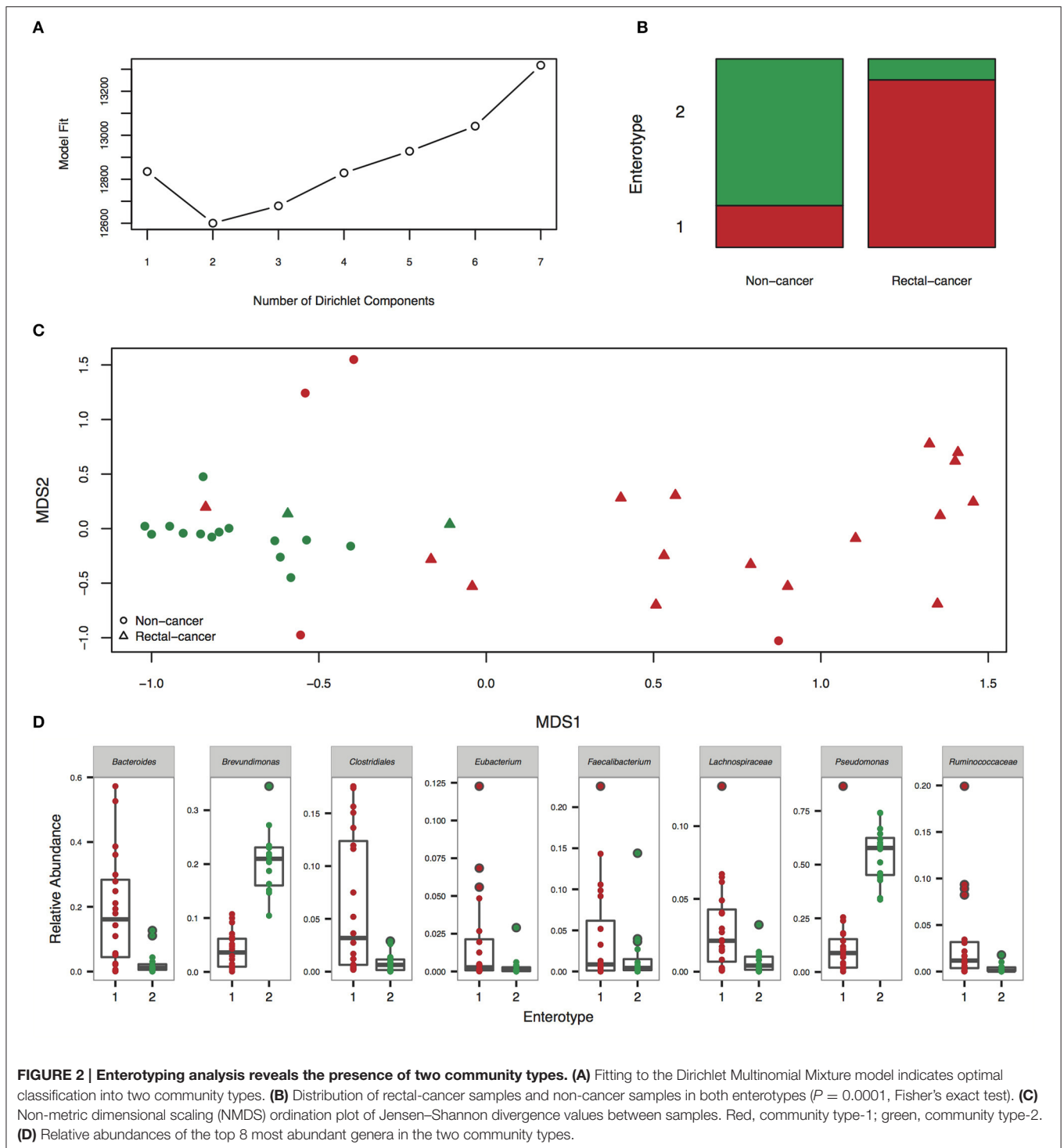
ddPCR Confirms the Higher Counts of *B. fragilis* in Tumor Samples

As two OTUs classified as *B. fragilis* were among the smallest *p*-values found and with the highest fold change between the groups, we designed a specific ddPCR assay for *B. fragilis* in order to verify the validity of the results using an alternative approach. As can be seen in **Figures 4A,B**, we observed the expected correlation ($R^2 = 0.78$) between both methods and confirmed the higher ratio of *B. fragilis*/human DNA in rectal cancer samples, validating the results of our sequencing approach (P-value $= 0.04$, Wilcoxon Rank-Sum Test). To further evidence

the presence of *B. fragilis* in tumor specimens, we performed an immunohistochemistry assay on 3 rectal-cancer samples using an anti-*B. fragilis* LPS antibody and found that this bacterium was present in rectal-cancer tissue (**Figures 4C,D**).
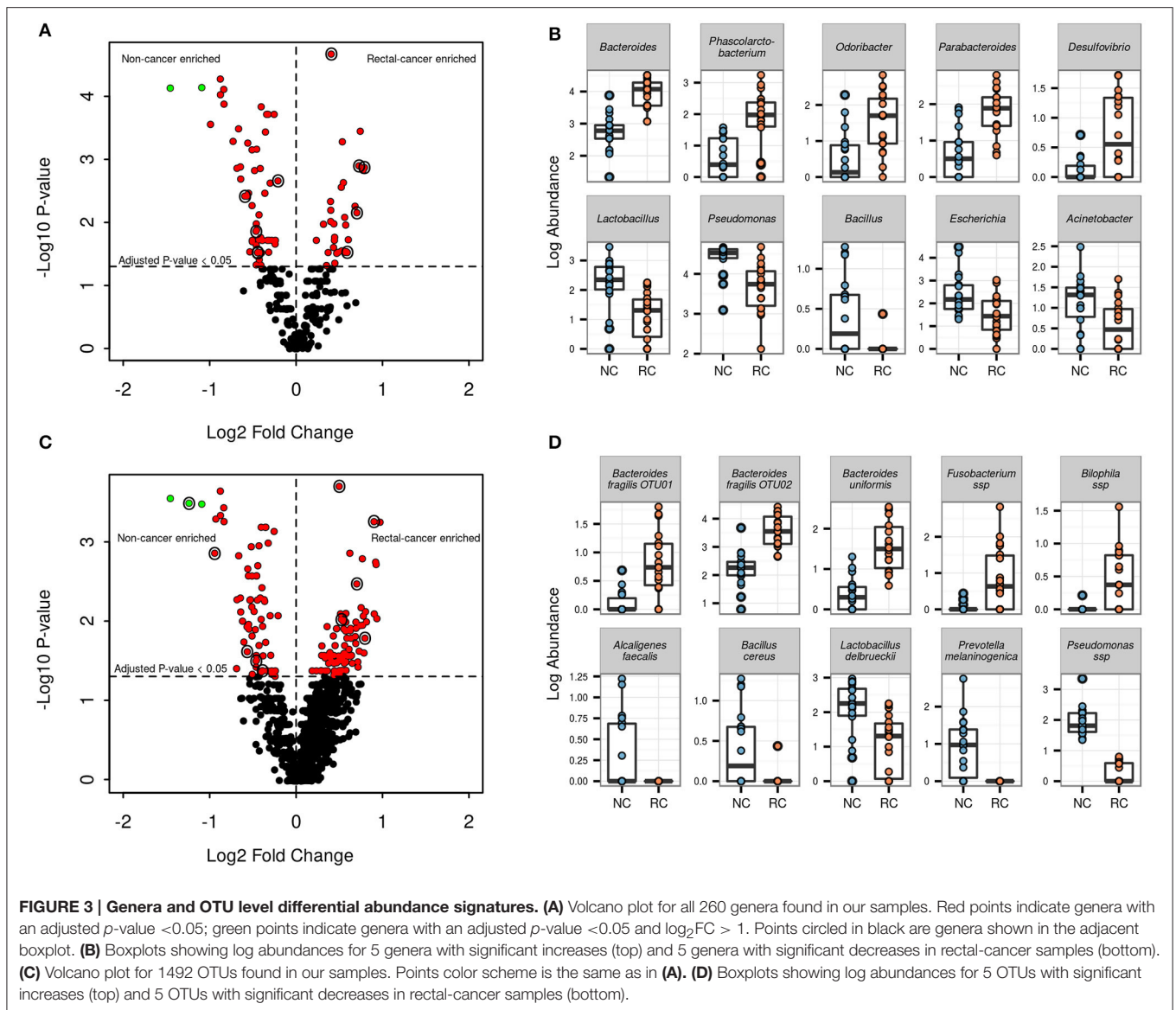
DISCUSSION

In face of the microbiota gradient found in the human digestive tract (Zhang et al., 2014; Gao et al., 2015; Flemer et al., 2016) and the possibility that tissue-associated microorganisms could play a more direct role in immunomodulation and cancer

FIGURE 2 | Enterotyping analysis reveals the presence of two community types. (A) Fitting to the Dirichlet Multinomial Mixture model indicates optimal classification into two community types. **(B)** Distribution of rectal-cancer samples and non-cancer samples in both enterotypes ($P = 0.0001$, Fisher's exact test). **(C)** Non-metric dimensional scaling (NMDS) ordination plot of Jensen–Shannon divergence values between samples. Red, community type-1; green, community type-2. **(D)** Relative abundances of the top 8 most abundant genera in the two community types.

development, we investigated bacterial populations present in tissue biopsies, which may be relevant to pathological processes. Instead of studying colon and rectum samples together, our work is more specific as it is focused and contains only rectal tumors. Whereas, we achieved high 16S rRNA coverage from a large spectrum of bacteria from cancer samples, before

any therapeutic intervention, we also see limitations, such as our relatively small sample size of 36 individuals. However, effect size analysis (Kelly et al., 2015) between both groups revealed an ω^2 ranging from 0.13 to 0.26, depending on the metric of pairwise distance, with PERMANOVA p-values <0.001 and power of 1 (data not shown), a finding that indicates
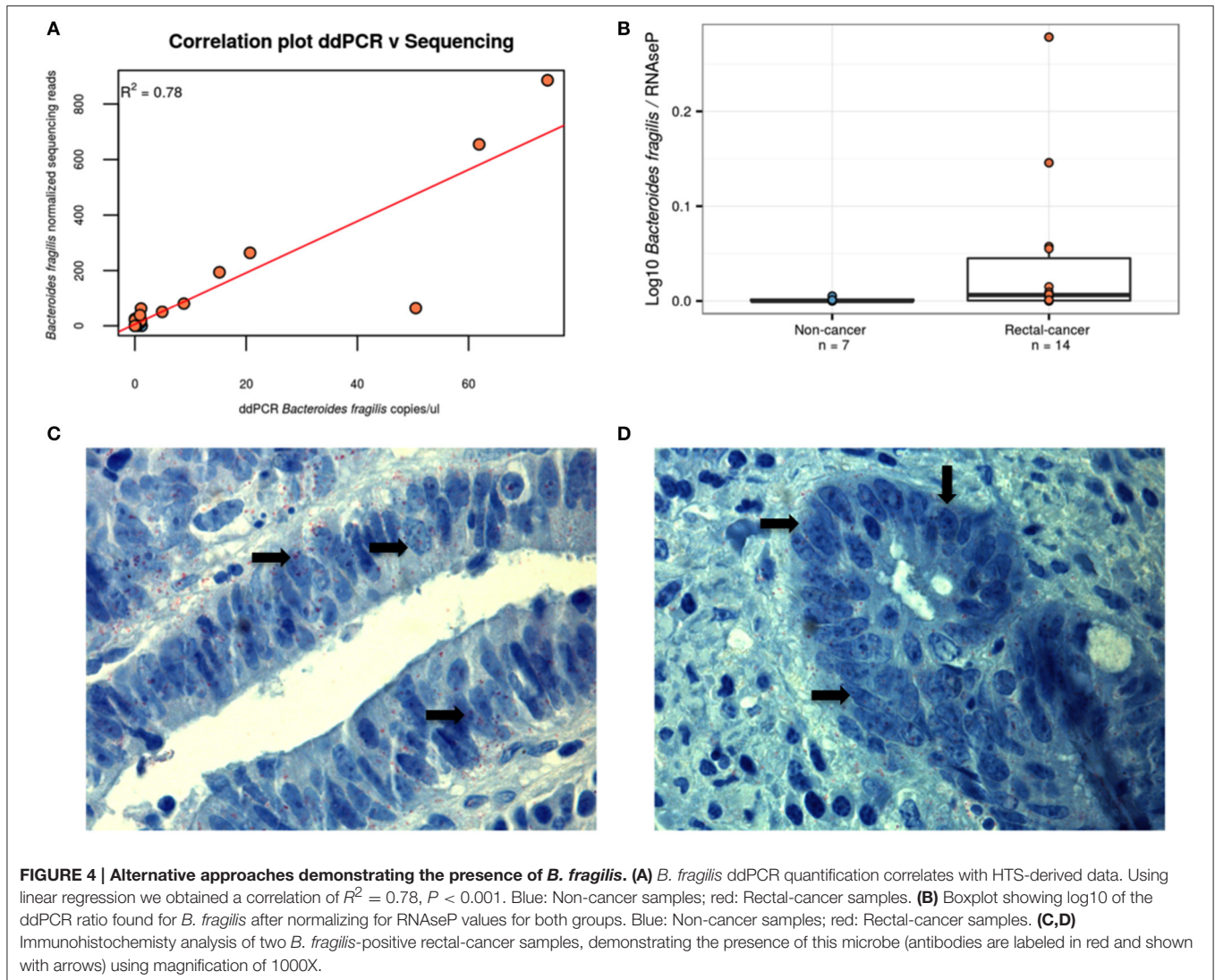
FIGURE 3 | Genera and OTU level differential abundance signatures. (A) Volcano plot for all 260 genera found in our samples. Red points indicate genera with an adjusted *p*-value <0.05; green points indicate genera with an adjusted *p*-value <0.05 and log$_2$FC > 1. Points circled in black are genera shown in the adjacent boxplot. **(B)** Boxplots showing log abundances for 5 genera with significant increases (top) and 5 genera with significant decreases in rectal-cancer samples (bottom). **(C)** Volcano plot for 1492 OTUs found in our samples. Points color scheme is the same as in **(A)**. **(D)** Boxplots showing log abundances for 5 OTUs with significant increases (top) and 5 OTUs with significant decreases in rectal-cancer samples (bottom).

that this sample size allows the observation of significant microbial differences between our two sample groups. We also need to point out that, whereas the primer pair used here gives a good coverage of most phyla, it has a poor coverage of the two closely related bacteria phyla *Lentisphaerae* and *Verrucomicrobia*.

In our study, we observed increased species-diversity and -richness among rectal-cancer samples. Higher species-diversity and -richness were seen in rectal tissue samples from adenomas compared to normal samples (Sanapareddy et al., 2012) and CRCs vs. adenomas (Nakatsu et al., 2015) and increased richness was found in CRCs compared to both adenomas and controls (Mira-Pascual et al., 2015). However, when looking at fecal samples, studies have had conflicting results. One study found increased diversity of both genes and genera along the adenoma-carcinoma transition

(Feng et al., 2015), whereas another found a decrease in diversity when comparing carcinoma samples and normal controls (Ahn et al., 2013) and a third found no differences between controls, adenomas and carcinomas (Zeller et al., 2014). It is noteworthy to state that these fecal studies grouped proximal and distal colon cancers together with rectal cancers, which could have led to differences in their results. We should note that the five cases of early-stage lesions (pT2) showed, on average, intermediate microbial richness, when compared to non-cancer biopsies and a more advanced neoplastic stage (pT3). This suggests that increased species richness of cancer lesions could have an early role in rectal carcinogenesis.

Inter-individual microbial community heterogeneity of the human gut is influenced by spatial distribution, micro-heterogeneity, host genetics, dietary preferences,

FIGURE 4 | Alternative approaches demonstrating the presence of *B. fragilis*. (A) *B. fragilis* ddPCR quantification correlates with HTS-derived data. Using linear regression we obtained a correlation of $R^2 = 0.78$, $P < 0.001$. Blue: Non-cancer samples; red: Rectal-cancer samples. (B) Boxplot showing log10 of the ddPCR ratio found for *B. fragilis* after normalizing for RNAseP values for both groups. Blue: Non-cancer samples; red: Rectal-cancer samples. (C,D) Immunohistochemisty analysis of two *B. fragilis*-positive rectal-cancer samples, demonstrating the presence of this microbe (antibodies are labeled in red and shown with arrows) using magnification of 1000X.

and mucin content (Eckburg et al., 2005; Hong et al., 2011; Zhang et al., 2014), and has posed a long-standing challenge when investigating microbial signatures implicated in CRC tumorigenesis. However, our results show that despite the high inter-individual differences, a common microbial community pattern appears to emerge, as shown in the PCoA analysis that clustered non-cancer and rectal-cancer groups separately (**Figure 1D**), suggesting a common dysbiotic setting related to this neoplasia.

We performed a global analysis of high-level phenotypical differences for bacteria identified in both groups. We highlight the higher abundance of anaerobic bacteria in the RC group in agreement with a previous study (Warren et al., 2013) and the reduction of biofilm-forming bacteria. The latter is a finding that may point to barrier breakage that would contribute to rectal colonization by relevant bacteria (Reid et al., 2001) (Supplementary Figure 3).

The alterations we found at the phylum level include higher levels of *Cyanobacteria* (possibly *Melainabacteria*) (Soo et al., 2014), *Actinobacteria*, *Bacteroidetes*, *OD1*, *Proteobacteria*, and *Planctomycetes* in the RC-group. We should note an important abundance difference for bacteria of the candidate phylum *OD1* (*Parcubacteria*). These highly adapted organisms have not been isolated *in vitro* yet; they have small genomes (<1 Mb) and reduced metabolic properties identified in a range of anoxic environments. The absence of biosynthetic capabilities and DNA repair enzymes, derived from the genomic analyses of some *OD1* bacteria, suggests a role as ectosymbionts (Nelson and Stegen, 2015). However, the putative role of these microbes in rectal cancer remains to be determined. A second phylum, *Planctomycetes*, which are atypical bacteria (Fuerst and Sagulenko, 2011) relatively close to *Verrucomicrobia* (Hou et al., 2008) and more frequently observed in aquatic environments (such as saltwater, fresh water, and acidic mud), also showed potential as a

biomarker for RC, with striking differences between the groups.

Interestingly, our study also indicated the differential abundance of more specific microbes after comparing NC and RC groups. *B. fragilis*, a symbiotic organism common to the human intestinal tract, was found to be more abundant in rectal-cancer samples seen by 16S rRNA HTS and confirmed by ddPCR. Other studies that investigated tissue-associated bacteria also found increased abundance of *B. fragilis* in tumor samples (Wang et al., 2012; Zeller et al., 2014; Nakatsu et al., 2015). *B. fragilis* has been identified as an important human intestinal symbiont and has been suggested to act as a "keystone pathogen" in the development of CRC (Hajishengallis et al., 2012). *B. fragilis* is an obligate anaerobe and is a minority member of the normal colonic microbiota with a propensity for mucosal adherence (Sears et al., 2014). Previous reports have linked enterotoxigenic *B. fragilis* (ETBF) to human diarrheal illnesses and increased tumorigenesis in an IL-23-dependent and STAT3-dependent manner (Wick et al., 2014). The toxin fragylisin, produced by ETBF, is a zinc-dependent metalloprotease that triggers NF-kB signaling and cleaves E-cadherin, and has been suggested to be oncogenic (Wu et al., 2009). Bacterial genera known for their role in butyrate production, such as *Ruminococcus, Roseburia,* and *Butyricimonas* were more abundant among rectal-cancers, differing from results reported so far. An explanation for this difference could involve the fact that most data has been derived from fecal samples and/or grouping different anatomical tumor sites (such as proximal, distal, and rectal). An OTU assigned to *Bilophila,* a bile-resistant, strictly anaerobic bacterial genus, was also more abundant among rectal-cancer samples, and evidence suggests that products of bacterial bile acid conjugation, secondary bile acids, are carcinogenic (McGarr et al., 2005; Ridlon et al., 2014). *Desulfovibrio*, a commensal sulfate-reducing bacterium, may contribute to mucosal inflammation through hydrogen sulfide production, a resulting by-product of sulfated mucin metabolism (Earley et al., 2015). *Phascolarctobacterium,* known to produce propionate via succinate fermentation, was also increased among cancer samples. On the other hand, we found that *L. delbrueckii* was more abundant in non-cancer samples. Probiotic *Lactobacilli* can modify the enteric flora and are thought to have a beneficial effect on enterocolitis. Treatment of IL-10-deficient mice with the probiotic *Lactobacillus salivarius* ssp. reduced the intensity of mucosal inflammation and the incidence of colon cancer from 50 to 10%. These effects were accompanied by significant reductions in fecal coliform, enterococci, and *Clostridium perfringens* levels (O'Mahony et al., 2001). This study exemplifies the effect of changes at the flora level on the development of inflammation, and supports the hypothesis that there are "protective" species and "harmful" species in the normal bacterial flora.

After identifying relevant cancer-related microorganisms, the next steps of microbiome studies will certainly involve microbial manipulations to reduce disease-associated agents, or increase the frequency of protective and health-associated microbes. This can be achieved through diet, exemplified by a previous study using animal models that showed taurine consumption lead to a reduction of *Proteobacteria* (especially *Helicobacter*), as well as an elevation in short-chain fatty acids (SCFA) and a reduction in fecal lipopolysaccharides (LPS) (Yu et al., 2016). Duque et al., recently demonstrated, using SHIME® (Simulator of the Human Microbial Ecosystem), that the consumption of non-pasteurized fresh orange juice was able to significantly increase levels of *Lactobacillus* spp., *Enterococcus* spp., *Bifidobacterium* spp., and *Clostridium* spp. and to reduce enterobacteria (Duque et al., 2016).

Long before associations between cancer and the microbial flora started to be uncovered, diet recommendations—including low consumption of red meat and fat, and high ingestion of fibers and vegetables—have been recognized as protective against the development of colorectal cancer. Current evidences suggest that diet recommendations may be effective, together with tissue environment and host-related factors, because they also help shape the gut microbiota (Sonnenburg and Bäckhed, 2016). Further research may show that treatment of rectal dysbiosis may contribute to the prevention of inflammation-induced rectal carcinoma development and aid in chemotherapy and overall treatment response (Yang and Pei, 2006).

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fcimb.2016.00179/full#supplementary-material

Availability of Supporting Data
Nucleotide sequences used for this study have been deposited in the SRA under accession SRP077097.

REFERENCES

Ahn, J., Sinha, R., Pei, Z., Dominianni, C., Wu, J., Shi, J., et al. (2013). Human gut microbiome and risk for colorectal cancer. *JNCI J. Natl. Cancer Inst.* 105, 1907–1911. doi: 10.1093/jnci/djt300

Allais, L., Kerckhof, F.-M., Verschuere, S., Bracke, K. R., De Smet, R., Laukens, D., et al. (2016). Chronic cigarette smoke exposure induces microbial and inflammatory shifts and mucin changes in the murine gut. *Environ. Microbiol.* 18, 1352–1363. doi: 10.1111/1462-2920.12934

Arthur, J. C., Gharaibeh, R. Z., Mühlbauer, M., Perez-Chanona, E., Uronis, J. M., McCafferty, J., et al. (2014). Microbial genomic analysis reveals the essential role of inflammation in bacteria-induced colorectal cancer. *Nat. Commun.* 5, 4724. doi: 10.1038/ncomms5724

Binefa, G., Rodríguez-Moranta, F., Teule, À., and Medina-Hayas, M. (2014). Colorectal cancer: from prevention to personalized medicine. *World J. Gastroenterol.* 20, 6786. doi: 10.3748/wjg.v20.i22.6786

Bingham, S. A. (2000). Diet and colorectal cancer prevention. *Biochem. Soc. Trans.* 28, 12–16. doi: 10.1042/bst0280012

Boulangé, C. L., Neves, A. L., Chilloux, J., Nicholson, J. K., and Dumas, M.-E. (2016). Impact of the gut microbiota on inflammation, obesity, and metabolic disease. *Genome Med.* 8:42. doi: 10.1186/s13073-016-0303-2

Calinski, T., and Harabasz, J. A. (1974). Dendrite method for cluster analysis. *Commun Stat. Simul Comput* 3, 1–27. doi: 10.1080/03610917408548446

Caporaso, J. G., Bittinger, K., Bushman, F. D., DeSantis, T. Z., Andersen, G. L., and Knight, R. (2010b). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26, 266–267. doi: 10.1093/bioinformatics/btp636

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010a). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 2010, 335–336. doi: 10.1038/nmeth.f.303

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., et al. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 42, D633–D642. doi: 10.1093/nar/gkt1244

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72, 5069–5072. doi: 10.1128/AEM.03006-05

Duboc, H., Rajca, S., Rainteau, D., Benarous, D., Maubert, M.-A., Quervain, E., et al. (2013). Connecting dysbiosis, bile-acid dysmetabolism and gut inflammation in inflammatory bowel diseases. *Gut* 62, 531–539. doi: 10.1136/gutjnl-2012-302578

Duque, A. L. R. F., Monteiro, M., Adorno, M. A. T., Sakamoto, I. K., and Sivieri, K. (2016). An exploratory study on the influence of orange juice on gut microbiota using a dynamic colonic model. *Food Res. Int.* 84, 160–169. doi: 10.1016/j.foodres.2016.03.028

Durbán, A., Abellán, J. J., Jiménez-Hernández, N., Ponce, M., Ponce, J., Sala, T., et al. (2011). Assessing gut microbial diversity from feces and rectal mucosa. *Microb. Ecol.* 61, 123–133. doi: 10.1007/s00248-010-9738-y

Earley, H., Lennon, G., Balfe, A., Kilcoyne, M., Clyne, M., Joshi, L., et al. (2015). A Preliminary study examining the binding capacity of *Akkermansia muciniphila* and *Desulfovibrio* spp., to colonic mucin in health and ulcerative colitis. *PLoS ONE* 10:e0135280. doi: 10.1371/journal.pone.0135280

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., et al. (2005). Diversity of the human intestinal microbial flora. *Science* 308, 1635–1638. doi: 10.1126/science.1110591

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* 10, 996–998. doi: 10.1038/nmeth.2604

Feng, Q., Liang, S., Jia, H., Stadlmayr, A., Tang, L., Lan, Z., et al. (2015). Gut microbiome development along the colorectal adenoma–carcinoma sequence. *Nat. Commun.* 6:6528. doi: 10.1038/ncomms7528

Flemer, B., Lynch, D. B., Brown, J. M., Jeffery, I. B., Ryan, F. J., Claesson, M. J., et al. (2016). Tumour-associated and non-tumour-associated microbiota in colorectal cancer. *Gut.* doi: 10.1136/gutjnl-2015-309595. [Epub ahead of print].

Fuerst, J. A., and Sagulenko, E. (2011). Beyond the bacterium: planctomycetes challenge our concepts of microbial structure and function. *Nat. Rev. Microbiol.* 9, 403–413. doi: 10.1038/nrmicro2578

Galluzzi, L., Buqué, A., Kepp, O., Zitvogel, L., and Kroemer, G. (2015). Immunological effects of conventional chemotherapy and targeted anticancer agents. *Cancer Cell* 28, 690–714. doi: 10.1016/j.ccell.2015.10.012

Gao, Z., Guo, B., Gao, R., Zhu, Q., and Qin, H. (2015). Microbiota disbiosis is associated with colorectal cancer. *Front. Microbiol.* 6:20. doi: 10.3389/fmicb.2015.00020

Gevers, D., Kugathasan, S., Denson, L. A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., et al. (2014). The treatment-naive microbiome in new-onset crohn's disease. *Cell Host Microbe* 15, 382–392. doi: 10.1016/j.chom.2014.02.005

Gur, C., Ibrahim, Y., Isaacson, B., Yamin, R., Abed, J., Gamliel, M., et al. (2015). Binding of the Fap2 Protein of *Fusobacterium nucleatum* to human inhibitory receptor TIGIT protects tumors from immune cell attack. *Immunity* 42, 344–355. doi: 10.1016/j.immuni.2015.01.010

Hajishengallis, G., Darveau, R. P., and Curtis, M. A. (2012). The keystone-pathogen hypothesis. *Nat. Rev. Microbiol.* 10, 717–725. doi: 10.1038/nrmicro2873

Hennig, C. (2015). *fpc: Flexible Procedures for Clustering:R Package Version 2.1–10*.

Holmes, I., Harris, K., and Quince, C. (2012). Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS ONE* 7:e30126. doi: 10.1371/journal.pone.0030126

Hong, P.-Y., Croix, J. A., Greenberg, E., Gaskins, H. R., and Mackie, R. I. (2011). Pyrosequencing-based analysis of the mucosal microbiota in healthy individuals reveals ubiquitous bacterial groups and micro-heterogeneity. *PLoS ONE* 6:e25042. doi: 10.1371/journal.pone.0025042

Hou, S., Makarova, K. S., Saw, J. H., Senin, P., Ly, B. V., Zhou, Z., et al. (2008). Complete genome sequence of the extremely acidophilic methanotroph isolate V4, Methylacidiphilum infernorum, a representative of the bacterial phylum Verrucomicrobia. *Biol. Direct* 3:26. doi: 10.1186/1745-6150-3-26

Kelly, B. J., Gross, R., Bittinger, K., Sherrill-Mix, S., Lewis, J. D., Collman, R. G., et al. (2015). Power and sample-size estimation for microbiome studies using pairwise distances and PERMANOVA. *Bioinformatics* 31, 2461–2468. doi: 10.1093/bioinformatics/btv183

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41:e1. doi: 10.1093/nar/gks808

Koren, O., Knights, D., Gonzalez, A., Waldron, L., Segata, N., Knight, R., et al. (2013). A guide to enterotypes across the human body: meta-analysis of microbial community structures in human microbiome datasets. *PLoS Comput. Biol.* 9:e1002863. doi: 10.1371/journal.pcbi.1002863

Kostic, A. D., Chun, E., Robertson, L., Glickman, J. N., Gallini, C. A., Michaud, M., et al. (2013). *Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* 14, 207–215. doi: 10.1016/j.chom.2013.07.007

Kostic, A. D., Gevers, D., Pedamallu, C. S., Michaud, M., Duke, F., Earl, A. M., et al. (2012). Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res.* 22, 292–298. doi: 10.1101/gr.126573.111

Lane, D. J. (1991). "16S/23S rRNA sequencing," in *Nucleic Acid Techniques in Bacterial Systematics*. ed E. G. M. Stackebrandt (Chichester: John Wiley and Sons), 115–175.

Leclercq, S., Matamoros, S., Cani, P. D., Neyrinck, A. M., Jamar, F., Stärkel, P., et al. (2014). Intestinal permeability, gut-bacterial dysbiosis, and behavioral markers of alcohol-dependence severity. *Proc. Natl. Acad. Sci.U.S.A.* 111, E4485–E4493. doi: 10.1073/pnas.1415174111

Lozupone, C., and Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71, 8228–8235. doi: 10.1128/AEM.71.12.8228-8235.2005

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2015). *Cluster Analysis Basics and Extensions. R package version 2.0.1. CRAN*. Available online at: https://cran.r-project.org/web/packages/cluster/citation.html

McGarr, S. E., Ridlon, J. M., and Hylemon, P. B. (2005). Diet, anaerobic bacterial metabolism, and colon cancer: a review of the literature. *J. Clin. Gastroenterol.* 39, 98–109.

McMurdie, P. J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* 8:e61217. doi: 10.1371/journal.pone.0061217

Mira-Pascual, L., Cabrera-Rubio, R., Ocon, S., Costales, P., Parra, A., Suarez, A., et al. (2015). Microbial mucosal colonic shifts associated with the development of colorectal cancer reveal the presence of different bacterial and archaeal biomarkers. *J. Gastroenterol.* 50, 167–179. doi: 10.1007/s00535-014-0963-x

Moore, W. E., and Moore, L. H. (1995). Intestinal floras of populations that have a high risk of colon cancer. *Appl. Environ. Microbiol.* 61, 3202–3207.

Nakatsu, G., Li, X., Zhou, H., Sheng, J., Wong, S. H., Wu, W. K., et al. (2015). Gut mucosal microbiome across stages of colorectal carcinogenesis. *Nat. Commun.* 6:8727. doi: 10.1038/ncomms9727

Nelson, W. C., and Stegen, J. C. (2015). The reduced genomes of Parcubacteria (OD1) contain signatures of a symbiotic lifestyle. *Front. Microbiol.* 6:173. doi: 10.3389/fmicb.2015.00713

Oksanen, J., Blanchet, F., Kindt, R., Legendre, P., and O'Hara, R. (2016). *Vegan: Community Ecology Package. R Package 23-3 2016*:Available online at: https://cran.r-project.org/web/packa

O'Mahony, L., Feeney, M., O'Halloran, S., Murphy, L., Kiely, B., Fitzgibbon, J., et al. (2001). Probiotic impact on microbial flora, inflammation and tumour development in IL-10 knockout mice. *Aliment. Pharmacol. Ther.* 15, 1219–1225. doi: 10.1046/j.1365-2036.2001.01027.x

Pitt, J. M., Marabelle, A., Eggermont, A., Soria, J.-C., Kroemer, G., and Zitvogel, L. (2016). Targeting the tumor microenvironment: removing obstruction to anticancer immune responses and immunotherapy. *Ann. Oncol.* 8, 1482–1492. doi: 10.1093/annonc/mdw168

Price, M. N., Dehal, P. S., and Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* 26, 1641–1650. doi: 10.1093/molbev/msp077

Reid, G., Howard, J., and Gan, B. S. (2001). Can bacterial interference prevent infection? *Trends Microbiol.* 9, 424–428. doi: 10.1016/S0966-842X(01)02132-1

R Foundation (2011). *R Foundation for Statistical Computing*. Vienna AI 3-900051-07-0: R Development Core Team.

Ridlon, J. M., Kang, D.-J., Hylemon, P. B., and Bajaj, J. S. (2014). Bile acids and the gut microbiome. *Curr. Opin. Gastroenterol.* 30, 332–338. doi: 10.1097/MOG.0000000000000057

Riley, D. R., Sieber, K. B., Robinson, K. M., White, J. R., Ganesan, A., Nourbakhsh, S., et al. (2013). Bacteria-human somatic cell lateral gene transfer is enriched in cancer samples. *PLoS Comput. Biol.* 9:e1003107. doi: 10.1371/journal.pcbi.1003107

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65.

Rubinstein, M. R., Wang, X., Liu, W., Hao, Y., Cai, G., Han, Y. W., et al. (2013). *Fusobacterium nucleatum* promotes colorectal carcinogenesis by modulating E-cadherin/β-catenin signaling via its FadA adhesin. *Cell Host Microbe* 14, 195–206. doi: 10.1016/j.chom.2013.07.012

Sabino, J., Vieira-Silva, S., Machiels, K., Joossens, M., Falony, G., Ballet, V., et al. (2016). Primary sclerosing cholangitis is characterised by intestinal dysbiosis independent from IBD. *Gut.* 10, 1681–1689. doi: 10.1136/gutjnl-2015-311004

Sanapareddy, N., Legge, R. M., Jovov, B., McCoy, A., Burcal, L., Araujo-Perez, F., et al. (2012). Increased rectal microbial richness is associated with the presence of colorectal adenomas in humans. *ISME J.* 6, 1858–1868. doi: 10.1038/ismej.2012.43

Sears, C. L., Geis, A. L., and Housseau, F. (2014). *Bacteroides fragilis* subverts mucosal biology: from symbiont to colon carcinogenesis. *J. Clin. Invest.* 124, 4166–4172. doi: 10.1172/JCI72334

Shannon, C. E. (1948). The mathematical theory of communication. 1963. *MD Comput.* 14, 306–317.

Shapiro, H., Thaiss, C. A., Levy, M., and Elinav, E. (2014). The cross talk between microbiota and the immune system: metabolites take center stage. *Curr. Opin. Immunol.* 30, 54–62. doi: 10.1016/j.coi.2014.07.003

Simpson, E. H. (1949). Measurement of diversity. *Nature* 163, 688–688. doi: 10.1038/163688a0

Sonnenburg, J. L., and Bäckhed, F. (2016). Diet-microbiota interactions as moderators of human metabolism. *Nature* 535, 56–64. doi: 10.1038/nature18846

Soo, R. M., Skennerton, C. T., Sekiguchi, Y., Imelfort, M., Paech, S. J., Dennis, P. G., et al. (2014). An expanded genomic representation of the phylum cyanobacteria. *Genome Biol. Evol.* 6, 1031–1045. doi: 10.1093/gbe/evu073

Tamas, K., Walenkamp, A. M. E., De Vries, E. G. E., Van Vugt, M. A. T. M., Beets-Tan, R. G., Van Etten, B., et al. (2015). Rectal and colon cancer: not just a different anatomic site. *Cancer Treat. Rev.* 41, 671–679. doi: 10.1016/j.ctrv.2015.06.007

Tibshirani, R., and Walther, G. (2005). Cluster validation by prediction strength. *J. Comput. Graph. Stat.* 14, 511–528. doi: 10.1198/106186005X59243

Tong, J., Liu, C., Summanen, P., Xu, H., and Finegold, S. M. (2011). Application of quantitative real-time PCR for rapid identification of *Bacteroides fragilis* group and related organisms in human wound samples. *Anaerobe* 17, 64–68. doi: 10.1016/j.anaerobe.2011.03.004

Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-tieulent, J., and Jemal, A. (2015). Global cancer statistics, 2012. *CA Cancer J. Clin.* 65, 87–108. doi: 10.3322/caac.21262

Turnbaugh, P. J., Ridaura, V. K., Faith, J. J., Rey, F. E., Knight, R., and Gordon, J. I. (2009). The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci. Transl. Med.* 1, 6ra14. doi: 10.1126/scitranslmed.3000322

Vétizou, M., Pitt, J. M., Daillere, R., Lepage, P., Waldschmitt, N., Flament, C., et al. (2015). Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. *Science* 350, 1079–1084. doi: 10.1126/science.aad1329

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi: 10.1128/AEM.00062-07

Wang, T., Cai, G., Qiu, Y., Fei, N., Zhang, M., Pang, X., et al. (2012). Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *ISME J.* 6, 320–329. doi: 10.1038/ismej.2011.109

Warren, R. L., Freeman, D. J., Pleasance, S., Watson, P., Moore, R. A., Cochrane, K., et al. (2013). Co-occurrence of anaerobic bacteria in colorectal carcinomas. *Microbiome* 1:16. doi: 10.1186/2049-2618-1-16

Wick, E. C., Rabizadeh, S., Albesiano, E., Wu, X., Wu, S., Chan, J., et al. (2014). Stat3 activation in murine colitis induced by enterotoxigenic *Bacteroides fragilis. Inflamm. Bowel Dis.* 20, 821–834. doi: 10.1097/MIB.0000000000000019

Wu, N., Yang, X., Zhang, R., Li, J., Xiao, X., Hu, Y., et al. (2013). Dysbiosis signature of fecal microbiota in colorectal cancer patients. *Microb. Ecol.* 66, 462–470. doi: 10.1007/s00248-013-0245-9

Wu, S., Rhee, K.-J., Albesiano, E., Rabizadeh, S., Wu, X., Yen, H.-R., et al. (2009). A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. *Nat. Med.* 15:1016–1022. doi: 10.1038/nm.2015

Yang, L., and Pei, Z. (2006). Bacteria, inflammation, and colon cancer. *World J. Gastroenterol.* 12, 6741–6746. doi: 10.3748/wjg.v12.i42.6741

Yu, H., Guo, Z., Shen, S., and Shan, W. (2016). Effects of taurine on gut microbiota and metabolism in mice. *Amino Acids* 48, 1601–1617. doi: 10.1007/s00726-016-2219-y

Zeller, G., Tap, J., Voigt, A. Y., Sunagawa, S., Kultima, J. R., Costea, P. I., et al. (2014). Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol. Syst. Biol.* 10, 766–766. doi: 10.15252/msb.20145645

Zhang, Z., Geng, J., Tang, X., Fan, H., Xu, J., Wen, X., et al. (2014). Spatial heterogeneity and co-occurrence patterns of human mucosal-associated intestinal microbiota. *ISME J.* 8, 881–893. doi: 10.1038/ismej.2013.185

Zitvogel, L., Galluzzi, L., Smyth, M. J., and Kroemer, G. (2013). Mechanism of action of conventional and targeted anticancer therapies: reinstating immunosurveillance. *Immunity* 39, 74–88. doi: 10.1016/j.immuni.2013.06.014