Tumor tissue-specific bacterial biomarker panel for colorectal cancer: *Bacteroides massiliensis, Alistipes species, Alistipes onderdonkii, Bifidobacterium pseudocatenulatum, Corynebacterium appendicis*

Rizwana Hasan1,2 · Sudeep Bose2 · Rahul Roy1 · Debarati Paul2 · Saumitra Rawat3 · Pravin Nilwe4 · Neeraj K. Chauhan5 · Sangeeta Choudhury1

Received: 28 December 2021 / Revised: 26 April 2022 / Accepted: 28 April 2022 / Published online: 26 May 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Human microbiome studies have shown diversity to exist among different ethnic populations. However, studies pertaining to the microbial composition of CRC among the Indian population have not been well explored. We aimed to decipher the microbial signature in tumor tissues from North Indian CRC patients. Next-generation sequencing of tumor and adjacent tissue-derived bacterial 16S rRNA V3-V4 hypervariable regions was performed to investigate the abundance of specific microbes. The expression profile analysis deciphered a decreased diversity among the tumor-associated microbial communities. At the phyla level, Proteobacteria was differentially expressed in CRC tissues than adjacent normal. Further, DeSeq2 normalization identified 4 out of 79 distinct species (*p* < 0.005) only in CRC, *Bacteroides massiliensis*, *Alistipes onderdonkii*, *Bifidobacterium pseudocatenulatum*, and *Corynebacterium appendicis*. Thus, the findings suggest that microbial signatures can be used as putative biomarkers in diagnosis, prognosis and treatment management of CRC.

Keywords CRC · Microbiome · Next-generation sequencing (NGS) · 16S rRNA · Microbial species · Dysbiosis

Introduction

Colorectal cancer (CRC) is a challenging health problem and ranks third in terms of incidence and second in terms of mortality; third most common cancer among men and second most common in women. Overall 1.9 million new cases of CRC and 935,000 deaths were estimated to occur worldwide in 2020 (Sung et al. 2021). The incidence rate is steadily increasing in many countries including Eastern Europe, South Eastern and South Central Asia, and South America (Arnold et al. 2020, 2017). India contributes to the top 5 frequent cases reported worldwide. In countries undergoing a significant transition, with an increasing Human Development Index (HDI), there appears to be a constant rise in the incidence rate of CRC (Fidler et al. 2016). Major risk factors are low fiber diet, lack of regular physical activity, obesity, tobacco, smoking and use of heavy alcohol. In addition, people with certain hereditary cancer syndromes or family history of colorectal cancer have high risk of developing the disease. Since CRC symptoms are difficult to identify because early indicators include constipation, diarrhea, changes in stool color, blood in the stool, bleeding from the rectum, excessive gas, abdominal cramps, and abdominal pain (Siegel et al. 2020); therefore, the key strategies are its timely screening, early diagnosis, and prevention.

In this direction over the last decade, microbiome patterns have been explored as biomarkers to diagnose and predict prognosis (Chen et al. 2012; Raskov et al. 2017). Most literature cited are from the Western population (high
incidence countries), and only a few handful studies have been reported from the Indian perspective. Gupta et al. identified a panel of potential microbial taxonomy and gene markers that discriminated Indian CRC fecal samples from that of the normal samples (Gupta et al. 2019). Exploratory studies pertaining to Indian gut microbiome interplay in healthy individuals remains to be understood. Das et al. examined the gut microbiota in 84 healthy individuals from the northern part of India (high altitude of Ladakh). This study demonstrated region-specific differences in α-/β-
diversity, with Firmicutes dominance followed by Bacteroidetes, Actinobacteria, and Proteobacteria (Das et al. 2018). A study by Dhakan et al. evaluated the gut microbiota in two geographically and socio-demographically distinct cohorts from central (Bhopal) and southern (Kerala) India. The study observed enrichment of genus Prevotella among the vegetarian population and Bacteroides, Ruminococcus, and Faecalibacterium among the omnivorous population (Dhakan et al. 2019). In 2018, Dubey AK et al. carried out a study including 14 geographical locations across India [Landscape of Gut Microbiome-Pan India Exploration (LogMPIE)]. The study revealed 36 unique organisms in northern India, 149 in southern, 95 in western, and 62 in eastern Indian population. The most predominant organisms were Prevotella copri and Faecalibacterium prausnitzii (Dubey et al. 2018).

Although diverse taxonomical microbiome structure has been shown to be associated with different regions of India among healthy individuals, there is a lack of studies pertaining to CRC tissues in Indian patients. Thus, in the present study, the V3–V4 region of 16S rRNA was sequenced using an Ion Torrent™ platform to investigate the microbiome pattern in CRC tumor tissue. This microbial composition derived may provide a clue to CRC diagnosis.

Materials and methods

Patient selection

CRC patients (histopathologically confirmed) undergoing resection, above 18 years, either sex, and were not on antibiotics < 30 days were enrolled in the study. Patients who received radio-/chemo-therapy before surgery and/or on immnosuppressive drug treatments and age > 18 and < 80 years were excluded from the study. Pregnant women were also excluded. Tissue samples from both tumor area (n = 5) and adjacent normal (AN) area (10 cm from tumor site; histologically normal) (n = 5) were aseptically collected and stored at − 80 °C. Ethical consent from institutional ethics committee was obtained (SGRH-EC/06/16/1000) and each patient provided written informed consent.

DNA extraction

Microbial DNA was extracted from tumor/adjacent normal tissue samples using the QIAamp DNA Mini Kit (Qiagen, Valencia, CA, USA) according to the modified protocol from ~ 25 mg of tissue. In brief, each sample was incubated in 180 μl lysozyme (10 mg/ml, Sigma-Aldrich, USA) for 40 min at 37 °C, and further incubated at 56 °C for 16 h with 20 μl proteinase K, until the tissue was completely lysed. Buffer AL was added to this mixture and incubated for 10 min at 70 °C; then, DNA has been eluted according to the manufacturer’s instructions. Microbial DNA quantified using Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, USA). DNA samples were stored at − 20 °C until further processing.

16S rRNA amplicon library generation

16S rRNA hypervariable regions were amplified from extracted DNA by using primer pair Probio_Uni and Probio_Rev (Milani et al. 2013), which targets the V3–V4 region of 16S rRNA gene sequence [520F/802R] (Ventura et al. 2009). The 5’ends of the forward primers were fused with the A-Adaptor (Thermofisher Scientific), 10 base pair barcode key sequence plus barcode adaptor, whereas the reverse primers were fused with the truncated P1-adaptor sequence (trP1), respectively. The complete list of the primers which is used in this study is mentioned in Table 1.

The primers were diluted in a low Tris–EDTA solution (1:10). The fusion amplicons were prepared with 5 ng of bacterial DNA from each sample. In brief, AmpliTag Gold 360 MM (Thermofisher Scientific) containing premixed and premeasured components and 5 pmol of fusion primers were used for 25 μl PCR amplification reaction. The PCR conditions were maintained at 5 min at 94 °C, 35 cycles of 30 s at

| Primer   | Adapter sequence | Key   | Barcode | Barcode adapter | Primer sequence (5′–3′)          |
|----------|------------------|-------|---------|----------------|----------------------------------|
| Probio Uni | CCA TCT CAT CTT GGG TGT TCC GCAC | TCAG  | TTAAACCTC | GAT  | CTTACGGGGRSCGACGACG |
| Probio Rev | CCA TCT CAT CTT GGG CAG TCG GTG CAT | CCTTCCTATGGGCACTCGGTGAT | ATTACCGGGGCTGCT |
| 520F     | CCA TCT CAT CTT GGG TGT TCC GCAC | TCAG  | TTAAACCTC | GAT  | AYTTGGYD TAA AAGNG  |
| 802R     | CCA TCT CAT CTT GGG CAG TCG GTG CAT | TCAG  | TTAAACCTC | GAT  | TACNVGGGT A TCT AATC |

Springer
94 °C, 30 s at 55 °C and 90 s at 72 °C, followed by 10 min at 72 °C. Amplification was carried out using a 9700 Thermocycler (Applied Biosystems, USA). The PCR amplicon was purified using AMPure XP reagent (Beckman Coulter, USA), and the concentration was determined with Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, USA). The respective size distribution of the amplicon was verified with Agilent 2100 Bioanalyzer using a high-sensitivity DNA kit (Agilent, Santa Clara, CA, USA). Amplicon libraries were diluted to 100 pM, and an equimolar pool was prepared for clonal amplification.

**Template preparation and sequencing**

Template preparation for libraries on Ion Sphere™ was performed as per OneTouch 2 protocols and reagents (Thermo Fisher Scientific, USA). Briefly, library amplicons were clonally amplified onto ion sphere particles (ISPs) through emulsion PCR and then enriched for template-positive ISPs. Ion S5 system emulsion PCR reactions utilized the Ion 520TM (Thermo Fisher Scientific, USA). Following recovery, enrichment was completed by selectively binding the ISPs containing amplified library amplicons to streptavidin-coated magnetic beads, removing empty ISPs through washing steps, and denaturing the template strands to allow the collection of template-positive ISPs. For all reactions, these steps were accomplished using the Thermo Fisher Scientific’s ES module of Ion OneTouch 2.

The templated ISPs were loaded on Ion 520 chip kit using the manufacturer’s protocol (Thermo Fisher Scientific, USA). Sequencing was performed with the Ion 520 OT2 (Thermo Fisher Scientific, USA) using the 200 bp chemistry with 500 flow (125 cycle) for the V3 region and 400 bp chemistry with 800 flow (200 cycles) for the V4 region run format. Supplement Fig. SI-1 represents the NGS run summary.

**Ion Reporter platform: metagenomics workflow**

Combining the two primer pools (V3–V4) resulted in broad-range sequence-based identification of bacteria from mixed samples. Ion Reporter Software enabled rapid identification (at genus or species level) of microbes present in samples, using both curated Greengenes and premium curated MicroSEQ ID 16S rRNA reference databases. The Ion Reporter metagenomics workflow provided primer information, classification information, and mapping information.

**16S rRNA sequence data analysis**

Ion Reporter Software 5.10, 16S Metagenomics workflow version 1.0. (Thermo Fisher Scientific, USA) and QIIME2 (https://qiime2.org) were used for DNA sequences processing. Analysis of the sample data was performed utilizing the Ion Reporter Software, which is based on the QIME’s open-source bioinformatics platform to perform diversity analyses (α and β diversity) and visualizations in the form of a Krona diagram. Unaligned binary data files (Binary Alignment Map) created by the Ion Torrent PGM were transferred to an in-house Ion Reporter server (Thermo Fisher Scientific, USA) and analyzed using default settings. Each respective sequence was assigned based on Ribosomal Database Project (RDP) taxonomy. Consensus reads were normalized by converting OTUs for each sample to a percentage of the reads for a given sample. Taxonomy was assigned with a minimum alignment threshold of 97%. The identity score of sequence alignment for genus and species was set at 99% or higher.

**Alpha diversity analysis**

Chao1 index determined the species richness in the samples based on the number of OTU’s. The Shannon and Simpson index was utilized to reflect community diversity, including species richness and species evenness. Relative abundance of phyla and α-diversity metrics were analyzed using the Paleontological Statistics Software Package (PAST). The univariate frequency distribution was conducted to verify data normalization. Boxplots showing the Shannon, Simpson, and Chao1 diversity indices, the upper whisker starts from 75th percentile to the highest value within the 1.5× interquartile range (IQR), the lower whisker extends from 25th percentile to the lowest value within 1.5× IQR of the hinge. Any data points beyond the end of the whiskers were outliers. The ns sign indicates no statistical significance in the pairwise comparison between CRC and AN; p values were obtained using the One-sample t and Wilcoxon test (GraphPad Prism 9).

**Beta diversity analysis**

Principal coordinates analysis (PCoA) was performed to explore the microbial species similarities and/or dissimilarities between tumor and AN samples. In the 3D PCoA plot, the close algorithm (distance matrix) between samples represented similar microbial species.

**MA plot analysis**

MA plot represents the scattered plot of log₂ fold-change (M on the y-axis) and means (A on the x-axis) between two samples. The base mean was the average of normalized count
values, dividing by size factors, taken for all samples. The log2Fold Change is the size estimate.

**OTU Venn chart**

A Venn diagram visually displays the number of common/unique OTUs in the tissue samples. The microbiomes of different samples were obtained by combining with the OTU’s representing the species. Based on the OTU abundance, OTU of each group was listed, and the Venn diagram was drawn using Bioinformatics and Evolutionary Genomics Venn Diagram online tool.

**Heat map and hierarchical clustering**

HeatMap differential abundance was constructed using DESeq2 data with significant OTU’s ($p < 0.05$). The gathered data from DESeq2 were analyzed using NG-CHM GUI 2.20.2 BUILDER for building a Heat Map where the cluster analysis based on each OTU table to create dendrograms using Ward's clustering and Euclidean distance was carried out. The reason for this analysis was to recognize similarities in microorganisms within particular variables (Ryan et al. 2019). The joining of the Euclidean distance as the distance measure and Ward's method as a linkage rule was adopted. The color gradient key displays a linear scale of relative abundance.

**Differential expression analysis**

DESeq2 software was used to normalize by calculating the geometric mean across the 10 samples from five patients and determining the differential expression analysis using the fold-change expression of the bacterial genes between CRC and AN.

**Results**

**Clinical findings**

In this pilot study, demographic data such as age, gender, tumor site and size, tumor stage, and tumor pathology are described in detail in Table 2. Clinical symptoms of anorexia, nausea, vomiting, constipation/diarrhea, abdominal pain, weight loss, stool occult blood findings were noted. The male: female ratio was 4:1; tumor staging was determined according to the AJCC TNM system in CRC (Amin et al. 2017). Most of the patients belonged to stage II or higher, and 4 out of the 5 were obese.

Patients were found to be anemic; mean ± SD hemoglobin was lower than normal (8.7 ± 2.2 g/dl; normal range: 13–17 g/dl). The level of CEA was nearly twice (6.3 ± 3.7 ng/ml; normal range: 0–3 ng/ml), and 3/5 patients showed low levels of protein and albumin (6.2 ± 0.8 g/dl and 3.1 ± 1.0 g/dl, respectively) (Supplementary Table SI-2).

**Taxonomical composition in CRC and adjacent normal tissue**

**Ion Torrent sequencing and pre-processing of reads**

A total of 7.6 million raw reads were generated using the Ion Torrent™ platform. The sequencing indexed on an Ion 520 Chip provided 1.47 GB of data with a mode reads of 240 bp (Supplementary Fig. SI-1). Chips showed 93% loading with 100% enrichment. Pre-processing workflow consisting of a high-quality filtering approach resulted in 89% of the final library (Supplementary Table SI-1).

**Alpha diversity**

Shannon was used to quantify the distribution (evenness) and Simpson's indexes were applied to determine the species diversity; Chao1 was employed to estimate the abundance of community richness in CRC and AN tissue samples. As shown in Fig. 1, the Shannon index and Simpson index showed fewer species-level diversity among CRC tissue samples than AN tissues. Chao1 index showed a similar observation of lower species richness in CRC than that of AN.

**Beta diversity**

Beta diversity two-dimensional coordinate graph was used to measure the similarity or dissimilarity between the microbial

| Patients | Age/sex | Stage | Stage grouping | Localisation | Histopathological findings | Diet | BMI | Obesity |
|---------|---------|-------|----------------|--------------|--------------------------|------|-----|---------|
| 1       | 48/m    | IIb   | pT3pN0pMx      | Rectum       | Moderately differentiated adenocarcinoma | Non-V | 32.5 | Obese   |
| 2       | 82/m    | IIb   | pT3pN0pMx      | Colon        | Moderately differentiated adenocarcinoma | Non-V | 32.6 | Obese   |
| 3       | 83/m    | IV    | pT4apN1pMx     | Colon        | poorly differentiated adenocarcinoma    | Veg   | 20.2 | Normal  |
| 4       | 32/m    | IIIb  | pT3pN1pMx      | Caecum       | Moderately differentiated adenocarcinoma, | Non-V | 26.8 | overweight |
| 5       | 52/f    | IIIc  | pT3pN2pMx      | Rectum       | Moderately differentiated adenocarcinoma, | Non-V | 34.3 | Obese     |
Family of CRC and AN samples. Figure 1d shows close coordinates of 4 CRC samples (CRC P1,2,3,5) except for CRC-P4, whereas all AN samples were observed to be highly diverse in their representation of the microbial family. This PCoA graph showed a significant distinction in bacterial family composition between the CRC and AN tissue samples of the five patients.

**Microbial abundance and distribution**

Analysis of the V3–V4 region of filtered reads showed 17 phyla, 28 classes, 55 orders, 149 families, 222 genus, and 361 species (Supplementary Fig. SI-3). Figure 2 depicts the relative abundance of phyla in both tumor and adjacent normal tissue. The five most abundant phyla were Firmicutes (42.9%), Proteobacteria (38.4%), Actinobacteria (7.8%), Bacteroidetes (8.3%), and Fusobacteria (1.4%) comprised of 98.9% (SI-2). Proteobacteria: Firmicutes ratio in CRC was 1:0.6 and in AN was 1:1.1 (Supplementary Fig. SI-2). Detailed depictions using Krona graphs representing phylum, class, order, family, genus, and species are available in Supplementary Fig. SI-5. As per the metagenomics analysis, the percentage of each phyla abundance in CRC and AN was represented.

![Fig. 1 A Indices of microbial diversity. Box plot showing species-level microbial diversity of five different patient samples of CRC (red) and AN (blue). The alpha diversity was measured by diversity indices, including the Shannon index (a), Simpson index (b), and Chao1 (c). Alpha diversity distances calculated using phylotype relative abundance measurements between AN and CRC tissues demonstrate that the microbial richness of AN tissues is higher than their CRC tissue, while the diversity has no statistical significance (ns) between the two groups. The indices in investigated individuals were computed using the PAST software. B Represents β diversity Principal-coordinate analysis (PCoA), Euclidean plot at Family level, in which each dot represents a sample, and each color represents a group: red for group CRC and blue for group AN. The plot describes the similarities and dissimilarities between samples of five patients (P1-P5). PC1 is the principal coordinate component causing the most considerable difference in samples, with an explanatory value of 64.16%, and PC2 and PC3, 22.82% and 6.65%, respectively (colour figure online).](image-url)
in Supplementary Fig. SI-4. The taxonomic compositions of the remaining 1.02% were majorly represented by Acidobacteria, Chloroflexi, Deinococcus-Thermus, Elusimicrobia, Lentisphaerae, Nitrospinae, Nitrospirae, Spirochaetes, Synergistetes, Tenericutes, and Verrucomicrobia as shown in Supplementary Fig. SI-4. Cyanobacteria (0.14%) was abundant in CRC and Synergistetes (0.3%) in AN among these rare phyla.

**Differentially enriched operational taxonomic units in tissues of CRC and adjacent normal**

The DESeq2 data based on the MA plot (Fig. 3) illustrates the statistically differential gene expression pattern at species level between CRC and AN on a logarithmic scale of base 2, wherein each data point represent a bacterial species. The central (faded shade) clustering indicates a similar expression pattern of species between CRC and AN, whereas expression of highly distinct species or less distinct species as dark/light shaded spots, respectively.

To find the relationship between and among CRC and AN at the species level, the data represented as Venn diagram (Fig. 4) depict both common and unique OTUs at the species level. 210 species were common in both CRC and AN; 79 species were present only in CRC and not in AN, while 72 species were unique to AN.

With a cut-off of > 100 OTU counts, 6 species were significantly abundant (<0.05 from DEseq data) in CRC—*Bacteroides massiliensis*, *Alistipes* sp. *Alistipes onderdonkii*, *Bifidobacterium pseudocatenulatum*, *Corynebacterium appendicis*, and *Acidiphilium* sp. Distinct abundance signature in AN showed significant presence of *Bacillus* sp., *Veillonella atypica*, *Oribacterium* sp. from the Firmicutes phyla and *Paracoccus sphaerophysae*, *Campylobacter gracilis* and *Pasteurella pneumotropica* belonging from phyla proteobacteria (Table 3).

Further, the data comprising of significant ($p<0.05$) OTU were analyzed using DESeq2 after normalized logarithmic transformation and represented by Heatmap (Fig. 5). The analysis showed differential species abundance in CRC and AN wherein enrichment of *Ruminococcus callidus*, *Blautia producta*, *Micrococcus lylae*, *Butyrivibrio crossotus*, and *Corynebacterium_appendicis* ($p<0.05$) was found in CRC, and *Streptococcus intermedius*, *Bacteriodes ovatus*, and *Bacteroides intestinalis* ($p<0.05$) were found in AN.
et al. 2019; Dubey et al. 2018). One of the likely reasons could be the omnivorous diet of Indians that are rich in carbohydrates (Dhakan et al. 2018). This correlates with studies by Dhakan et al. and others. The majority of the literature shows microbiome studies conducted in fecal samples. This can pose a challenge as fecal matter does not capture all the gut microbes and, in particular, mucosal adherent microbes (Eckburg et al. 2005; Rivadeneira et al. 2010). Pathogenic bacteria with invasive properties that can proliferate within the anaerobic tumor and its microenvironment most likely represent the tumor-associated signature (Zmora et al. 2018). Thus, studying the microbiome of tumor tissue may help understand the progression of the disease. The present study attempted to decipher the CRC tumor-associated microbial signature.

In this study, the relative abundance of *Prevotella copri* and *Faecalibacterium prausnitzii* was found to be among the top 10 abundant species in both CRC and adjacent normal tissues. This correlates with studies by Dhakan et al. and Dubey et al. One of the likely reasons could be the omnivorous diet of Indians that are rich in carbohydrates (Dhakan et al. 2019; Dubey et al. 2018).

**CRC microbial signature**

Using metagenomics (16S rRNA NGS), this study identified the microbial composition and diversity in tumor/adjacent normal tissues pairs and uncovered microbial signatures inextricably linked to CRC. An abundance of Proteobacteria and Firmicutes phylum was observed in both CRC and AN. When the distribution of each phylum were further studied, Proteobacteria were more abundant than Firmicutes. Literature has reported a higher abundance of Proteobacteria and Fusobacteria in tumor tissues and a lower abundance of Bacteroidetes, Firmicutes, and Actinobacteria (Kim et al. 2018; Ringel et al. 2015; Shah et al. 2018; Wang et al. 2020). Colibactin produced by *E. coli* promoted colon tumorigenesis in ApcMin/+ and Il10−/− mice models, suggesting the role of Proteobacteria phyla-induced peroxisome proliferator-activated receptor-gamma (PPAR-γ) activity (Tomkovich et al. 2017). In fact, altered microbial composition in intestinal dysbiosis was observed due to the relative higher abundance of phyla Proteobacteria in the dextran sulfate sodium-induced murin colitis model (Hughes et al. 2017). Additionally, Proteobacteria-induced pro-inflammatory molecules such as lipopolysaccharides has shown to increase fat storage, and thus associated with obesity, an etiologic factor of increased risk for CRC development (Rizzatti et al. 2017). In our study, eighty percent (4 out of 5) of the enrolled CRC patients were obese.

QIIME2 analysis showed lower species richness and evenness (α-diversity) in tumor than their adjacent area, but a distinct bacterial family composition (β-diversity) in the tumor samples of CRC tissue. A study by He et al. (2021), found similar results of decreasing microbial diversity and abundance in CRC than that normal individual.

The 16S sequencing data analysis showed distinct 79 species in CRC and 72 in AN. Based on cut-off (> 100 OTU), 4 species significantly abundant (< 0.005) were *Bacteroides massiliensis, Alistipes onderdonkii, Bifidobacterium pseudocatenulatum, Corynebacterium appendicis*. *Bacteroides massiliensis* was reported to be highly abundant in CRC patients (Ozawa et al. 1979) and Ulcerative colitis patients (Wu et al. 2018). Recently, Parker et al. reviewed the emerging role of different species of *Alistipes genus*. *Alistipes species* have contributed to tumor pathogenesis via the IL-6/STAT3 pathway (Moschen et al. 2016). The study by Goetz et al. found that the protein Lipocalin 2 (LCN 2), produced by *Alistipes species*, binds to high-affinity iron-chelating compounds (siderophores), which causes reduction of iron availability (Goetz et al. 2002; Parker et al. 2020). In the 5 CRC patients in our study, hemoglobin level were extremely low. One of the reasons could be significant high abundance of *Alistipes species*.

**Adjacent normal microbial signature**

The high abundance of species from Firmicutes and Proteobacteria was observed in the AN region of the 5 CRC patients enrolled in the study (*Bacillus* sp., *Veillonella atypica, Paracoccus sphaerophysae, Campylobacter gracilis, Pasteurella pneumotropica, Oribacterium* sp.). *Bacillus*
Table 3 Abundant species in CRC and AN

| Genus/species in CRC | p value | Genus/species in AN | p value |
|----------------------|---------|---------------------|---------|
| Corynebacterium appendicis | 1.28E-05 | Pasteurella pneumotropica | 0.00027 |
| Bacteroides massiliensis | 0.00024 | Campylobacter gracilis | 0.00052 |
| Bifidobacterium pseudocatenulatum | 0.00092 | Paracoccus sphaerophysae | 0.00032 |
| Alistipes onderdonkii | 0.00164 | Veillonella atypica | 0.00144 |

Fig. 5 An OTU table heat map shows taxonomy assignment for each OTU from sample characterized by amplicon sequencing of 16S rRNA gene (V3–V4 region) within CRC and AN. The OTU heat map displays OTU counts per sample, where the counts are colored based on the contribution of each OTU to the total OTU count present in the sample (green: contributes a low percentage of OTUs to sample; red: contributes a high percentage of OTUs). Vertical clustering (hierarchical clustering) indicates the similarity in the richness of different species among different samples. The closer the Euclidean distance between two species, the shorter the branch length, indicating a more significant similarity in richness between the two species. The heat map was built using NG-CHM BUILDER. Scale on the right indicates the breakpoints associated with the colors (colour figure online)
sp. belonging to Firmicute phyla are known to be involved in gut homeostasis by promoting the growth of other beneficial microbes and suppressing pathogen/pathogen-induced inflammatory response of intestinal mucosa (Khochamit et al. 2015; Nagal et al. 1996; Ozawa et al. 1979). Similarly, Veillonella atypica ferments lactate to propionate and acetate (Mashima and Nakazawa 2014) and stimulates the growth of a wide range of organisms through metabolic complementation (Zhou et al. 2015). Oribacterium sp. belonging to Firmicutes phyla plays a significant role in repairing damaged mucous tissue, as shown by Zelante and team. The study reported that Oribacterium species binds to the host aryl-hydrocarbon receptor inducing the production of short-chain fatty acids, and thus maintains intestinal homeostasis (Zelante et al. 2013).

**Conclusion**

The mortality rate of colorectal cancer is significantly high. The major limitation is its late identification, advanced stage of the disease with late diagnosis, and resistance to treatment. With evolving metagenomics technology, the genetic mutations of cancer and the expression pattern of microbiome can play a critical role in CRC diagnosis and prediction. This preliminary study deciphered a panel of bacteria specific to CRC tumor tissue, which can enable the use of microbial biomarkers for early diagnosis of CRC. The study needs further validation in stratified groups that includes normal mucosa and adenomas.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s00203-022-02954-2.

**Acknowledgements** The authors would like to thank Parul Chugh (SGRH) for statistical help and the co-workers at Department of Research, SGRH.

**Author contributions** All authors contributed towards writing the manuscript and have given approval to the final version of the manuscript. RH performed collection of human samples/isolation/NGS assays and framing the manuscript. SB guided, framed the manuscript, inferred the data and edited the manuscript. RR performed the metagenomics data analysis and prepared the figures. SR enrolled the patients as per inclusion criteria, provided the clinical samples and demographics of the patients. DB contributed in manuscript proof reading. PN & NKS assisted in experimental standardization and NGS data acquisition and cleaning. SC conceptualized, guided, analyzed the data, framed and edited the manuscript.

**Funding** This study was supported by the grant provided by Research and Development program (RDP Grant No.: 4.9.36-011), Sir Ganga Ram Hospital (SGRH), India.

**Declarations**

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethics approval** This study was reviewed and approved by Sir Ganga Ram Hospital Ethics Committee (Ethics approval code: EC/06/16/1000). The patients/participants provided their informed consent to participate in this study.

**References**

Amin MB, Greene FL, Edge SB., et al. (2017) The Eighth Edition AJCC Cancer Staging Manual: continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging. CA Cancer J Clin 67(2):93–99. https://doi.org/10.3322/caac.21388

Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jamal A, Bray F (2017) Global patterns and trends in colorectal cancer incidence and mortality. Gut 66(4):683–691. https://doi.org/10.1136/gutjnl-2015-310912

Arnold M, Abnet CC, Neale RE, et al (2020) Global burden of 5 major types of gastrointestinal cancer. Gastroenterology 159(1):335–349. https://doi.org/10.1053/j.gastro.2020.02.068

Chen W, Liu F, Ling Z, Tong X, Xiang C (2012) Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. PLoS ONE 7(6):e39743. https://doi.org/10.1371/journal.pone.0039743

Das B, Ghosh TS, Kedia S et al (2018) Analysis of the gut microbiome of rural and urban healthy Indians living in sea level and high altitude areas. Sci Rep 8(1):10104. https://doi.org/10.1038/s41598-018-28550-3

Dhakan DB, Maji A, Sharma AK, et al (2019) The unique composition of Indian gut microbiome, gene catalogue, and associated fecal metabolome deciphered using multi-omics approaches. Gigascience. https://doi.org/10.1093/gigascience/giz004

Dubey AK, Uppadhyaya N, Nilawe P et al (2018) LogMPIE, pan-India profiling of the human gut microbiome using 16s rRNA sequencing. Sci Data 5:180232. https://doi.org/10.1038/sdata.2018.232

Eckburg PB, Bik EM, Bernstein CN et al (2005) Diversity of the human gut microbial flora. Science 308(5728):1635–1638.

Fidler MM, Soerjomataram I, Bray F (2016) A global view on cancer incidence and national levels of the human development index. Int J Cancer 139(11):2436–2446. https://doi.org/10.1002/ijc.30382

Goetz DH, Holmes MA, Borregaard N, Bluhm ME, Raymond KN, Strong RK (2002) The neutrophil lipocalin NGAL is a bacteriostatic agent that interferes with siderophore-mediated iron acquisition. Mol Cell 10(5):1033–1043. https://doi.org/10.1016/s1097-2765(02)00708-6

Gupta A, Dhakan DB, Maji A, et al (2019) Association of Flavonifractor plautii, a flavonoid-degrading bacterium, with the gut microbiome of colorectal cancer patients in India. mSystems. https://doi.org/10.1128/mSystems.00438-19

He T, Cheng X, Xing C (2021) The gut microbial diversity of colon cancer patients and the clinical significance. Bioengineered 12(1):7046–7060. https://doi.org/10.1080/21655979.2021.1972077

Helmink BA, Khan M, Hermann A, Gopalakrishnan V, Wargo JA (2019) The microbiome, cancer, and cancer therapy. Nat Med 25(3):377–388. https://doi.org/10.1038/s41591-019-0377-7
