



OPEN ACCESS

EDITED BY

Franck Carbonero,
Washington State University Health Sciences
Spokane, United States

REVIEWED BY

Tarini Shankar Ghosh,
University College Cork, Ireland
Tapas Ranjan Behera,
Cleveland Clinic, United States

*CORRESPONDENCE

Tam Thi Thanh Tran
✉ tran-thi-thanh.tam@usth.edu.vn

RECEIVED 21 February 2024

ACCEPTED 04 December 2024

PUBLISHED 24 December 2024

CITATION

Nhung PTT, Le HTT, Nguyen QH, Huyen DT,
Quyen DV, Song LH, Van Thuan T and
Tran TTT (2024) Identifying fecal microbiota
signatures of colorectal cancer in a
Vietnamese cohort.

Front. Microbiol. 15:1388740.
doi: 10.3389/fmicb.2024.1388740

COPYRIGHT

© 2024 Nhung, Le, Nguyen, Huyen, Quyen,
Song, Van Thuan and Tran. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Identifying fecal microbiota signatures of colorectal cancer in a Vietnamese cohort

Pham Thi Tuyet Nhung^{1,2}, Hang Thi Thu Le³,
Quang Huy Nguyen³, Dao Thi Huyen⁴, Dong Van Quyen^{3,5},
Le Huu Song^{2,4}, Tran Van Thuan¹ and Tam Thi Thanh Tran^{3*}

¹Hanoi Medical University, Hanoi, Vietnam, ²108 Military Central Hospital, Hanoi, Vietnam,

³Department of Life Sciences, University of Science and Technology of Hanoi, Vietnam Academy of
Science and Technology, Hanoi, Vietnam, ⁴Vietnamese-German Center for Medical Research
(VG-CARE), Hanoi, Vietnam, ⁵Molecular Microbiology Lab, Institute of Biotechnology (IBT), Vietnam
Academy of Science and Technology (VAST), Hanoi, Vietnam

Background: Colorectal cancer (CRC) is among the top three causes of global cancer mortality. In Vietnam, CRC is the third leading cause of death in women and the fourth cause of cancer mortality in men. A large number of metagenomic studies have reported the relationship between altered composition and function of the gut microbiota with CRC, but this relationship in low- and middle-income countries including Vietnam (with an estimated population of 100.3 million people in 2023, ranking 16th largest country by population in the world) is not well-explored.

Methods: We collected clinical data and fecal samples from 43 CRC patients and 44 healthy control subjects. The total community DNA of microorganisms was extracted from the fecal samples and analyzed for microbiota composition using Illumina MiSeq amplicon sequencing targeting the V3–V4 region of the 16S rRNA gene.

Results: We identified a significant difference in the overall fecal microbiota composition between CRC patients and healthy controls, and we detected several CRC-associated microbial signatures in fecal samples of Vietnamese patients with CRC, which overlapped with signatures from other countries and meta-analyses. Although patients with ($n = 8$) and without ($n = 35$) type 2 diabetes (T2D) exhibited distinct gut microbiota composition compared to healthy controls, increased relative abundances of putatively pathogenic species including *Parvimonas micra*, *Peptostreptococcus stomatis*, and *Prevotella intermedia* were consistent biomarkers for CRC. In contrast, several health-associated species were significantly depleted in CRC patients such as *Lactobacillus johnsonii* and *Bifidobacterium longum* in CRC/non-T2D patients, *Ruminococcus* species, *Bacteroides uniformis*, and *Phascolarctobacterium faecium* in CRC/T2D patients, and *Butyricicoccus pullicaecorum* in both CRC groups combined.

Conclusion: Our findings confirm alterations in gut microbiota composition in CRC in a pilot Vietnamese cohort and highlight several gut microbial taxa that may have inhibitory or driver roles in CRC. This and future studies will enable the development of cancer diagnostics and treatment strategies for CRC in Vietnam, with a focus on targeting the microbiota.

KEYWORDS

colorectal cancer, fecal microbiota, cancer signatures, Vietnamese patients, 16S rRNA amplicon sequencing

Background

Colorectal cancer (CRC) is one of the most common cancers worldwide, including in Vietnam. According to Globocan 2020, this cancer is the third most commonly diagnosed cancer and the second most lethal cancer worldwide (Sung et al., 2021). A systematic analysis of CRC from 1990 to 2017 across 195 countries revealed that the incidence and mortality rates of CRC have decreased in developed countries. This decline is attributed to the introduction of screening tests, early detection programs, and the adoption of best practices in CRC treatment and management. In contrast, the opposite trend was observed in many low-income and middle-income countries, including Vietnam (Collaborators, 2019). In Vietnam, CRC is the most common cancer of the gastrointestinal tract; the estimated number of new cancer cases in the year 2020 was more than 6,448 rectal cancer and 9,399 colon cancer, and the number of CRC deaths was reported at ~8,000 people. CRC becomes the fifth leading cause of cancer mortality in Vietnam, accounting for 9% of all new cancer cases. It is the third and fourth leading cause of cancer mortality in women and men, respectively (Sung et al., 2021). Although the majority of CRC occurs in individuals aged over 50 years, early-onset CRC (<50 years) has increased in prevalence in recent years, suggesting as-yet-unknown environmental risk factors (Stoffel and Murphy, 2020). Apart from age-related risk, environmental factors including Western dietary habits, smoking, high intake of red meat and processed meat, heavy alcohol use, weight gain and obesity, and low physical activity can increase the risk of CRC development (Collaborators, 2019). Among the environmental risk factors, the gut microbiota plays a potential role in causing sporadic CRC. The gut microbiota is a community of microbes living in the intestine and acting as a virtual organ. Gut microbes actively interact with the colonic epithelium and contribute to maintaining gut homeostasis (Tlaskalova-Hogenova et al., 2014; Gagniere et al., 2016).

Since the majority of microbes in the human alimentary tract are present in the large intestine, many researchers are focusing on the association of the gut microbiota with intestinal carcinogenesis and its potential role in cancer treatment. Numerous studies have revealed the differences in the gut microbial composition between healthy individuals and patients with CRC and described specific bacterial markers for colorectal cancer screening (Feng et al., 2015; Baxter et al., 2016; Yachida et al., 2019). Several bacterial candidates from seven genera including *Fusobacterium*, *Peptostreptococcus*, *Porphyromonas*, *Prevotella*, *Parvimonas*, *Bacteroides*, and *Gemella* are associated with CRC and can potentially be used as key markers in detecting CRC (Wong and Yu, 2019; Ternes et al., 2020). Enrichment of *Gemella*, *Peptostreptococcus*, *Fusobacterium nucleatum*, *Leptotrichia*, *Selenomonas sputigena*, and *Campylobacter rectus* was observed in tumor biopsies of 43 Vietnamese CRC patients when compared to control biopsies of 25 individuals with non-cancerous colorectal polyps (Tran et al., 2022). The major challenge of gut microbiome research in humans is the normal range of differences in the

composition of gut microbiota between study cohorts, between disease conditions, or even between individuals. A meta-analysis of 2,500 individuals with different diseases, including CRC, showed that the composition of the gut microbiome was most strongly influenced by regional factors such as country or continent. Additionally, the microbiome-disease signature for each condition varied from study to study (Ghosh et al., 2020). Therefore, in this pilot project, we expanded our understanding of the alteration of the gut microbiota in CRC in non-Western countries by specifically targeting the microbiota signatures in fecal samples of Vietnamese patients with CRC.

Methods

Study participants and clinical data collection

A total of 43 patients newly diagnosed with CRC at the 108 Military Central Hospital and 44 healthy controls from the community, who were in generally good health and showed no symptoms of CRC, were recruited for this study during the years 2022–2023. In addition, the following inclusion criteria were used for the selection of CRC and healthy subjects: aged 50–79 years, male or female, and body mass index (BMI; range: 18.5–30). If any of the following conditions applied, a participant was considered ineligible to participate in this study: complicating diseases, mental illness, evidence of a biochemical or structural abnormality of the digestive tract, irritable bowel syndrome, inflammatory bowel disease, taking any antibiotics within 4 weeks prior to recruitment, or adherence to a restricted diet.

Participants' demographic characteristics such as age, sex, BMI, health status, and medication history were collected and controlled for the data analysis. A family history of cancer and gastrointestinal and allergic disorders was also recorded. The results of blood chemistry tests and clinicopathological characteristics of the CRC patients, including tumor size, TNM (tumor, node, and metastasis) stage, tumor location, metastasis status, tumor recurrence, and survival status, were extracted from patients' medical records.

Fecal sample collection and total genomic DNA extraction

All participants were asked to self-collect fecal samples at home or at the hospital with a plastic specimen container. Fresh feces were transported to the laboratory on dry ice to prevent bacterial growth within 4 h or were kept in cold storage (~4°C) for 24–48 h before arriving at the laboratory. Upon arrival at the laboratory, the fecal samples were subsampled into aliquots to prevent further unnecessary freeze-thaw cycles and were frozen for long-term storage at –80°C.

Total genomic DNA was extracted from the fecal samples using the Qiagen DNeasy® PowerSoil® Pro Kit (QIAGEN, Germany) with a bead-beating step. Briefly, 0.25 g of fecal samples were homogenized mechanically in two intervals (10 min at a speed of 25 Hz) in a PowerBead Pro Tube containing CD1 solution and beads using a TissueLyser II (QIAGEN, Germany). The subsequent

Abbreviations: CRC, colorectal cancer; CEA, carcinoembryonic antigen; BMI, body mass index; SCFAs, short-chain fatty acids; T2D, type 2 diabetes; TNM, tumor, node, and metastasis.

steps of genomic DNA extraction were then performed according to the manufacturer's extraction instructions. Genomic DNA was measured using the NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific, USA). The genomic DNA was kept at -20°C for further use.

Amplification and sequencing of the bacterial 16S rRNA gene

The extracted genomic DNA was used for PCR amplification with a bacterial primer pair (5'-CCTACGGRRBGCASCAGKVRVGAAT-3' and 5'-GGACTACNVGGGTWTCTAATCC-3') for the 16S rRNA gene V3–V4 hypervariable region. The PCR master mix contained 1 μM of each forward and reverse primer, 2 μl of dNTPs, 2.5 μl of TransStart buffer, 0.5 μL of TransStart Taq DNA polymerase, 20 ng of the extracted DNA, and 25 μl of nuclease-free water. The PCR program was as follows: 1. 94°C for 3 min, 2. 95°C for 5 s, 3. 57°C for 90 s, 4. 72°C for 10 s, and 5. 72°C for 5 min, with steps 2–4 repeated for 24 cycles. The amplicons from each sample were then incorporated with indexed adapters using limited cycle PCR and were purified using magnetic beads. The purified 16S amplicon libraries were quantified using the Tecan Infinite 200 Pro microplate reader and were pooled together at equimolar concentration. The 250 bp paired-end sequencing was processed on the pooled libraries using the Illumina MiSeq Platform (Illumina, USA) at Genewiz, Inc. (South Plainfield, NJ, USA). Raw 16S rRNA gene sequencing data is available at the NCBI Sequence Read Archive (SRA), under BioProject PRJNA1077687.

Bioinformatics analysis of the 16S amplicon data

Analysis of 16S amplicon sequencing data was carried out as in our previous study (Tran et al., 2019). Briefly, after merging the 2x250 bp paired-end reads with FLASH v1.2.8 (Magoc and Salzberg, 2011), the merged reads were trimmed forward primers and reverse primers with cutadapt v1.8.3 (Martin, 2011) and truncate_reverse_primer.py script from Quantitative Insights Into Microbial Ecology (QIIME) v1.9.1 (Caporaso et al., 2010). The QIIME's split_libraries_fastq.py script was used to filter out low-quality reads with the quality threshold at Q19 and concatenate reads from different samples into one file. Next, the clustering of reads into operational taxonomic units (OTUs) was processed using USEARCH v8.1 software (Edgar, 2010). The reads were pre-processed using usearch -derep_fulllength, sortbylength, and sortysize. A default threshold of 97% sequence identity was set for usearch -cluster_otus to form OTU representative sequences. Chimeras were eliminated from the OTU representative sequences using the usearch -uchime_ref with the ChimeraSlayer reference database. The usearch -usearch_global was applied to map all reads to the OTU representative sequences with an identity threshold of 97% to create the OTU table. Taxonomy assignments were obtained using mothur v1.36.1 (Schloss et al., 2009) with the Ribosomal Database Project (RDP) trainset 18 (Cole et al., 2009) and SPINGO v1.3 (Allard et al., 2015) for species classification. Alpha-diversity and beta-diversity were computed using QIIME's alpha_rarefaction.py and beta_diversity.py scripts on a rarefied OTU table at the minimum OTU reads across all samples as 70,034 reads per sample.

TABLE 1 Descriptive statistics of CRC patients and healthy controls.

Characteristic	Categories	CRC patients (n = 43)	Healthy controls (n = 44)	P-value
Age at diagnosis	Mean \pm SD (years)	64 \pm 6	61 \pm 6	0.046^a
	50–60, no. (%)	15 (34.9%)	15 (34.1%)	0.051 ^b
	60–70, no. (%)	16 (30.8%)	25 (56.8%)	
	70–80, no. (%)	12 (23.1%)	4 (9.1%)	
BMI	Mean \pm SD (kg/m^2)	22.7 \pm 2.2	23.2 \pm 1.8	0.142 ^a
	Normal weight, no. (%)	35 (67.3%)	39 (88.6%)	0.383 ^c
	Overweight, no. (%)	8 (15.4%)	5 (11.4%)	
Sex				
	Male, no. (%)	24 (46.2%)	24 (54.5%)	0.394 ^c
	Female, no. (%)	19 (36.5%)	20 (45.5%)	
Comorbidity				
	T2D, no. (%)	8 (15.4%)	0 (0%)	0.002^c
	Hypertension, no. (%)	16 (30.8%)	6 (13.6%)	0.014^c
Family history of CRC, no. (%)		2 (4.7%)	0 (0%)	0.241 ^c

Age, weight, height, and BMI are presented with mean \pm standard deviation (SD).

^aMann–Whitney U-test.

^bChi-squared test.

^cFisher's exact test.

Statistically significant results (p < 0.05) are shown in bold.

Statistical analysis

Statistical analysis was performed using R v4.2.1 software ([R Core Team, 2016](#)). The significant differences in clinical measures and alpha-diversity indices were determined using the Mann-Whitney *U*-test for comparison of two independent groups and the Kruskal-Wallis test for comparison of more than two groups. Comparisons of categorical variables were calculated using Fisher's exact test for two groups and chi-squared test for >2 groups. Principal coordinate analysis (PCoA) was used to visualize the similarities in beta diversity between the samples. Permutational multivariate analysis of variance (PERMANOVA) tests were conducted to assess the significant differences in beta diversity between the groups. The differential taxonomic abundances between CRC patients and healthy individuals were calculated using DESeq2 ([Love et al., 2014](#)), controlling for age, sex, BMI, and hypertension. A *p*-value adjusted using the Benjamini-Hochberg correction, <0.05 , was considered statistically significant. The Spearman correlation between the relative abundance of OTUs and the clinical traits was conducted using the cor.test function in R.

Results

Clinical characteristics of study participants

The clinical descriptors of the CRC patients and healthy controls enrolled in this study are presented in [Table 1](#). We included 43 CRC patients and 44 healthy controls, with an average age of 64 years (age range, 53–76 years) for CRC patients and an average of 61 years (age range, 51–75 years) for healthy controls. The ages were divided into three groups; the majority of healthy controls (56.8%) were 60–70 years old, while the patients with CRC were 50–60 and 60–70 years old, accounting for 65.7%. There was no significant difference between CRC patients and healthy controls with respect to BMI, sex, or family history of CRC (Mann-Whitney *U*-test, *p*-value > 0.05). A high prevalence of T2D and hypertension, which are common comorbidities in people aged over 50, was recorded in CRC patients compared to healthy individuals.

All patients with CRC suffered from abdominal pain. The coexistence of other symptoms included rectal bleeding, diarrhea, and/or weight loss. Among the 43 patients, five patients (11.6%) had either liver metastasis or peritoneal metastasis, while two patients (4.7%) experienced multiple organ metastases. A total of 62.5% of tumors in CRC patients were reported as 2–5 cm in size and 39.5% were 5–10 cm; all tumors were resectable. The majority of the tumors (46.5%) occurred in the left colon, followed by the right colon, the rectum, and the transverse colon, accounting for 23.3, 20.9, and 9.3%, respectively. According to the TNM classification, the recruited patients ranged from early to advanced stages, with the majority (72.1%) being in stages II and III. Cancer recurred in 11.6% of the patients, and three patients (7%) died within 2 years of diagnosis ([Table 2](#)). In addition to clinical features, blood biochemical characteristics were investigated in CRC patients. The mean values of most parameters in both male and female CRC patients were within the normal range. This indicates that CRC itself was not the cause of the blood test changes. On the contrary, carcinoembryonic antigen (CEA) level, a prognostic marker for

TABLE 2 Clinicopathological characteristics of CRC patients.

Characteristic	Categories	CRC patients (<i>n</i> = 43)
Tumor size (cm)	Median (Range)	4.5 (2–10)
	2–5 cm	26 (60.5%)
	5–10 cm	17 (39.5%)
TNM stage*, no. (%)	Stage I	7 (16.3%)
	Stage II	13 (30.2%)
	Stage III	18 (41.9%)
	Stage IV	5 (11.6%)
Tumor location, no. (%)		
	Left colon	20 (46.5%)
	Right colon	10 (23.3%)
	Transverse colon	4 (9.3%)
	Rectum	9 (20.9%)
Metastasis, no. (%)		
	One organ	5 (11.6%)
	>1 organ	2 (4.7%)
Tumor recurrence, no. (%)		5 (11.6%)
Survival status, no. (%)		
	Living	40 (93%)
	Dead	3 (7%)

*TNM stage: Tumor, node, and metastasis (TNM) stage.

CRC, was elevated in the serum of 21 patients (48.8%), especially patients with metastases ([Supplementary Table 1](#)).

Gut microbiota diversity in CRC patients and healthy controls

Total genomic DNA was extracted from fecal samples of individual participants and underwent 16S rRNA gene amplicon sequencing for the identification of bacterial diversity and composition. We observed a significant increase in alpha diversity measured using the Chao1, Shannon's, Simpson's, and phylogenetic diversity indices in Vietnamese patients with CRC compared to healthy controls (Mann-Whitney *U*-test, *p*-value < 0.05 ; [Figure 1A](#)). The alpha-diversity indices did not differ by the participant's age category (50–60, 60–70 vs. 70–80 years), BMI category (normal weight vs. overweight), sex (male vs. female), and blood pressure (hypertension vs. non-hypertension), although the Simpson's index was slightly lower in the participants with T2D (median = 0.93; IQR = 0.90–0.94) than the participants without T2D (median = 0.94; IQR = 0.93–0.96; *p*-value = 0.098). Furthermore, alpha diversity in the CRC group was not influenced by any clinical traits, except for the BMI category and T2D status. In contrast, the Chao1 index differed among age categories in the healthy control group ([Supplementary Table 2](#)).

Upon assessment of the overall gut microbiota composition, beta diversity represented by unweighted UniFrac distance

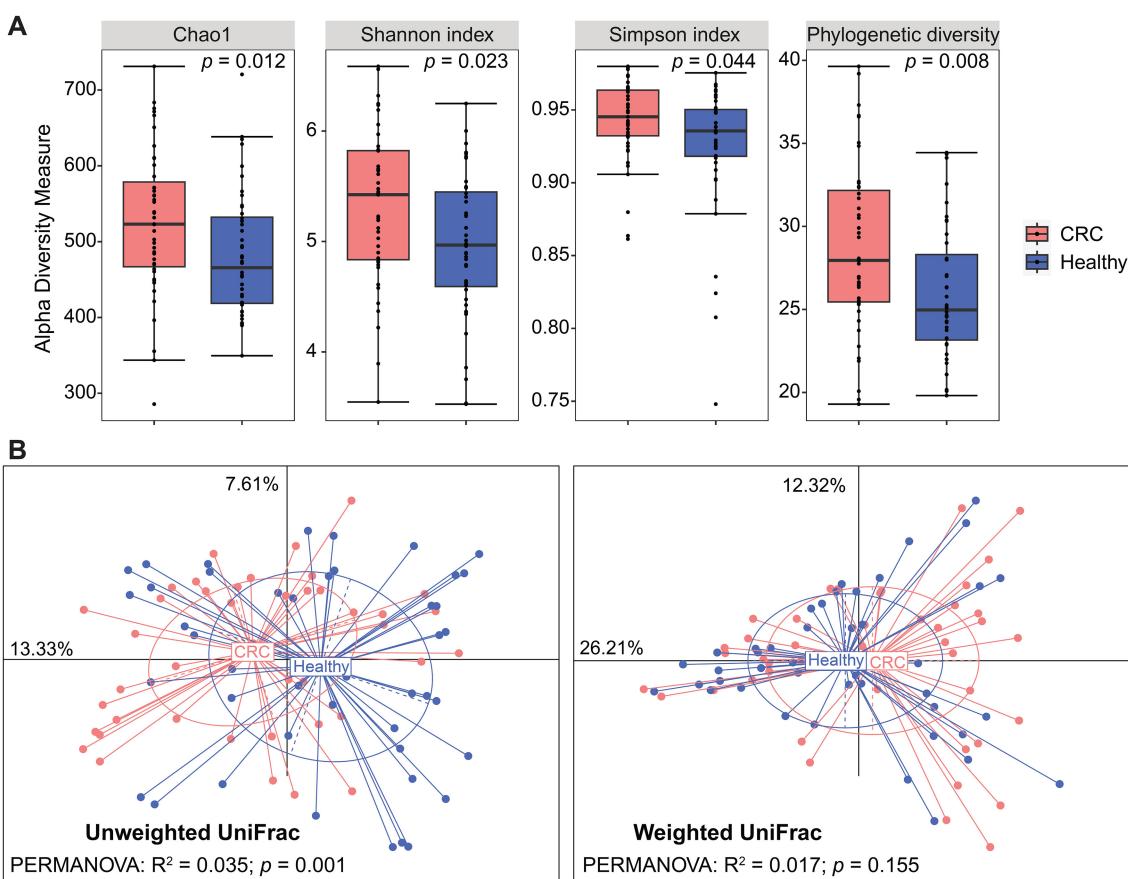


FIGURE 1

Diversity of the gut bacterial communities differs between CRC patients and healthy controls. **(A)** Boxplots of four alpha-diversity indices visualize the difference in the species richness and evenness of CRC patients compared to healthy controls. *P*-values were determined using the Mann-Whitney *U*-test. **(B)** Principal coordinate analysis (PCoA) plots based on the unweighted and weighted UniFrac distances. Permutational multivariate analysis of variance (PERMANOVA) tests were conducted to assess significant differences between CRC and healthy groups.

revealed two fairly distinct clusters based on CRC condition (PERMANOVA: $R^2 = 0.035$; $p = 0.001$), whereas no significance was detected between CRC patients and healthy controls based on the weighted UniFrac distance (PERMANOVA: $R^2 = 0.017$; $p = 0.155$; Figure 1B). In addition, there were no significant differences by the participant's age category, BMI category, sex, T2D condition, or blood pressure in either UniFrac distance. However, unweighted and weighted UniFrac distances were discriminatory in terms of BMI categories in the CRC cohort, with a low coefficient of determination (PERMANOVA: $R^2 = 0.037$ and 0.058, respectively; $p < 0.05$ for both; Supplementary Figure 1A and Table 3). Within the healthy subjects' samples, there was a significant separation among age categories based on the weighted UniFrac distance (PERMANOVA: $R^2 = 0.098$; $p = 0.01$), and sex had a relevant impact on the weighted UniFrac distance (PERMANOVA: $R^2 = 0.043$; $p = 0.015$; Supplementary Figure 1B and Table 3).

Intestinal microbial associations with colorectal cancer

We observed high variability in microbiome composition at the family and genus levels among individuals in both

the CRC and control groups. The most dominant families in the participants' gut were *Lachnospiraceae*, *Oscillospiraceae*, *Bacteroidaceae*, *Enterobacteriaceae*, and *Peptostreptococcaceae*, with a mean relative abundance of 74.6% (range = 44.01–82.18%, SD = 11.24; Supplementary Figure 2A). At the genus level, the relative abundance means of the top five genera were 16.81% (SD = 10.63) for *Blautia*, 5.69% (SD = 4.83) for *Faecalibacterium*, 4.68% (SD = 6.22) for *Phocaeicola*, 4.21% (SD = 7.46) for *Escherichia/Shigella*, and 4% (SD = 5.05) for *Bacteroides* (Supplementary Figure 2B). Next, we identified 6 classified families and 18 classified genera that significantly correlated with the CRC condition using DESeq2, controlling for age, sex, BMI, and hypertension (Figure 2A). Compared to the healthy individuals, *Porphyromonadaceae*, *Turicibacteraceae*, *Eubacteriales incertae sedis XIII*, *Akkermansiaceae*, and *Peptoniphilaceae* were found to be increased in CRC samples, while *Micrococcaceae* was negatively associated with CRC condition. Similarly, the genera *Agathobacter*, *Parasutterella*, *Citrobacter*, *Muribaculum*, *Butyrivibrio*, and *Lactobacillus* were underrepresented in the CRC patients. On the contrary, 12 out of 18 genera were positively associated with CRC, with the top genera showing the highest log₂ fold changes in abundance in CRC compared to healthy controls, namely *Mogibacterium*, *Porphyromonas*, *Peptostreptococcus*, *Prevotella*,

TABLE 3 Association between beta-diversity and clinical characteristics.

	Unweighted UniFrac		Weighted UniFrac	
	R ²	p	R ²	p
All samples (CRC and healthy samples)				
<i>CRC condition:</i> CRC vs. healthy	0.035	0.001	0.017	0.155
<i>Age category:</i> 50–60, 60–70 vs. 70–80	0.025	0.267	0.026	0.292
<i>BMI category:</i> normal weight vs. overweight	0.014	0.152	0.009	0.617
<i>Sex:</i> male vs. female	0.015	0.103	0.011	0.462
<i>T2D condition:</i> T2D vs. non-T2D	0.012	0.371	0.012	0.362
<i>Blood pressure:</i> hypertension vs. non-hypertension	0.011	0.539	0.006	0.897
CRC samples				
<i>Age category:</i> 50–60, 60–70 vs. 70–80	0.045	0.580	0.066	0.121
<i>BMI category:</i> normal weight vs. overweight	0.037	0.032	0.058	0.017
<i>Sex:</i> male vs. female	0.020	0.744	0.014	0.821
<i>T2D condition:</i> T2D vs. non-T2D	0.027	0.248	0.032	0.198
<i>Blood pressure:</i> hypertension vs. non-hypertension	0.020	0.722	0.006	0.994
<i>Tumor size:</i> 2–5 cm vs. 5–10 cm	0.024	0.428	0.019	0.626
<i>TNM stage:</i> four stages	0.066	0.700	0.074	0.413
<i>Tumor location:</i> four locations	0.069	0.562	0.062	0.640
<i>Metastasis:</i> yes vs. no	0.021	0.661	0.023	0.426
<i>Tumor recurrence:</i> yes vs. no	0.019	0.810	0.012	0.906
<i>Survival status:</i> living vs. dead	0.021	0.630	0.014	0.821
<i>CEA:</i> CEA-normal vs. CEA-elevated	0.022	0.624	0.025	0.381
Healthy samples				
<i>Age category:</i> 50–60, 60–70 vs. 70–80	0.054	0.176	0.098	0.010
<i>BMI category:</i> normal weight vs. overweight	0.021	0.645	0.035	0.144
<i>Sex:</i> male vs. female	0.043	0.015	0.019	0.627
<i>Blood pressure:</i> hypertension vs. non-hypertension	0.025	0.327	0.028	0.261

R² and p-values were obtained from permutational multivariate analysis of variance (PERMANOVA) tests. Statistically significant (p < 0.05) is shown in bold.

and *Parvimonas* (log₂ fold change > 2.5, adjusted p-value < 0.001; [Figure 2A](#) and [Supplementary Table 3](#)).

A comparison of bacterial species detected a lower abundance of three species and a higher abundance of 13 species in the CRC patients than in the healthy controls ([Figure 2](#) and [Supplementary Table 4](#)). Among those, *Parvimonas micra*, *Peptostreptococcus stomatis*, and *Prevotella intermedia* showed the greatest variance between the CRC and healthy groups (log₂ fold change > 4, adjusted p-value < 0.001). In addition, we found significant enrichment of three *Bacteroides* (*B.*) species (*B. caccae*, *B. celulosilyticus*, and *B. salyersiae*) and three *Parabacteroides* (*P.*) species (*P. distasonis*, *P. merdae*, and *P. goldsteinii*) in patients with CRC. On the contrary, CRC patients had lower abundance levels of two health-associated species including *Lactobacillus johnsonii* and *Butyrivibrio pullicaecorum* than the healthy individuals (log₂ fold change < -2, adjusted p-value < 0.05).

Impact of type 2 diabetes on fecal microbiota signatures of colorectal cancer

T2D is a common comorbidity in patients with CRC, and this metabolic disease has been reported to be associated with alteration in the gut microbiota community ([Umirah et al., 2021](#)). Thus, we further evaluated CRC-related gut microbiota signatures in the presence of T2D. A comparison between CRC patients without T2D (*n* = 35) and those with T2D (*n* = 8) revealed significant differences in 17 genera and 15 species ([Figure 3A](#)). Among the significantly associated species, the top six species that showed the most significant fold change between the two groups were *Ruminococcus flavefaciens*, *Peptococcus niger*, *Akkermansia muciniphila*, *Butyrivibrio pullicaecorum*, *Phascolarctobacterium faecium*, and *Porphyromonas endodontalis* (log₂ fold change > 3.9, adjusted p-value < 0.05; [Figure 3A](#) and [Supplementary Figure 3](#)).

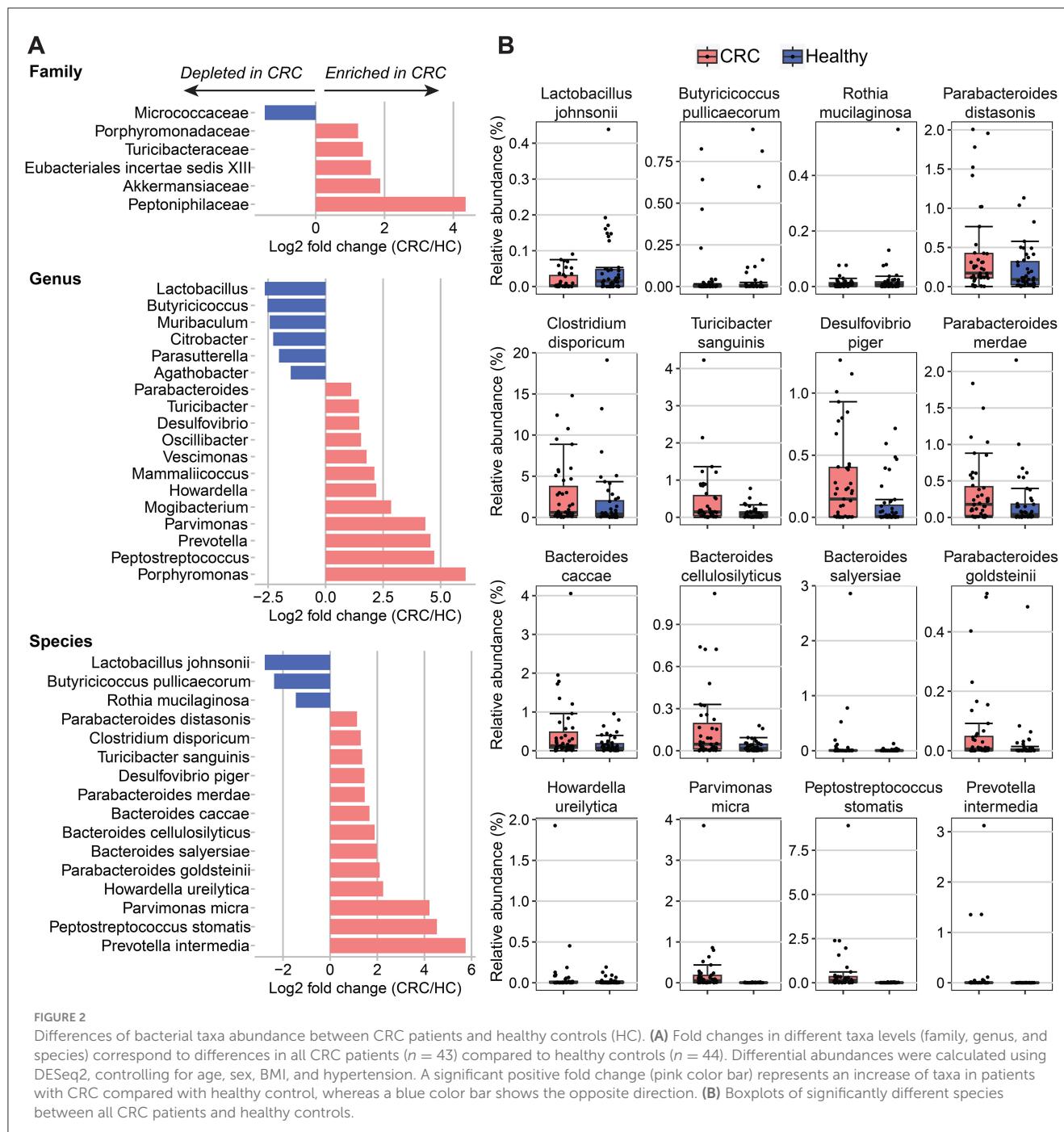
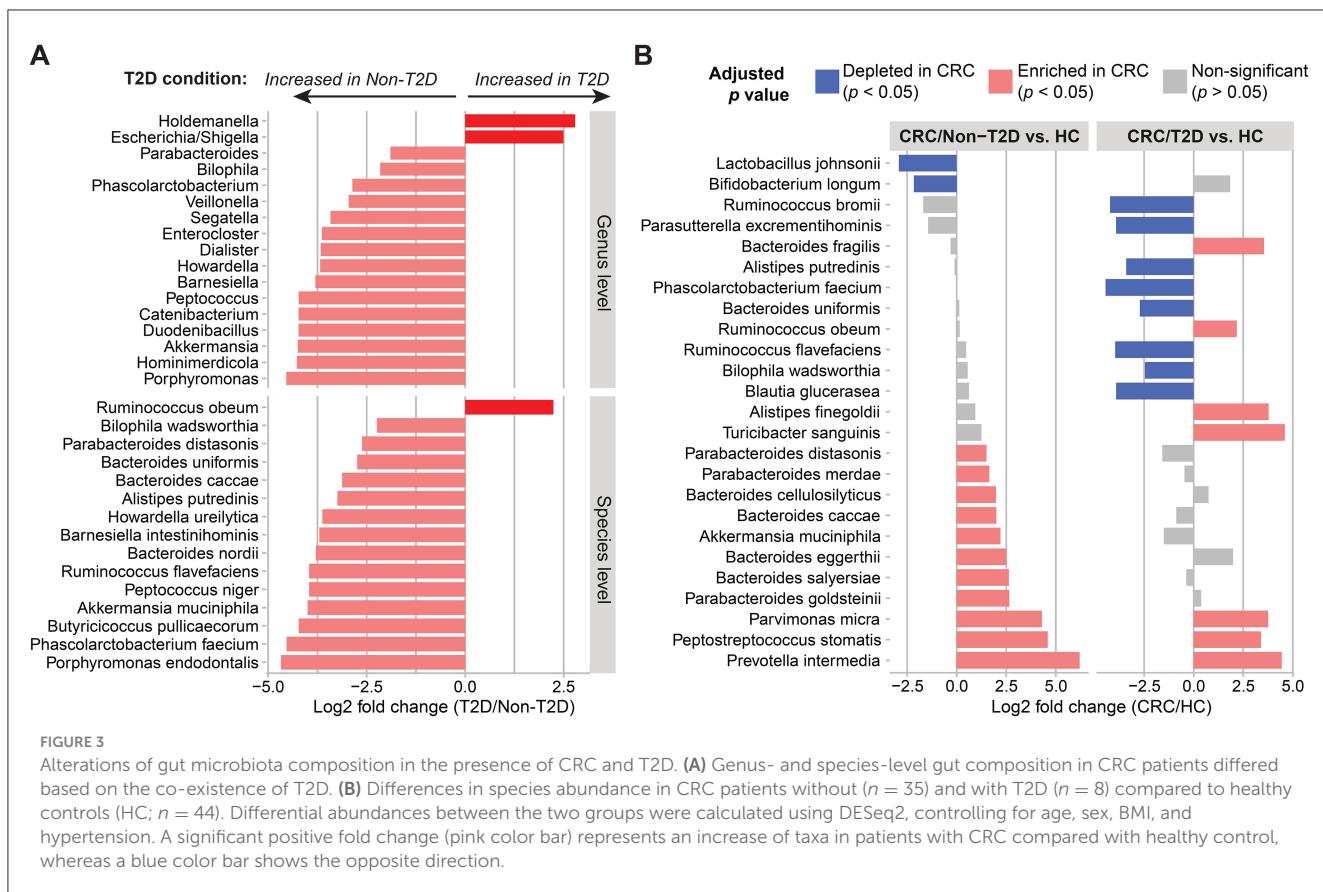


FIGURE 2

Differences of bacterial taxa abundance between CRC patients and healthy controls (HC). **(A)** Fold changes in different taxa levels (family, genus, and species) correspond to differences in all CRC patients ($n = 43$) compared to healthy controls ($n = 44$). Differential abundances were calculated using DESeq2, controlling for age, sex, BMI, and hypertension. A significant positive fold change (pink color bar) represents an increase of taxa in patients with CRC compared with healthy control, whereas a blue color bar shows the opposite direction. **(B)** Boxplots of significantly different species between all CRC patients and healthy controls.

Next, in comparison to the healthy control group, the families *Turicibacteraceae* and *Peptoniphilaceae* and six genera including *Turicibacter*, *Mogibacterium*, *Parvimonas*, *Prevotella*, *Peptostreptococcus*, and *Porphyromonas* were consistently enriched in CRC patients with and without T2D (log₂ fold change > 1, adjusted p -value < 0.05; [Supplementary Figure 4](#)). At the species level, only *Parvimonas micra*, *Peptostreptococcus stomatis*, and *Prevotella intermedia* were associated with CRC patients with either non-T2D or T2D. On the other hand, the potential health-associated species were not shared between CRC/non-T2D and CRC/T2D. For example, the two species *Lactobacillus johnsonii*

and *Bifidobacterium longum* were significantly increased in healthy individuals when compared to CRC/non-T2D. Meanwhile, the combination of CRC and T2D caused the change in the addition of species, increasing *Turicibacter sanguinis*, *Alistipes finegoldii*, *Bacteroides fragilis*, and *Ruminococcus obeum* and reducing *Phascolarctobacterium faecium*, *Ruminococcus bromii*, *Ruminococcus flavefaciens*, *Parasutterella excrementihominis*, *Blautia glucerasea*, *Alistipes putredinis*, *Bacteroides uniformis*, and *Bilophila wadsworthia* ([Figure 3B](#) and [Supplementary Table 4](#)). However, a larger sample size for the CRC/T2D group would be needed to confirm this finding. Together, these results suggested



that both CRC and T2D contributed significantly to the alterations of the gut microbiota composition.

Correlation of gut microbial taxa and clinical traits in CRC patients

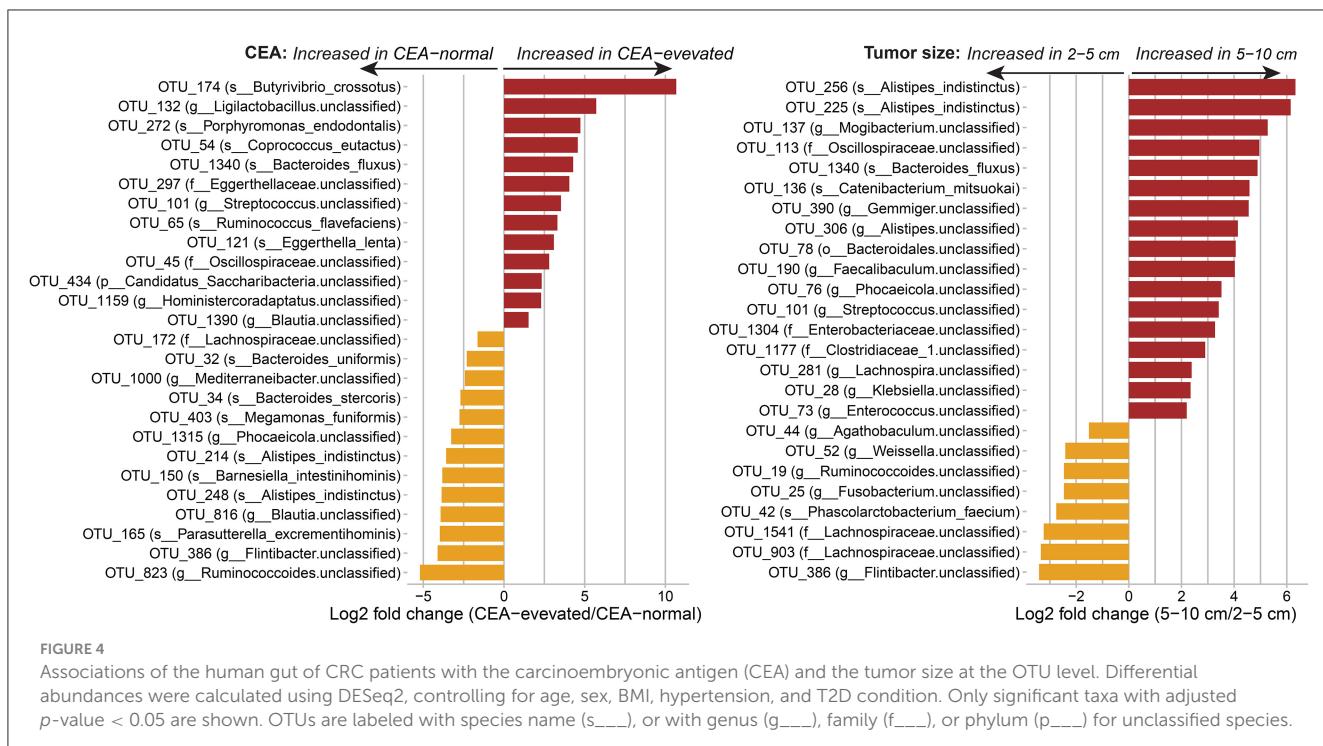
We examined whether the composition of the human gut of CRC patients differed based on clinical traits such as CEA level and tumor size. Since the T2D condition was found to be correlated with changes in the gut composition in CRC patients, we adjusted for the presence of T2D along with age, sex, BMI, and hypertension when conducting the differential abundance analysis of the gut microbiota for the level of CEA and the tumor size using DESeq2. Overall, 26 OTUs showed significant variation between the CEA-normal and CEA-elevated groups, while 25 OTUs were significantly associated with the tumor size (Figure 4). The CEA level and the tumor size shared three common microbial OTUs: OTU_1340, belonging to the species *Bacteroides fluxus*; OTU_101 and OTU_386, belonging to the genus *Streptococcus* and the genus *Flintibacter*, respectively.

Next, the Spearman correlation analysis in the CRC subgroups (CRC/non-T2D and CRC/T2D) revealed the differences in OTUs correlated with the CEA level and the tumor size (Supplementary Figure 5 and Supplementary Table 5). We observed strong correlations between microbial OTUs and the CEA level and the tumor size in the CRC/T2D group (Spearman's

correlation coefficient, $0.7 < |\rho| < 0.9$; nominal p -values < 0.05); meanwhile, moderate relationships were shown in the CRC/non-T2D group (Spearman's correlation coefficient, $0.3 < |\rho| < 0.6$; nominal p -values < 0.05). Specifically, the relative abundances of several OTUs belonging to the *Lachnospiraceae* were inversely correlated with tumor size in both CRC/non-T2D (OTU_1275, OTU_1278, and OTU_1541) and CRC/T2D (OTU_83, OTU_176, OTU_385, OTU_1104, OTU_1198, and OTU_1541). Furthermore, the OTUs that correlated with either the CEA level or the tumor size were separated into three distinct clusters by the hierarchical cluster analysis. Cluster 1 was composed of five out of a total of 9 OTUs that were positively associated with the tumor size in CRC/T2D samples, whereas the majority of negative associations with CEA level and tumor size were observed in cluster 2 and cluster 3 in both the CRC subgroups, except for the association between the CEA level and its correlated OTUs in the CRC/T2D group. However, the microbial abundance heatmap of the correlated OTUs showed no clear separation of CRC samples according to TNM stage, tumor location, and metastasis (Supplementary Figure 5).

Discussion

We present the fecal microbiota profiles from Vietnamese patients diagnosed with CRC compared to healthy individuals using 16S rRNA amplicon sequencing. We also extended the investigation between the gut microbiota profiles and various



clinical characteristics including age, BMI, sex, comorbidities (T2D and hypertension), and clinicopathological characteristics in the CRC group. Compared to other clinical characteristics, the CRC condition is the main factor that drives the divergence of the gut microbiota diversity and composition in this study cohort.

Increased alpha diversity, species richness, and evenness were observed in CRC patients compared to healthy controls. Several previous studies reported a significantly lower alpha diversity related to CRC patients (Ahn et al., 2013; Ai et al., 2019), but contradictory results were also observed (Young et al., 2021). We hypothesized that the presence of multiple pathogenic bacteria in our CRC patients may contribute to higher alpha diversity. Further research in a large sample size is needed to identify precisely whether alpha diversity is negatively or positively correlated with CRC. In the present study, CRC condition along with BMI, age, and sex was found to be associated with the variability of the overall gut microbiota composition, in line with previous reports showing divergence of gut microbial profiles depending on host clinical features (Ghosh et al., 2020). Therefore, adjustments for potential confounders such as age, sex, BMI, and hypertension were implemented when comparing different microbial taxa between CRC patients and healthy controls.

A comparison of our findings with several published studies across various cohorts revealed both global and regional CRC-associated taxonomic signatures at the species level (Supplementary Table 6). In particular, we observed the over-representation of a set of global microbial signatures including *Parvimonas micra*, *Peptostreptococcus stomatis*, and *Prevotella intermedia* species in the CRC patients with and without T2D in comparison with the healthy individuals in our study. The presence of *Parvimonas micra* is associated with the upregulation of miR-218-5p, which inhibits the expression of protein tyrosine

phosphatase receptor, thereby activating the Ras/ERK/c-Fos pathway and promoting CRC cell proliferation (Chang et al., 2023). Members of the genus *Peptostreptococcus*, *Peptostreptococcus stomatis* are two of the four universal fecal microbial signatures for CRC that have been found in Chinese, Danish, Austrian and French cohorts (Yu et al., 2017). *Prevotella intermedia* has been described to be more abundant in colorectal adenocarcinoma tissues as well as in the fecal samples of CRC patients, which may promote migration and invasion of CRC tumor cells (Avuthu and Guda, 2022; Lo et al., 2022; Tito et al., 2024). In addition, the relative dominance of *Bacteroides* and *Parabacteroides* species (*B. caccae*, *B. cellulosilyticus*, *B. salyersiae*, *P. distasonis*, *P. merdae*, and *P. goldsteinii*) was found in the CRC patients without T2D, which partly resembled previous findings (Feng et al., 2015; Obon-Santacana et al., 2022; Iadsee et al., 2023). These *Parabacteroides* species could be regarded as regional-specific microbial signatures, as they were present in only a few Asian cohorts. For example, *Parabacteroides distasonis* was observed in the Vietnamese cohort (our study) as well as in Thai and Indian cohorts (Gupta et al., 2019; Iadsee et al., 2023). Meanwhile, discrepancies in *Bacteroides* species except *Bacteroides fragilis* were noted across different study cohorts (Supplementary Table 6), suggesting that these signatures may represent country-specific microbial signatures. Both beneficial and pathogenic roles are reported for *Bacteroides* and *Parabacteroides* species (Ezeji et al., 2021; Zafar and Saier, 2021). For example, colonic mucus-degrading bacteria such as *B. caccae* were positively associated with a low-fiber diet, and they apparently obtain an energy source by degrading the intestinal mucus barrier, which could lead to increased pathogen susceptibility (Desai et al., 2016). Similarly, even though *Parabacteroides* species such as *Parabacteroides goldsteinii* are considered novel next-generation probiotics (Wu et al., 2019), they

can act as opportunistic pathogens that are positively correlated with the early stage of Crohn's disease and Lynch syndrome (Mori et al., 2019; Ma et al., 2022). *Parabacteroides distasonis* has been shown to exert a protective role by suppressing colon tumorigenesis and maintaining the intestinal epithelial barrier in a murine model (Koh et al., 2020). Since this species was one of the CRC-enriched species in the CRC/non-T2D patients, it was not considered for anti-cancer properties in this study. Nevertheless, we observed an over-representation of this species in the CRC patients without T2D compared to the CRC patients with T2D, suggesting a protective effect against T2D in CRC patients. Among the enriched species that were reported in the CRC/T2D patients compared to the healthy group, we also found a *Bacteroides* species, *Bacteroides fragilis*. The association between *Bacteroides fragilis* and the induction of interleukin-17, which leads to colonic tumor formation, has been reported in a mouse model (Wu et al., 2009). In our study cohort, we did not detect the prevalence of the opportunistic pathogen *Fusobacterium nucleatum* in either CRC patients with or without T2D. This species is widely regarded as a biomarker for CRC in many cohorts as well as in another Vietnamese cohort study (Tran et al., 2022). We acknowledge that the study by Tran et al. (2022) and the current study differed in several aspects, including the CRC patient cohorts from two different regions of Vietnam, sample collection times, sample types, genomic DNA extraction, and sequencing platforms. Thus, additional validation studies are required to confirm specific fecal microbial biomarkers for the early screening of CRC.

In contrast, a decreasing trend was found for some health-associated species including *Lactobacillus johnsonii* and *Bifidobacterium longum* in the CRC/non-T2D samples and *Butyricoccus pullicaeorum* in both CRC patients with/without T2D combined, which highlight the potential importance of health-associated species for human gut health. Indeed, strains of *Lactobacillus johnsonii*, recognized as a potential probiotic of the genus *Lactobacillus*, have been shown to have multiple beneficial effects such as modulating innate immune response and enhancing gut barrier integrity (Zheng et al., 2021; Zhang et al., 2023). The anti-CRC potential of the butyrate-producing *Butyricoccus pullicaeorum* and the acetate-producing *Bifidobacterium longum* has been clinically demonstrated (Chang et al., 2020; Liang et al., 2023). Short-chain fatty acids (SCFAs) such as acetate and butyrate produced by those bacteria are reported to reinforce the intestinal defense of host epithelial cells and to regulate the immune response by a process involving regulatory T cells and IL-10-producing T cells (Fukuda et al., 2011; Singh et al., 2014). Unlike the gut microbiota profiles of CRC patients without T2D, *Phascolarctobacterium faecium*, *Ruminococcus bromii*, *Ruminococcus flavefaciens*, *Parasutterella excrementihominis*, *Blautia glucerasea*, *Alistipes putredinis*, *Bacteroides uniformis*, and *Bilophila wadsworthia* were depleted in CRC/T2D patients. Previous research presented a significant decrease in the genera *Bacteroides* and *Parasutterella* in CRC patients compared to healthy controls (Wang et al., 2012). In addition, Li L. et al. reported that *Bacteroides uniformis* and *Phascolarctobacterium faecium* were negatively associated with the T2D phenotype and were considered to play a

pivotal role in the regulation of insulin secretion and glucose metabolism (Li et al., 2020). The genus *Ruminococcus*, including *Ruminococcus bromii* and *Ruminococcus flavefaciens*, is widely known for the degradation of resistant starch, producing SCFAs (Chassard et al., 2010; Ze et al., 2012).

While the current study provides insight into the association between gut microbiota profiles and CRC patients in Vietnamese cohorts, there are some limitations. First, the study cohort enrolled participants in north and north-central coastal regions, which does not represent gut microbiota profiles for all the different geographical regions of Vietnam. It would be interesting to investigate if any fecal microbiota biomarkers for CRC are common or unique to specific regions in Vietnam. Second, on average, the age of the healthy control group was lower than that of the CRC group because the study participants, especially the healthy individuals, were randomly selected within a wide age range (aged 50–79 years) and younger people were more likely associated with better health. To rectify this, differences in the gut microbial taxa were adjusted for the age factor for all analyses by linear models implemented in DeSeq2. Third, although a modest sample size was recruited for this pilot study, statistical power remained inadequate when studying correlations between the gut microbial taxa and clinical traits in CRC patients. Finally, we have not investigated the association between CRC and the functional capacity of the gut microbiota due to the limitations of 16S rRNA gene amplicon sequencing. A larger study in both gut microbiota composition and functional profiles, ideally sampling from multiple regions that capture gut microbiota diversity across different geographical regions of Vietnam, will ultimately provide a comprehensive picture for fully understanding the dynamic interaction of the gut microbiota and clinical characteristics of CRC patients.

Conclusion

The current study provided new insights into the alterations of the gut microbiota composition and fecal microbial signatures associated with Vietnamese patients diagnosed with CRC. These findings lay the foundations for the application of next-generation sequencing technology in early CRC detection and the use of personalized microbiota-targeted therapy for the prevention and treatment of CRC in Vietnam.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary material](#).

Ethics statement

The studies involving humans were approved by Hanoi Medical University Institutional Review Board, Hanoi Medical University, Hanoi, Vietnam. The studies were conducted in accordance

with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

PTTN: Conceptualization, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. HTTL: Methodology, Writing – review & editing. QHN: Methodology, Writing – review & editing. DTH: Methodology, Writing – review & editing. DVQ: Writing – original draft, Writing – review & editing. LHS: Methodology, Writing – original draft, Writing – review & editing. TTV: Writing – original draft, Writing – review & editing. TTTT: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 108.04-2021.22. This study was supported by the 108 Military Central Hospital, the Vietnamese-German Center for Medical Research (VG-CARE), and the University of Science and Technology of Hanoi.

References

- Ahn, J., Sinha, R., Pei, Z., Dominian, C., Wu, J., Shi, J., et al. (2013). Human gut microbiome and risk for colorectal cancer. *J. Natl. Cancer Inst.* 105, 1907–1911. doi: 10.1093/jnci/djt300
- Ai, D., Pan, H., Li, X., Gao, Y., Liu, G., and Xia, L. C. (2019). Identifying gut microbiota associated with colorectal cancer using a zero-inflated lognormal model. *Front. Microbiol.* 10:826. doi: 10.3389/fmicb.2019.00826
- Allard, G., Ryan, F. J., Jeffery, I. B., and Claesson, M. J. (2015). SPINGO: a rapid species-classifier for microbial amplicon sequences. *BMC Bioinformat.* 16:324. doi: 10.1186/s12859-015-0747-1
- Avuthu, N., and Guda, C. (2022). Meta-analysis of altered gut microbiota reveals microbial and metabolic biomarkers for colorectal cancer. *Microbiol. Spectr.* 10:e0001322. doi: 10.1128/spectrum.00013-22
- Baxter, N. T., Ruffin, M. T. t., Rogers, M. A., and Schloss, P. D. (2016). Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. *Genome Med.* 8:37. doi: 10.1186/s13073-016-0290-3
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303
- Chang, S. C., Shen, M. H., Liu, C. Y., Pu, C. M., Hu, J. M., and Huang, C. J. (2020). A gut butyrate-producing bacterium *Butyrivibrio pullicaeorum* regulates short-chain fatty acid transporter and receptor to reduce the progression of 1,2-dimethylhydrazine-associated colorectal cancer. *Oncol. Lett.* 20:327. doi: 10.3892/ol.2020.12190
- Chang, Y., Huang, Z., Hou, F., Liu, Y., Wang, L., Wang, Z., et al. (2023). Parvimonas micra activates the Ras/ERK/c-Fos pathway by upregulating miR-218-5p to promote colorectal cancer progression. *J. Exp. Clin. Cancer Res.* 42:13. doi: 10.1186/s13046-022-02572-2
- Chassard, C., Delmas, E., Robert, C., and Bernalier-Donadille, A. (2010). The cellulose-degrading microbial community of the human gut varies according to the presence or absence of methanogens. *FEMS Microbiol. Ecol.* 74, 205–213. doi: 10.1111/j.1574-6941.2010.00941.x
- Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., et al. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucl. Acids Res.* 37, D141–D145. doi: 10.1093/nar/gkn879
- Collaborators, G. B. D. C.C. (2019). The global, regional, and national burden of colorectal cancer and its attributable risk factors in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol. Hepatol.* 4, 913–933. doi: 10.1016/S2468-1253(19)30345-0
- Desai, M. S., Seekatz, A. M., Koropatkin, N. M., Kamada, N., Hickey, C. A., Wolter, M., et al. (2016). A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell* 167, 1339–1353 e1321. doi: 10.1016/j.cell.2016.10.043
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461
- Ezeji, J. C., Sarikonda, D. K., Hopperton, A., Erkkila, H. L., Cohen, D. E., Martinez, S. P., et al. (2021). *Parabacteroides distasonis*: intriguing aerotolerant gut anaerobe with emerging antimicrobial resistance and pathogenic and probiotic roles in human health. *Gut Microbes* 13:1922241. doi: 10.1080/19490976.2021.1922241
- 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
- Fukuda, S., Toh, H., Hase, K., Oshima, K., Nakanishi, Y., Yoshimura, K., et al. (2011). Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* 469, 543–547. doi: 10.1038/nature09646
- Gagniere, J., Raisch, J., Veziant, J., Barnich, N., Bonnet, R., Buc, E., et al. (2016). Gut microbiota imbalance and colorectal cancer. *World J. Gastroenterol.* 22, 501–518. doi: 10.3748/wjg.v22.i2.501

Acknowledgments

We thank the study participants and our colleagues who helped facilitate our study. We also thank Paul W. O'Toole (APC Microbiome Ireland and School of Microbiology, University College Cork, Cork, Ireland) for his critical revision and helpful comments on the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2024.1388740/full#supplementary-material>

- Ghosh, T. S., Das, M., Jeffery, I. B., and O'Toole, P. W. (2020). Adjusting for age improves identification of gut microbiome alterations in multiple diseases. *Elife* 9:50240. doi: 10.7554/elife.50240
- Gupta, A., Dhakan, D. B., Maji, A., Saxena, R., Mahajan, S., Pulikkal, J., et al. (2019). Association of Flavonifractor plautii, a Flavonoid-Degrading Bacterium, with the Gut Microbiome of Colorectal Cancer Patients in India. *mSystems*. 4:e00438–19. doi: 10.1128/mSystems.00438-19
- Iadsee, N., Chuaypen, N., Techawiwattanaboon, T., Jinato, T., Patcharatrakul, T., Malakorn, S., et al. (2023). Identification of a novel gut microbiota signature associated with colorectal cancer in Thai population. *Sci. Rep.* 13:6702. doi: 10.1038/s41598-023-33794-9
- Koh, G. Y., Kane, A. V., Wu, X., and Crott, J. W. (2020). Parabacteroides distasonis attenuates tumorigenesis, modulates inflammatory markers and promotes intestinal barrier integrity in azoxymethane-treated A/J mice. *Carcinogenesis* 41, 909–917. doi: 10.1093/carcin/bgaa018
- Li, L., Li, C., Lv, M., Hu, Q., Guo, L., and Xiong, D. (2020). Correlation between alterations of gut microbiota and miR-122-5p expression in patients with type 2 diabetes mellitus. *Ann. Transl. Med.* 8:1481. doi: 10.21037/atm-20-6717
- Liang, J. Q., Zeng, Y., Lau, E. Y. T., Sun, Y., Huang, Y., Zhou, T., et al. (2023). A probiotic formula for modulation of colorectal cancer risk via reducing CRC-associated bacteria. *Cells* 12:91244. doi: 10.3390/cells12091244
- Lo, C. H., Wu, D. C., Jao, S. W., Wu, C. C., Lin, C. Y., Chuang, C. H., et al. (2022). Enrichment of *Prevotella intermedia* in human colorectal cancer and its additive effects with *Fusobacterium nucleatum* on the malignant transformation of colorectal adenomas. *J. Biomed. Sci.* 29:88. doi: 10.1186/s12929-022-00869-0
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8
- Ma, X., Lu, X., Zhang, W., Yang, L., Wang, D., Xu, J., et al. (2022). Gut microbiota in the early stage of Crohn's disease has unique characteristics. *Gut Pathog.* 14:46. doi: 10.1186/s13099-022-00521-0
- Magoc, T., and Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963. doi: 10.1093/bioinformatics/btr507
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10–12. doi: 10.14806/ej.17.1.200
- Mori, G., Orena, B. S., Cultrera, I., Barbieri, G., Albertini, A. M., Ranzani, G. N., et al. (2019). Gut microbiota analysis in postoperative lynch syndrome patients. *Front. Microbiol.* 10:1746. doi: 10.3389/fmicb.2019.01746
- Obon-Santacana, M., Mas-Lloret, J., Bars-Cortina, D., Criado-Mesas, L., Carreras-Torres, R., Diez-Villanueva, A., et al. (2022). Meta-analysis and validation of a colorectal cancer risk prediction model using deep sequenced fecal metagenomes. *Cancers* 14:174214. doi: 10.3390/cancers14174214
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09
- Singh, N., Gurav, A., Sivaprakasam, S., Brady, E., Padia, R., Shi, H., et al. (2014). Activation of Gpr109a, receptor for niacin and the commensal metabolite butyrate, suppresses colonic inflammation and carcinogenesis. *Immunity* 40, 128–139. doi: 10.1016/j.jimmuni.2013.12.007
- Stoffel, E. M., and Murphy, C. C. (2020). Epidemiology and mechanisms of the increasing incidence of colon and rectal cancers in young adults. *Gastroenterology* 158, 341–353. doi: 10.1053/j.gastro.2019.07.055
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 71, 209–249. doi: 10.3322/caac.21660
- Ternes, D., Karta, J., Tsenkova, M., Wilmes, P., Haan, S., and Letellier, E. (2020). Microbiome in colorectal cancer: how to get from meta-omics to mechanism? *Trends Microbiol.* 28, 401–423. doi: 10.1016/j.tim.2020.01.001
- Tito, R. Y., Verbandt, S., Aguirre Vazquez, M., Lahti, L., Verspecht, C., Llorens-Rico, V., et al. (2024). Microbiome confounders and quantitative profiling challenge predicted microbial targets in colorectal cancer development. *Nat. Med.* 30, 1339–1348. doi: 10.1038/s41591-024-02963-2
- Tlaskalova-Hogenova, H., Vannucci, L., Klimesova, K., Stepankova, R., Krizan, J., and Kverka, M. (2014). Microbiome and colorectal carcinoma: insights from germ-free and conventional animal models. *Cancer J.* 20, 217–224. doi: 10.1097/PPO.0000000000000052
- Tran, H. N. H., Thu, T. N. H., Nguyen, P. H., Vo, C. N., Doan, K. V., Nguyen Ngoc Minh, C., et al. (2022). Tumour microbiomes and Fusobacterium genomics in Vietnamese colorectal cancer patients. *NPJ Biofilms Microbiomes* 8:87. doi: 10.1038/s41522-022-00351-7
- Tran, T. T. T., Cousin, F. J., Lynch, D. B., Menon, R., Brulc, J., Brown, J. R., et al. (2019). Prebiotic supplementation in frail older people affects specific gut microbiota taxa but not global diversity. *Microbiome* 7:39. doi: 10.1186/s40168-019-0654-1
- Umireh, F., Neoh, C. F., Ramasamy, K., and Lim, S. M. (2021). Differential gut microbiota composition between type 2 diabetes mellitus patients and healthy controls: a systematic review. *Diabet. Res. Clin. Pract.* 173:108689. doi: 10.1016/j.diabres.2021.108689
- 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
- Wong, S. H., and Yu, J. (2019). Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. *Nat. Rev. Gastroenterol. Hepatol.* 16, 690–704. doi: 10.1038/s41575-019-0209-8
- 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
- Wu, T. R., Lin, C. S., Chang, C. J., Lin, T. L., Martel, J., Ko, Y. F., et al. (2019). Gut commensal *Parabacteroides goldsteinii* plays a predominant role in the anti-obesity effects of polysaccharides isolated from *Hirsutella sinensis*. *Gut* 68, 248–262. doi: 10.1136/gutjnl-2017-315458
- Yachida, S., Mizutani, S., Shiroma, H., Shiba, S., Nakajima, T., Sakamoto, T., et al. (2019). Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat. Med.* 25, 968–976. doi: 10.1038/s41591-019-0458-7
- Young, C., Wood, H. M., Fuentes Balaguer, A., Bottomley, D., Gallop, N., Wilkinson, L., et al. (2021). Microbiome analysis of more than 2,000 NHS bowel cancer screening programme samples shows the potential to improve screening accuracy. *Clin. Cancer Res.* 27, 2246–2254. doi: 10.1158/1078-0432.CCR-20-3807
- Yu, J., Feng, Q., Wong, S. H., Zhang, D., Liang, Q. Y., Qin, Y., et al. (2017). Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* 66, 70–78. doi: 10.1136/gutjnl-2015-309800
- Zafar, H., and Saier, M. H. Jr. (2021). Gut *bacteroides* species in health and disease. *Gut Microbes* 13, 1–20. doi: 10.1080/19490976.2020.1848158
- Ze, X., Duncan, S. H., Louis, P., and Flint, H. J. (2012). *Ruminococcus bromii* is a keystone species for the degradation of resistant starch in the human colon. *ISME J.* 6, 1535–1543. doi: 10.1038/ismej.2012.4
- Zhang, Z., Zhao, L., Wu, J., Pan, Y., Zhao, G., Li, Z., et al. (2023). The effects of *Lactobacillus johnsonii* on diseases and its potential applications. *Microorganisms* 11:102580. doi: 10.3390/microorganisms11102580
- Zheng, D., Wang, Z., Sui, L., Xu, Y., Wang, L., Qiao, X., et al. (2021). *Lactobacillus johnsonii* activates porcine monocyte derived dendritic cells maturation to modulate Th cellular immune response. *Cytokine* 144:155581. doi: 10.1016/j.cyto.2021.155581