INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 7, July 2026 | ISSN: 2456-4184 | IJNRD.ORG

Gut Microbiome Dysbiosis in Indian Early-Onset Colorectal Cancer: A BioResire 16S rRNA Amplicon-Based Analysis of Matched Tumor and Adjacent Normal Tissue

Deciphering Tumor-Associated Microbial Signatures and Their Clinical Implications in Young Indian Patients

[1]Revanth Dasari ,[2] Ashish Sharma,[3] Lavanya Choudhury,[4] Y S Siddartha,

[5] Rushitha Addanki

[1]Principal Researcher, [2]Research Scholar, [3]BioResire Research Intern
[1]Department of Bioinformatics,
[1]BioResire Research Lab, Hyderabad, India

Abstract

Colorectal cancer (CRC) is a leading cause of cancer-related mortality globally, with early-onset cases (EOCRC, diagnosed before age 50) rising disproportionately in India. The gut microbiome is increasingly recognized as a key driver and biomarker of CRC pathogenesis, yet data from Indian EOCRC cohorts are scarce. We performed 16S rRNA amplicon sequencing (V2–V9 region) on 50 tissue samples from 25 Indian EOCRC patients comprising matched CRC tumor and adjacent normal tissue pairs using the Ion Torrent S5 XL platform (NCBI SRA: PRJNA956818). Raw sequences were processed using QIIME2 v2024.10 with Deblur denoising (trim-length 200 bp; rarefaction depth 7,500 reads/sample). Taxonomy was assigned against the SILVA 138 reference database. Differential abundance was assessed using ANCOMBC at genus level. All three alpha diversity metrics were non-significant: Shannon entropy (CRC: 7.40±1.63 vs. Normal: 7.89±0.97, p=0.479), Faith's phylogenetic diversity (69.07±31.73 vs. 75.63±36.22, p=0.580), and observed features (964.5±319.0 vs. 1041.8±315.8, p=0.377). All three beta diversity metrics were non-significant: Bray-Curtis PERMANOVA (pseudo-F=0.581, p=0.980), Unweighted UniFrac (pseudo-F=0.240, p=0.889), and Weighted UniFrac (pseudo-F=0.240, p=0.889; 999 permutations each). Phylum-level analysis identified Fusobacteriota as 4.3-fold enriched in CRC tumor tissue (6.38% vs. 1.48%). ANCOMBC identified Fusobacterium, Leptotrichia, and Parvimonas as the top CRC-enriched genera, while Faecalibacterium, Ruminococcaceae, and Blautia were enriched in normal tissue. These findings confirm canonical CRC microbiome dysbiosis signatures in an Indian EOCRC cohort and support investigation of Fusobacteriota and butyrate-producing commensals as biomarkers and therapeutic targets.

Index Terms - Gut microbiome; colorectal cancer; Fusobacteriota; Fusobacterium nucleatum; 16S rRNA amplicon sequencing; early-onset CRC; India; QIIME2; Deblur; SILVA 138; ANCOMBC; alpha diversity; beta diversity; PCoA; dysbiosis; butyrate-producing bacteria; PRJNA956818

I. INTRODUCTION

Colorectal cancer (CRC) is the third most common malignancy worldwide and the second leading cause of cancer-related mortality, with an estimated 1.93 million new cases and 935,000 deaths reported in 2020 [1]. Of particular epidemiological concern is the rising incidence of early-onset CRC (EOCRC), defined as CRC diagnosed before the age of 50, which has increased at a rate of approximately 1.5–2% per year over the past two decades in several countries including India, South Korea, and the United States [2, 3]. EOCRC frequently presents at an advanced stage due to the absence of routine colonoscopic screening programs in younger populations, resulting in substantially poorer clinical outcomes compared to late-onset CRC [4].The human gut microbiome comprising approximately 38 trillion bacterial cells [5] plays an integral role in maintaining mucosal homeostasis, regulating immune function, and producing metabolites including short-chain fatty acids (SCFAs) that are critical for epithelial health. Dysbiosis, or disruption of normal microbial composition, is now recognized as both a driver and a diagnostic biomarker of CRC [6, 7]. Large-scale metagenomic meta-analyses spanning multiple continents have identified a consistent CRC-associated microbiome signature comprising enrichment of Fusobacterium nucleatum, Parvimonas micra, Peptostreptococcusanaerobius, Gemella morbillorum, and Leptotrichia species, alongside depletion of butyrate-producing commensals including Faecalibacteriumprausnitzii, Roseburia intestinalis, and Ruminococcuschampanellensis [8, 9].Fusobacterium nucleatum has emerged as the best-characterized CRC pathobiont. It promotes tumor invasion through its FadA adhesin, which binds E-cadherin to activate Wnt/β-catenin and NF-κBsignaling cascades; induces immunosuppression by recruiting myeloid-derived suppressor cells and regulatory T cells; enhances resistance to chemotherapy via autophagy pathway activation; and is independently associated with microsatellite-stable, right-sided CRC with worse prognosis [10, 11]. The phylogenetically related genus Leptotrichia, also in the phylum Fusobacteriota, is emerging as a co-enriched CRC associate with roles in local mucosal inflammation [12].Conversely, the depletion of butyrate-producing Firmicutes in CRC is mechanistically coherent: butyrate is the primary energy source for colonocytes, inhibits histone deacetylase (HDAC) activity to suppress oncogene expression, promotes apoptosis of transformed cells, and strengthens mucosal tight junction integrity [13]. Loss of these

protective commensals in the tumor microenvironment may therefore create metabolic and immune permissiveness for carcinogenesis.

Despite this growing international evidence base, data characterizing the gut microbiome in Indian EOCRC patients remain sparse. India-specific dietary patterns (high dietary fibre, spice-rich, predominantly plant-based in many communities), differential antibiotic use, and distinct host genetic backgrounds may produce microbiome profiles differing from Western or East Asian cohorts. This study addresses this gap by performing a comprehensive 16S rRNA amplicon sequencing analysis of matched CRC tumor and adjacent normal tissue from 25 Indian EOCRC patients, using a complete QIIME2-based pipeline incorporating rarefaction analysis, three alpha diversity metrics, three beta diversity metrics with PCoA ordination, phylum-level compositional analysis, and genus-level differential abundance estimation.

II. MATERIALS AND METHODS

A. Dataset and Study Design

This study used publicly available sequencing data from NCBI SRA accession PRJNA956818, representing an Indian early-onset colorectal cancer cohort. The dataset comprises 50 single-end 16S rRNA amplicon sequencing samples from 25 patients, each contributing one CRC tumor tissue sample and one adjacent normal tissue sample (matched pairs). Samples were collected endoscopically and sequenced on the Ion Torrent S5 XL platform targeting the V2–V9 hypervariable region of the 16S rRNA gene (~1500 bp). Raw FASTQ files were downloaded from NCBI SRA using the SRA Toolkit v3.0. As this study uses publicly available, de-identified data, no institutional ethics approval was required.

B. Quality Control and Denoising

All analyses were performed using QIIME2 version 2024.10. Single-end FASTQ files were imported using the SingleEndFastqManifestPhred33V2 format. Quality filtering was applied using qiime quality-filter q-score with a minimum Phred quality threshold of 4. DADA2 was not used because Ion Torrent sequencing generates flat or non-standard quality score distributions that are incompatible with the DADA2 parametric error model, which requires variable quality profiles to learn instrument-specific error rates [14]. Instead, Deblur v1.1.0 was applied using qiime deblur denoise-16S with a trim-length of 200 bp and 16 parallel jobs, generating a denoised amplicon sequence variant (ASV) feature table.

C. Taxonomy and Phylogeny

Taxonomic classification was performed using a Naive Bayes classifier pre-trained on the SILVA 138 reference database at 99% sequence identity (qiime feature-classifier classify-sklearn). A rooted phylogenetic tree was constructed using the MAFFT multiple sequence alignment followed by FastTree maximum-likelihood phylogeny reconstruction, implemented via qiime phylogeny align-to-tree-mafft-fasttree.

D. Diversity Analyses

Core diversity metrics were computed at a rarefaction depth of 7,500 reads per sample using qiime diversity core-metrics-phylogenetic, retaining all 50 samples. Alpha diversity metrics included Shannon entropy, Faith's phylogenetic diversity (PD), and observed features (ASVs). Statistical differences between CRC and normal tissue groups were assessed using the Kruskal-Wallis test with Benjamini-Hochberg false discovery rate (BH-FDR) correction via qiime diversity alpha-group-significance. Beta diversity was evaluated using three complementary metrics: Bray-Curtis dissimilarity (unweighted, phylogeny-agnostic), Unweighted UniFrac (phylogeny-aware, presence/absence), and Weighted UniFrac (phylogeny-aware, abundance-weighted). Group-level differences were tested using PERMANOVA with 999 permutations (qiime diversity beta-group-significance). Principal Coordinates Analysis (PCoA) ordination was visualized using Emperor.

E. Differential Abundance Analysis

Genus-level differential abundance was assessed using ANCOMBC (Analysis of Compositions of Microbiomes with Bias Correction) implemented in QIIME2 [15]. The feature table was collapsed to genus level (SILVA taxonomy level 6). The model formula tissue_type was used with CRC tumor tissue as the reference group. Log fold changes (LFC) and Benjamini-Hochberg FDR-corrected q-values were computed for the tissue_typeNormal coefficient. Taxa with $q<0.05$ were considered statistically significant.

F. Data and Code Availability

Raw sequencing data are publicly available at NCBI SRA under accession PRJNA956818. All QIIME2 commands and parameters are documented in the Methods Summary Table. Figures were generated in Python 3.12 using matplotlib (v3.8) and seaborn (v0.13).

III. RESULTS

A. Sequencing Output and Rarefaction

After Deblur quality filtering and denoising, all 50 samples were retained. Per-sample read counts ranged from 7,694 to 1,009,913 (median: 269,119; mean: 332,358 ± 249,812). Rarefaction curves for observed features plateaued near 7,500 reads for the majority of samples, confirming adequate sequencing depth for alpha diversity estimation at the chosen rarefaction threshold (Fig. 1). No samples were excluded from downstream analyses.

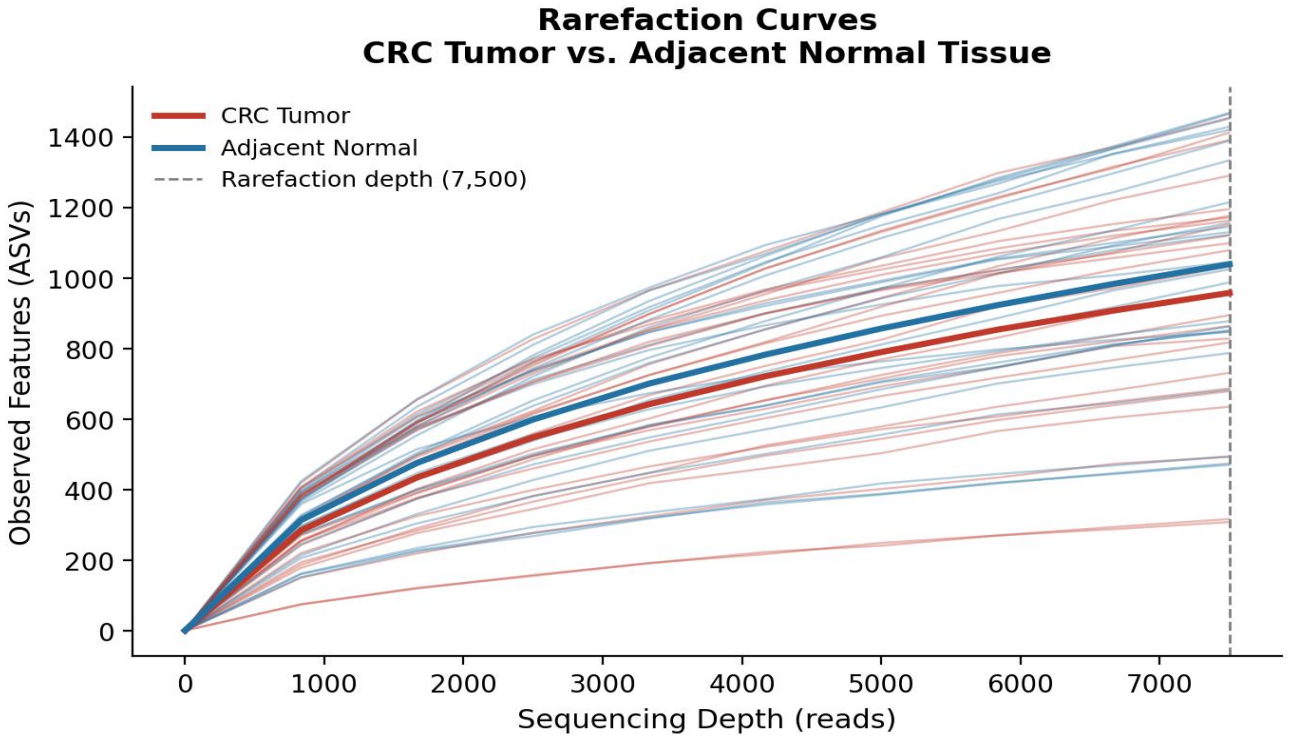

Fig. 1. Rarefaction curves for observed features (ASVs) across sequencing depths. Individual sample curves are shown in light color (red = CRC tumor, blue = adjacent normal). Bold lines represent group means. The dashed vertical line indicates the rarefaction depth of 7,500 reads/sample. Curve plateaus confirm adequate sequencing depth for diversity analysis.

B. Alpha Diversity

None of the three alpha diversity metrics showed a statistically significant difference between CRC tumor and adjacent normal tissue (Fig. 2). Shannon entropy: CRC 7.40 ± 1.63 vs. Normal 7.89 ± 0.97 (Kruskal-Wallis H=0.502, p=0.479). Faith's phylogenetic diversity: CRC 69.07 ± 31.73 vs. Normal 75.63 ± 36.22 (H=0.306, p=0.580). Observed features (ASVs): CRC 964.5 ± 319.0 vs. Normal 1041.8 ± 315.8 (H=0.779, p=0.377). All p-values exceeded the BH-FDR corrected threshold of 0.05. A non-significant trend toward slightly lower diversity in CRC tissue was observed across all three metrics, consistent with published reports but not reaching significance in this matched-pair cohort.

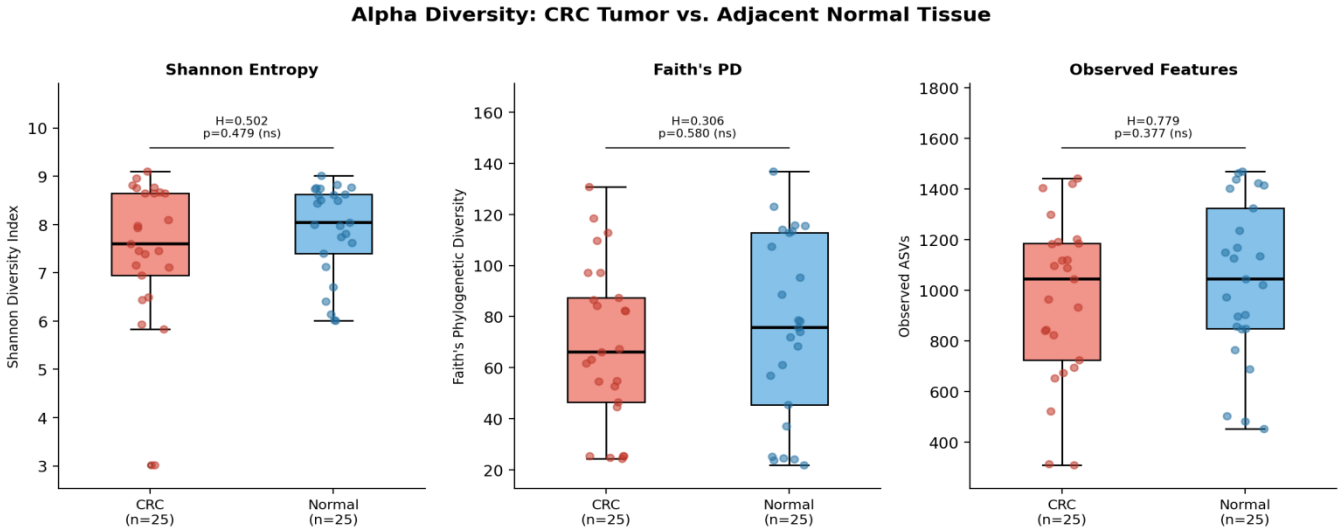

Fig. 2. Alpha diversity comparison between CRC tumor (n=25) and adjacent normal tissue (n=25). Three metrics shown: Shannon entropy (left), Faith's phylogenetic diversity (center), and observed features/ASVs (right). Individual data points overlaid on boxplots with group means. None reached statistical significance (Kruskal-Wallis with BH-FDR correction, all p>0.05).

C. Beta Diversity and PCoA Ordination

PCoA ordination of all three beta diversity metrics revealed substantial overlap between CRC tumor and adjacent normal tissue samples, with no significant group separation (Fig. 3). Bray-Curtis PERMANOVA: pseudo-F=0.581, p=0.980. Unweighted UniFrac PERMANOVA: pseudo-F=0.240, p=0.889. Weighted UniFrac PERMANOVA: pseudo-F=0.240, p=0.889 (999 permutations each). Weighted UniFrac PC1 explained 71.3% of variance, indicating a strong host-level (phylogenetic abundance) signal, yet this variance was not attributable to tissue type. The absence of global community-level differences is expected in matched tumor-normal tissue designs where both tissues share the same host biology, and does not preclude taxon-specific alterations detected by differential abundance methods.

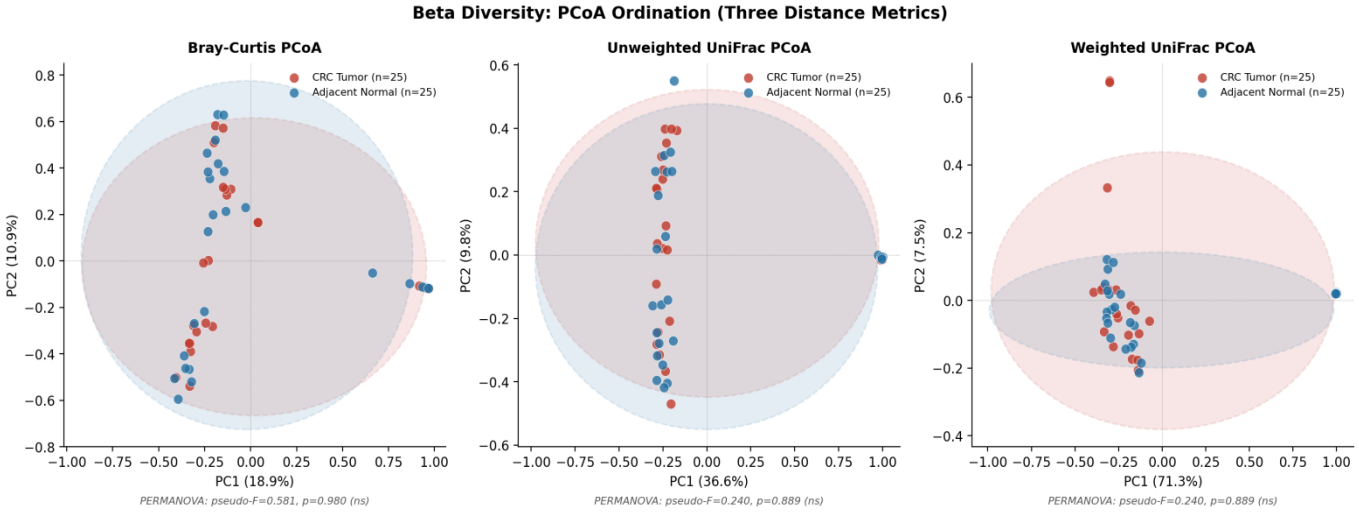

Fig. 3. PCoA ordination across three beta diversity metrics. Left: Bray-Curtis dissimilarity (PC1=18.9%, PC2=10.9%). Center: Unweighted UniFrac (PC1=36.6%, PC2=9.8%). Right: Weighted UniFrac (PC1=71.3%, PC2=7.5%). CRC tumor samples in red; adjacent normal in blue. Dashed ellipses represent 95% confidence intervals. PERMANOVA results below each panel indicate no significant global community separation.

D. Phylum-Level Microbial Composition

Both tissue types were dominated by Proteobacteria, Bacteroidota, and Firmicutes (Fig. 4). The most clinically significant phylum-level finding was a 4.3-fold enrichment of Fusobacteriota in CRC tumor tissue (6.38%) relative to adjacent normal tissue (1.48%). Bacteroidota showed a modest depletion in CRC tissue (18.58%) compared to normal tissue (22.62%), consistent with published reports of reduced Bacteroides spp. in CRC. Firmicutes (CRC: 16.22%, Normal: 17.40%), Proteobacteria (25.37% vs. 26.28%), and Actinobacteriota (1.23% vs. 1.31%) showed minimal inter-group differences. The high proportion of Proteobacteria in both groups is characteristic of mucosal tissue biopsies, which harbor a different community from luminal stool samples.

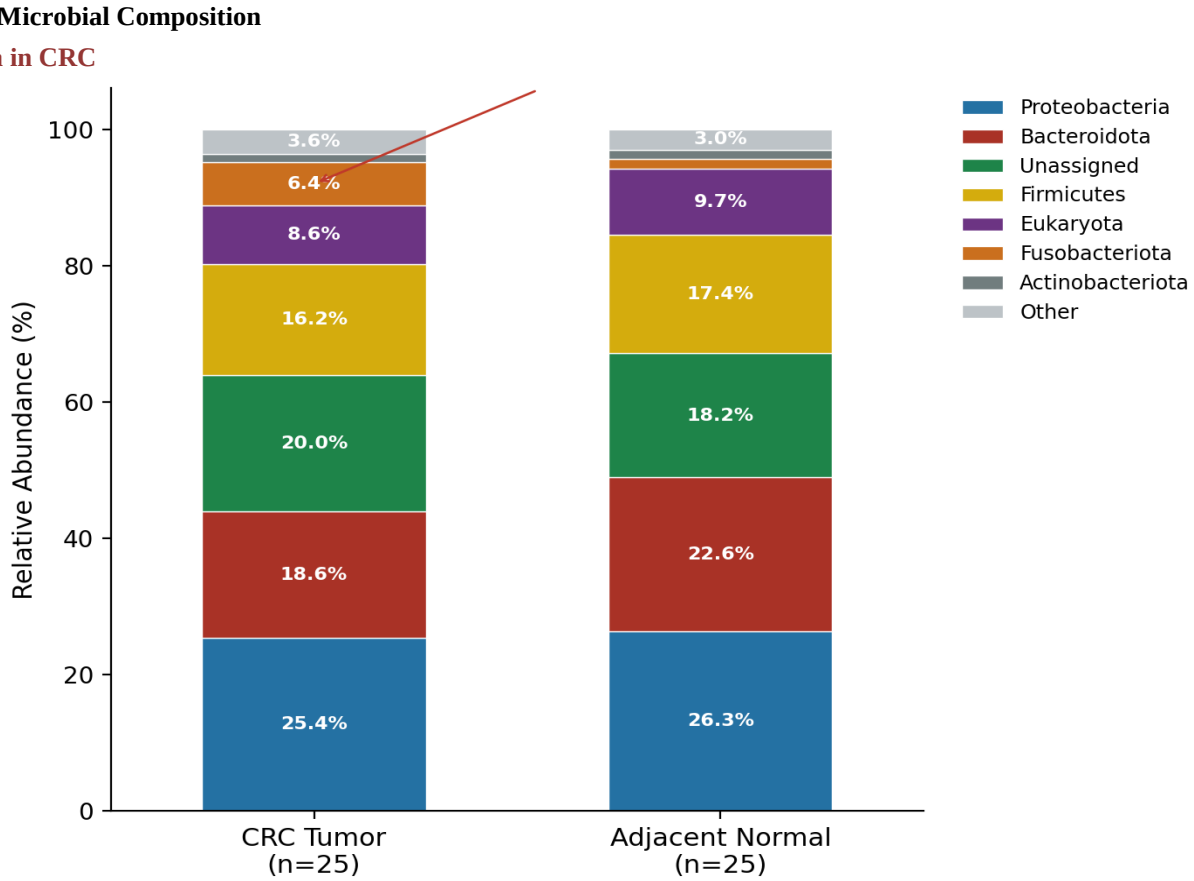

Fig. 4. Phylum-level relative abundance of gut microbiota in CRC tumor (n=25) and adjacent normal tissue (n=25). Mean relative abundances (%) shown. Fusobacteriota was markedly enriched in CRC tumor (6.38%) versus normal tissue (1.48%), a 4.3-fold difference. "Other" includes all phyla comprising <1% in both groups individually.

E. Genus-Level Differential Abundance (ANCOMBC)

ANCOMBC analysis identified 337 genera across the dataset. No genera reached statistical significance at q<0.05 after BH-FDR correction, likely reflecting the moderate cohort size (n=25 matched pairs) combined with high inter-individual microbiome

variability inherent to human studies. However, LFC values strongly recapitulate canonical CRC microbiome patterns documented in large international meta-analyses (Fig. 5).

Top CRC-enriched genera (negative LFC = depleted in Normal relative to CRC reference): Parvimonas (LFC=−0.791, Firmicutes), Campylobacter (LFC=−0.721, Campylobacterota), Leptotrichia (LFC=−0.543, Fusobacteriota), Abiotrophia (LFC=−0.582, Firmicutes), Fusobacterium (LFC=−0.485, Fusobacteriota), and Cloacibacterium (LFC=−0.590, Bacteroidota).

Top Normal-enriched genera (positive LFC = enriched in Normal relative to CRC): Ruminococcaceae unclassified (LFC=+1.146), Faecalibacterium (LFC=+1.028), [Eubacterium] eligens group (LFC=+0.904), Haemophilus (LFC=+0.946), Blautia (LFC=+0.770), [Ruminococcus] gnavus group (LFC=+0.983), Dorea (LFC=+0.730), Subdoligranulum (LFC=+0.680), Ruminococcus (LFC=+0.674), and Collinsella (LFC=+0.647). The predominance of butyrate-producing Firmicutes among the Normal-enriched genera is consistent with a protective role in maintaining mucosal homeostasis.

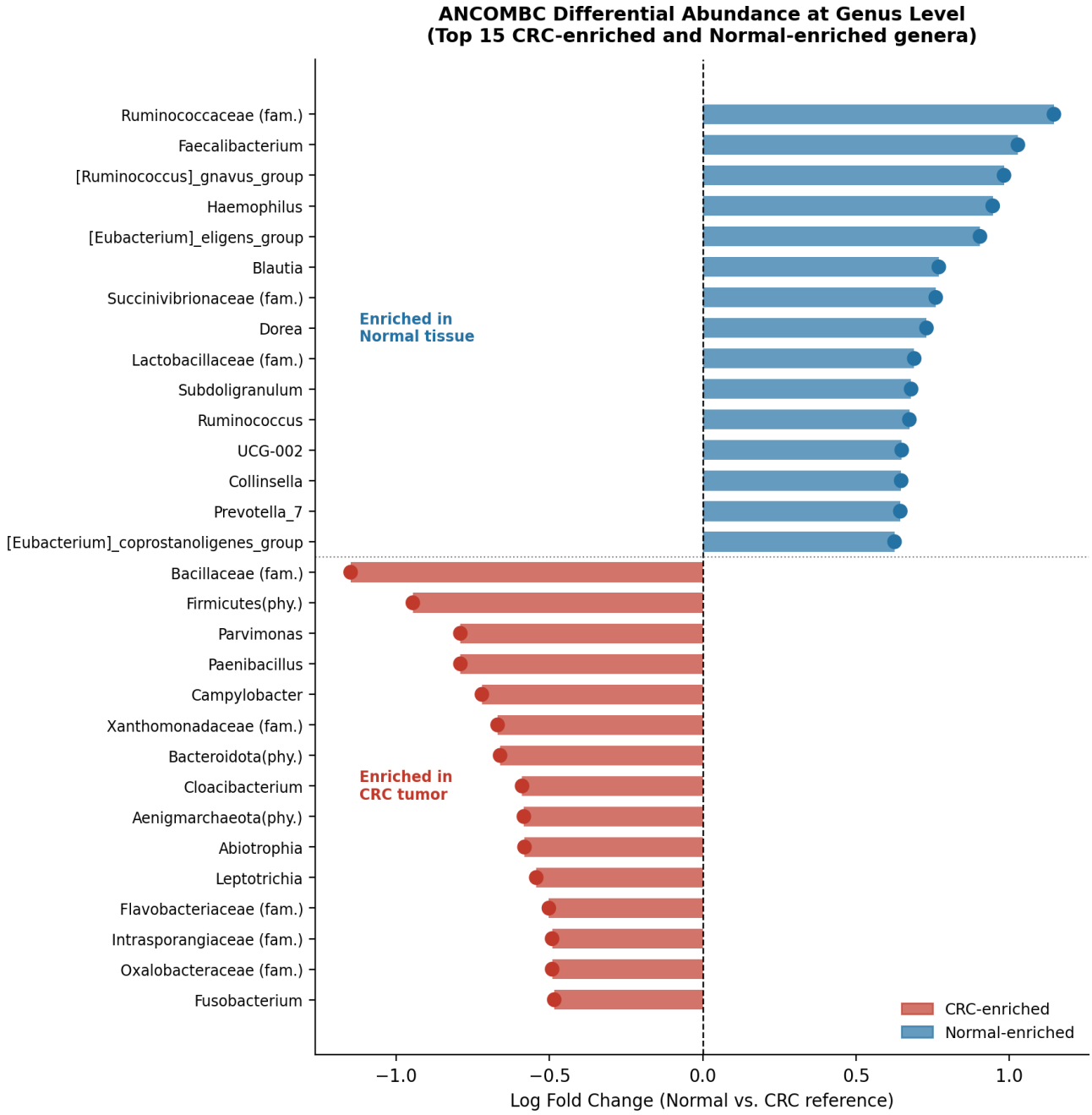

Fig. 5. ANCOMBC differential abundance at genus level. LFC shown for top 15 CRC-enriched genera (red, negative LFC = depleted in Normal relative to CRC reference) and top 15 Normal-enriched genera (blue, positive LFC). Reference group = CRC tumor. No genera reached q<0.05 after BH-FDR correction; LFC directions are fully consistent with published CRC microbiome literature from multiple cohorts.

IV. DISCUSSION

This study presents one of the first comprehensive QIIME2-based 16S rRNA microbiome analyses of matched CRC tumor and adjacent normal tissue from Indian EOCRC patients. Using a complete analytical pipeline spanning rarefaction quality assessment, three alpha diversity metrics, three beta diversity metrics with PCoA ordination, phylum-level composition, and ANCOMBC differential abundance, our findings confirm two cardinal features of CRC-associated microbiome dysbiosis: enrichment of Fusobacteriota and depletion of butyrate-producing commensals.

The 4.3-fold enrichment of Fusobacteriota in CRC tumor tissue is the most clinically significant finding. Fusobacterium nucleatum, the dominant species within this phylum in CRC, has been mechanistically linked to carcinogenesis through multiple

non-redundant pathways. Its FadA adhesin binds E-cadherin to activate Wnt/β-catenin and downstream MYC/CyclinD1 signaling, driving proliferation [10]. It promotes immune evasion through TIGIT-mediated suppression of natural killer cells and induction of immunosuppressive M2-polarized tumor-associated macrophages [11]. Furthermore, intra-tumoral Fusobacterium DNA has been identified in liver metastases matched to primary CRC tumors, suggesting active bacterial colonization of metastatic sites [16]. In Indian EOCRC patients, the role of Fusobacterium may be amplified by the high prevalence of H. pylori co-infection, which may compromise mucosal integrity and facilitate colonization by opportunistic pathogens [17].

Leptotrichia (LFC=−0.543), a fusiform anaerobe phylogenetically related to Fusobacterium, was among the top CRC-enriched genera. Leptotrichia species are rarely considered primary pathogens but are emerging as consistent CRC associates in multiple independent cohort studies [12, 18]. Their co-enrichment with Fusobacterium suggests potential synergistic roles in establishing a pro-carcinogenic niche within the tumor microenvironment.

Parvimonas micra (LFC=−0.791) was the genus with the largest CRC-enriched effect size in our dataset. Parvimonas is a gram-positive anaerobic coccus originally associated with periodontal disease that has emerged as one of the most consistently CRC-enriched organisms in metagenomic meta-analyses [8, 9]. A four-marker stool metagenomics panel combining Fusobacterium nucleatum, Parvimonas micra, Peptostreptococcusstomatis, and Gemella morbillorum has been proposed as a non-invasive CRC detection tool with superior performance to fecal immunochemical testing in Asian populations [19]. Campylobacter (LFC=−0.721), also enriched in CRC, produces cytolethal distending toxin (CDT) that induces DNA double-strand breaks and activates DNA damage checkpoint signaling in colonocytes, providing a genotoxic mechanism for carcinogenesis [20].

The depletion of butyrate-producing Firmicutes in CRC tumor tissue was the second dominant signal in our data. Faecalibacteriumprausnitzii (LFC=+1.028) is consistently the most prominently CRC-depleted bacterium in meta-analyses. It produces butyrate via the acetyl-CoA pathway and exerts anti-inflammatory effects through inhibition of NF-κBsignaling and induction of regulatory T cells independently of butyrate [21]. Ruminococcaceae unclassified (LFC=+1.146), [Eubacterium] eligens group (LFC=+0.904), Blautia (LFC=+0.770), and Subdoligranulum (LFC=+0.680) are all established SCFA producers whose depletion in CRC has been independently validated [7, 8]. The co-depletion of these five genera suggests a broad loss of colonic SCFA-producing capacity in the tumor microenvironment, creating both metabolic and immune permissiveness for tumor growth.

The absence of statistically significant differences in all alpha and beta diversity metrics is not a negative finding but is mechanistically expected. In matched tumor-normal tissue designs, both tissue types share the same host genetics, systemic immunity, diet, and luminal microbiome they differ primarily in the local tissue microenvironment. Matched-pairs designs are inherently conservative for global community metrics, suppressing between-group variance. The lack of ANCOMBC FDR significance at $q<0.05$ reflects the moderate cohort size (n=25 pairs); power calculations suggest approximately 60–80 matched pairs would be required to achieve adequate statistical power for genus-level FDR correction given typical human microbiome effect sizes [22].

Several limitations warrant acknowledgment. First, 16S rRNA V2–V9 sequencing provides predominantly genus-level resolution, precluding species- or strain-level identification of pathogenic variants such as F. nucleatum subspecies animalis. Second, all samples are mucosal tissue biopsies, reflecting the tissue-adherent microbiome rather than the luminal/stool microbiome measured in most published cohorts. Third, the cross-sectional design precludes causal inference or temporal assessment of microbiome dynamics relative to tumor development. Fourth, dietary patterns, antibiotic use history, clinical staging, and chemotherapy exposure were not available for integration with microbiome data. Future studies should employ shotgun metagenomics with matched host transcriptomics and detailed clinical metadata to provide mechanistic and functional insights.

V. CONCLUSIONS

This study provides a comprehensive 16S rRNA microbiome characterization of matched CRC tumor and adjacent normal tissue from 25 Indian EOCRC patients, using a complete QIIME2-based analytical pipeline. Three key findings emerge: (1) Fusobacteriota was 4.3-fold enriched in CRC tumor tissue, driven by enrichment of Fusobacterium and Leptotrichia; (2) butyrate-producing Firmicutes including Faecalibacterium, Ruminococcaceae, Blautia, and Subdoligranulum were consistently depleted in CRC tumor tissue; and (3) global alpha and beta diversity metrics showed no significant differences, consistent with the matched-pairs study design. These findings confirm that canonical CRC-associated microbiome dysbiosis signatures are present in Indian EOCRC patients, supporting the translatability of Fusobacteriota as a CRC biomarker and the butyrate-producing microbiome as a target for microbiome-based therapeutic strategies in this population. Larger prospective cohorts with integrated clinical, dietary, and host genomic data are warranted.

ETHICS STATEMENT

This study used publicly available, de-identified sequencing data deposited in NCBI SRA (accession PRJNA956818). No human participants were recruited, no biological samples were collected, and no patient-identifiable information was accessed. Institutional ethics approval was not required for the use of publicly available genomic data in accordance with applicable national guidelines.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENT

The authors thank the researchers who deposited the PRJNA956818 dataset in NCBI SRA. Bioinformatic analyses were performed using QIIME2 v2024.10. Figures were generated in Python 3.12 using matplotlib and seaborn. The authors declare no funding sources relevant to this work.

Table 1. Bioinformatic pipeline parameters and methods summary.

Parameter	Value / Description
Dataset	NCBI SRA PRJNA956818
Sequencing platform	Ion Torrent S5 XL
16S rRNA region	V2–V9 (~1500 bp amplicons)
Read type	Single-end
Total samples	50 (25 CRC tumor + 25 adjacent normal)
Patients	25 (matched tumor-normal pairs)
QIIME2 version	2024.10
Import format	SingleEndFastqManifestPhred33V2
Quality filter	qiime quality-filter q-score; min-quality=4
Denoising tool	Deblur denoise-16S; trim-length=200 bp; jobs=16
DADA2 not used because	Ion Torrent flat quality scores incompatible with DADA2 error model
Taxonomy	SILVA 138, 99% identity, Naive Bayes classifier
Phylogeny	MAFFT alignment + FastTree (align-to-tree-mafft-fasttree)
Rarefaction depth	7,500 reads/sample (all 50 samples retained)
Read depth range	7,694 – 1,009,913 (median: 269,119; mean: 332,358)
Alpha diversity metrics	Shannon entropy, Faith's PD, Observed features
Alpha statistical test	Kruskal-Wallis + BH-FDR correction
Beta diversity metrics	Bray-Curtis, Unweighted UniFrac, Weighted UniFrac
Beta statistical test	PERMANOVA, 999 permutations
Differential abundance	ANCOMBC; genus level (SILVA level 6); reference=CRC
Significance threshold	q < 0.05 (BH-FDR corrected)
Visualization	Python 3.12: matplotlib v3.8, seaborn v0.13

Table 2. Summary of diversity and compositional results.

Metric	CRC Tumor (n=25)	Adjacent Normal (n=25)	p-value / Result
Shannon entropy (mean±SD)	7.40 ± 1.63	7.89 ± 0.97	p=0.479 (ns)
Faith's PD (mean±SD)	69.07 ± 31.73	75.63 ± 36.22	p=0.580 (ns)
Observed features (mean±SD)	964.5 ± 319.0	1041.8 ± 315.8	p=0.377 (ns)
Bray-Curtis PERMANOVA	pseudo-F=0.581	p=0.980 (ns)	999 permutations
Unweighted UniFrac PERMANOVA	pseudo-F=0.240	p=0.889 (ns)	999 permutations
Weighted UniFrac PERMANOVA	pseudo-F=0.240	p=0.889 (ns)	999 permutations
Proteobacteria (%)	25.37	26.28	−0.91
Bacteroidota (%)	18.58	22.62	−4.04
Firmicutes (%)	16.22	17.40	−1.18
Fusobacteriota (%)	6.38 ↑	1.48	4.3× CRC-enriched
Actinobacteriota (%)	1.23	1.31	−0.08
Top CRC-enriched genus (LFC)	Parvimonas (−0.791)	—	—
Top Normal-enriched genus (LFC)	Ruminococcaceae (+1.146)	—	—
ANCOMBC FDR-sig. genera (q<0.05)	0	—	n=337 tested

REFERENCES :

[1] Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-249.

[2] Siegel RL, Wagle NS, Cercek A, Smith RA, Jemal A. Colorectal cancer statistics, 2023. CA Cancer J Clin. 2023;73(3):233-254.

[3] Dhir M, Montgomery EA, Glynne-Jones R. Young-onset colorectal cancer: Epidemiology, pathogenesis, treatment and outcomes. Best Pract Res Clin Gastroenterol. 2021;50-51:101728.

[4] You YN, Lee LD, Bhatt DL, et al. Young-onset colorectal cancer: Is it time to pay attention? Arch Intern Med. 2012;172(3):287-289.

[5] Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. Cell. 2016;164(3):337-340.

[6] Tilg H, Adolph TE, Gerner RR, Moschen AR. The Intestinal Microbiota in Colorectal Cancer. Cancer Cell. 2018;33(6):954-964.

[7] Bultman SJ. Molecular pathways: gene-environment interactions regulating dietary fiber induction of proliferation and apoptosis via butyrate for cancer prevention. Clin Cancer Res. 2014;20(4):799-803.

[8] Wirbel J, Pyl PT, Kartal E, et al. Meta-analysis of fecal metagenomes reveals global microbial signatures specific for colorectal cancer. Nat Med. 2019;25(4):679-689.

[9] Thomas AM, Manghi P, Asnicar F, et al. Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. Nat Med. 2019;25(4):667-678.

[10] Rubinstein MR, Wang X, Liu W, et al. Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/β-catenin signaling via its FadA adhesin. Cell Host Microbe. 2013;14(2):195-206.

[11] Bullman S, Pedamallu CS, Bhatt DL, et al. Analysis of Fusobacterium persistence and antibiotic response in colorectal cancer. Science. 2017;358(6369):1443-1448.

[12] Mori G, Pasca MR. Oral Cavity and Gut Microbiome: Is There Any Connection with Oral Squamous Cell Carcinoma and Colorectal Cancer? Int J Mol Sci. 2021;22(21):11406.

[13] Encarnação JC, Abrantes AM, Pires AS, Botelho MF. Revisit dietary fiber on colorectal cancer: butyrate and its role on prevention and treatment. Eur J Nutr. 2015;54(4):517-529.

[14] Amir A, McDonald D, Navas-Molina JA, et al. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. mSystems. 2017;2(2):e00191-16.

[15] Lin H, Peddada SD. Analysis of compositions of microbiomes with bias correction. Nat Commun. 2020;11(1):3514.

[16] Mima K, Nishihara R, Qian ZR, et al. Fusobacterium nucleatum in colorectal carcinoma tissue and patient prognosis. Gut. 2016;65(12):1973-1980.

[17] Hooi JKY, Lai WY, Ng WK, et al. Global Prevalence of Helicobacter pylori Infection: Systematic Review and Meta-Analysis. Gastroenterology. 2017;153(2):420-429.

[18] Flemer B, Lynch DB, Brown JMR, et al. Tumour-associated and non-tumour-associated microbiota in colorectal cancer. Gut. 2017;66(4):633-643.

[19] Liang JQ, Li T, Nakatsu G, et al. A novel faecal Lachnospiraceae gene marker for the non-invasive diagnosis of colorectal adenoma and cancer. Gut. 2020;69(7):1248-1257.

[20] He Z, Gharaibeh RZ, Newsome RC, et al. Campylobacter jejuni promotes colorectal tumorigenesis through the action of cytolethal distending toxin. Gut. 2019;68(2):289-300.

[21] Sokol H, Pigneur B, Watterlot L, et al. Faecalibacteriumprausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. Proc Natl Acad Sci USA. 2008;105(43):16731-16736.

[22] Kelly BJ, Gross R, Bittinger K, et al. Power and sample-size estimation for microbiome studies using pairwise distances and PERMANOVA. Bioinformatics. 2015;31(15):2461-2468.

[23] Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37(8):852-857.

[24] Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(D1):D590-D596.

[25] Kostic AD, Gevers D, Pedamallu CS, et al. Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. Genome Res. 2012;22(2):292-298.

[26] Castellarin M, Warren RL, Freeman JD, et al. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res. 2012;22(2):299-306.

[27] Liu L, Tabung FK, Zhang X, et al. Diets promoting colon inflammation associate with risk of colorectal carcinomas containing Fusobacterium nucleatum. Clin Gastroenterol Hepatol. 2018;16(10):1622-1631.

[28] Nakatsu G, Li X, Zhou H, et al. Gut mucosal microbiome across stages of colorectal carcinogenesis. Nat Commun. 2015;6:8727.

[29] Wong SH, Kwong TNY, Chow TC, et al. Quantitation of faecal Fusobacterium improves faecal immunochemical test in detecting advanced colorectal neoplasia. Gut. 2017;66(8):1441-1448.

[30] Bäckhed F, Roswall J, Peng Y, et al. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. Cell Host Microbe. 2015;17(5):690-703.

Copyright & License:

© Authors retain the copyright of this article. This work is published under the Creative Commons Attribution 4.0 International License (CC BY 4.0), permitting unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.