International Journal of Biomedical Investigation

journal homepage: https://ijbi.edwiserinternational.com

Original Article

Gut Microbial Dysbiosis in Colorectal Cancer: Insights from Shotgun Metagenomic Profiling

Raghavendra S*, Pooja M and Ramachandra Prasad LA

Department of Life Sciences, School of Sciences, Garden City University, Bengaluru, Karnataka, India.

ARTICLE INFO

Received 16 December 2025
Revised 10 January 2026
Available Online 23 January 2026

Keywords:
Gut microbiome
Colorectal cancer
Microbial dysbiosis
Differential abundance
Shotgun metagenomics

ABSTRACT

Objective: Changes in the gut microbiome have been progressively associated with colorectal cancer (CRC), but there is still a dearth of population-specific metagenomic data. The purpose of this work was to use shotgun metagenomic profiling to assess gut microbial dysbiosis linked to colorectal cancer and to find compositional changes that may have biological significance.

Methods: The NCBI Sequence Read Archive's publically accessible colorectal cancer and healthy control dataset provided the stool shotgun metagenomic samples used in this investigation. For analysis, a balanced selection of CRC and healthy control samples was chosen. Quality control, host read removal, and taxonomy categorization using Kraken2 with Bracken-based abundance calculation were applied to sequencing reads to get microbiome profiles followed by Differential abundance analysis of the profiles which was carried out using the DESeq2 framework and microbial community diversity was evaluated using α- and β-diversity measures.

Results: There was little indication of a decline in global diversity, and alpha-diversity indices revealed microbial richness and evenness that were nearly equal among CRC and control samples. However, a statistically significant variation in the overall community composition across groups was found using β-diversity analysis. Differential abundance testing revealed a clear microbial signature associated with colorectal cancer (CRC) samples in which opportunistic, pro-inflammatory and mucosa-associated taxa were observed to be significantly enriched and commensal beneficials such as Bifidobacterium sp. diminished. These findings suggested a significant restructuring of the gut microbiota rather than just an overall diversity decrease.

Conclusion: Various alterations in gut microbiome composition were associated with colorectal cancer. These changes reflected a consistent pattern characterized by reduction of saccharolytic commensals and increase in taxa implicated with anaerobic and proteolytic metabolism which could be candidate microbiome-based biomarkers for CRC.

Introduction

Colorectal cancer (CRC) represents a significant burden to world health, with over 1.9 million new cases and nearly 930,000 deaths estimated in 2020 worldwide, ranking second for cancer related mortality across the globe. These grim statistics highlight the extent and public health significance of the CRC burden in particular as the global burden is projected to rise with estimates that by 2040, new cases of CRC could be at 3.2 million annually with deaths rising to around 1.6 million a year. [1,2] Although genetics along with lifestyle/environmental factors (e.g., diet, obesity and inflammation) have long been recognized as major risk factors in CRC development, emerging evidence over the last decade has also implicated the gut microbial community which occupies this complex ecosystem. [3,4]

The human intestinal tract is inhabited by a dense and broadly variable micro-organismic community consisting of bacteria, archaea, viri and fungi collectively named gut microbiota. Under physiological conditions, this complex ecosystem fulfils vital tasks such as digestion and fermentation of nutrients from diet, generation of short-chain fatty acids (SCFAs), nutrition of colonic epithelial cells, integrity-supporting effect on mucosal barriers, immunomodulation and pathogen exclusion. [5,6] Of the SCFAs, those produced by healthful commensal bacteria, especially butyrate are significant in that they are the main fuel for healthy colonic epithelial cells and help to control inflammation and cell proliferation. [7]

Perturbations of intestinal ecosystem state, a condition commonly called "gut dysbiosis" have recently been implicated in the pathogenesis of colorectal tumour formation. [4] One common feature of dysbiosis in CRC is a decreased diversity of bacteria population and loss of protective taxa (e.g. butyrate-producing bacteria) as well as an overgrowth or enrichment of potentially harmful and pro-inflammatory microorganisms. [8,9] Several species and genera of bacteria have been repeatedly identified in association with CRC including Fusobacterium nucleatum, Bacteroides fragilis and Escherichia coli suggesting that they may contribute to the carcinogenesis through direct production of toxins, induction of chronic inflammation, disturbance of the mucosal barrier, induction or disruption modulation of inflammatory response or by generating genotoxic or pro-carcinogenic metabolic products. [10,11] Our ability to characterize gut microbiota at high taxonomic resolution i.e. species and even strain level and to infer functional potential by profiling microbial genes and

metabolic pathways has been transformed over the past few years by high-throughput sequencing, particularly shotgun metagenomic sequencing [3,12]. When compared to 16S-rRNA gene sequencing, metagenomics offers a more thorough and objective perspective of microbial populations, enhances repeatability and permits examination of microbial gene-level correlations with colorectal cancer. In fact, across many populations and cohorts, metagenomic investigations have found repeatable microbial fingerprints and microbial functional modules linked to colorectal cancer. [12]

Populations from Central or Eastern Europe, including Poland, have been included in relatively few metagenomic studies of colorectal cancer and the extent to which microbial signatures derived from Western-European or global cohorts apply to such populations remains understudied.

The majority of metagenomic studies of colorectal cancer (CRC) have been carried out in Western regions like North America and Western Europe. (13) Examining whether CRC-associated microbial signatures found elsewhere hold true in Polish cohorts or whether distinct, population-specific microbiome alterations exist is crucial because dietary patterns, lifestyle, environmental exposures and genetic background can vary significantly even within Europe. [14]

In this regard, the current work which examines the stool metagenomes of CRC patients and healthy controls attempts to support this endeavour by offering insights into changes in the gut microbiome linked to CRC in our group. This study characterizes microbial community structure, diversity, and differentially abundant taxa using a rigorous bioinformatics workflow, with findings interpreted in relation to known microbial mechanisms involved in colorectal cancer. The results aim to highlight microbiome-based biomarkers and microbial contributors to CRC in this population supporting future work on early detection, prevention, and microbiome-targeted interventions.

*Corresponding author: Raghavendra S, Department of Life Sciences, School of Sciences, Garden City University, Bengaluru, Karnataka, India.

https://doi.org/10.31531/2581-4745.1000175

Methodology

Data Acquisition and Sample Description

The NCBI SRA (Sequence Read Archive) metagenomic shotgun sequencing data used in this study are available under BioProject accession PRJNA1116523. Whole-genome shotgun metagenomic sequencing was performed to profile the human gut microbiome of patients with various types of cancers, in addition to colorectal cancer (CRC), and their corresponding age-sex-matched healthy controls. [15]

The NCBI SRA and BioSample records provided detailed metadata related to the BioProject. The dataset contained121 normal control samples and 39 CRC samples in it. 15 CRC samples and 15 normal control samples were chosen with matching age and sex to get balanced group size and adequate statistical power for differential abundance and diversity studies.The BioSample metadata's sample descriptions were used to confirm the clinical grouping. The NCBI's SRA toolkit was used to obtain the raw paired-end FASTQ files for the chosen samples, which were then used as the input for quality control, host read removal, taxonomy classification and other analyses. [16]

Quality Control and Pre-processing of Sequencing Reads

Raw paired-end FASTQ files obtained from the SRA were subjected to standard quality control and pre-processing steps prior to downstream metagenomic analysis. Adapter removal and quality filtering were performed using fastp which also generated HTML and JSON quality reports summarizing read quality and filtering statistics [17]. Read quality metrics were further assessed using FastQC to evaluate per-base sequence quality, GC content, sequence duplication levels and potential adapter contamination [18]. The resulting cleaned and high-quality reads were subsequently used for all downstream metagenomic processing steps.

Host Read Removal

The trimmed paired-end reads obtained from preprocessing were aligned against the Homo sapiens reference genome (hg38) using Bowtie2 in order to eliminate human derived sequences before microbial characterization. According to Langmead and Salzberg [19], Bowtie2 uses a quick gapped-alignment technique to effectively identify reads coming from the host genome. Only the unmapped read pairs were kept after reads that matched the human reference were eliminated. The subsequent processes used these non-host reads as their input.

Taxonomic Classification Using Kraken2 and Bracken

Non-host reads recovered after human-sequence removal were taxonomic classified using Kraken2, a k-mer based metagenomic classifier that provides taxonomic labels by matching query k-mers to a pre-built reference database [20]. The MiniKraken2 database was used for classification, resulting in a sample-level taxonomic report as well as a comprehensive output file. The Kraken2 report files were subjected to Bracken (Bayesian Re estimation of Abundances after Classification with Kraken) which produced revised abundance tables based on probabilistic re-allocation of k-mer assignments in order to increase the accuracy of species-level abundance estimations [21].

Construction of Taxonomic Abundance Matrices

Bracken produced species-level abundance data for every sample, which were combined into a single count matrix for further examination. A bespoke Python script that made use of the pandas module was used to parse Bracken output files. The fields relating to the fractional abundance and the estimated number of reads allocated to each species were retrieved for each sample. An abundance matrix with estimated read counts for each species across all samples was created by combining these per-sample tables by species name. Zeros were used to fill in the missing values caused by species that were absent from particular samples. For additional investigation, the resulting matrices were exported as tab-delimited files. [22]

Community Diversity and Ordination Analysis

The cleaned taxonomic count matrix and associated sample metadata were used to calculate microbial diversity metrics. The phyloseq framework was used to calculate alpha-diversity indicators, including observed richness, and non-parametric statistical tests were used in the same setting to evaluate group-wise comparisons between colorectal cancer and healthy control samples. [23] Bray-Curtis dissimilarity was used to measure beta diversity, and Principal Coordinates Analysis (PCoA) was used for ordination to show general variations in community structure among samples. The vegan package's implementation of PERMANOVA was used to statistically evaluate between-group compositional differences. [24]

Differential Abundance Analysis

Differential abundance testing between colorectal cancer and healthy control samples was performed using the DESeq2 framework, which models count data using the negative binomial distribution and applies shrinkage-based estimation of dispersion and fold change to improve stability in high-dimensional datasets. [25] Prior to analysis, the Bracken-derived count matrix and corresponding sample metadata were imported into R and converted into a DESeq2 dataset. The design formula included the clinical condition as the primary explanatory variable. Subsequently, the normalization of size-factors was carried out to adjust for differences in sequencing depth between samples and the model fitted to produce log2(fold-change) estimates and Wald test P-values.

The Benjamini–Hochberg false-discovery rate (FDR) was used for multiple-testing correction, and taxa were considered to be significantly differentially abundant when their adjusted P-value was <0.05. Normalized counts were exported for heatmap-based examination of important taxa, and variance-stabilizing transformation (VST) was utilized for exploratory visualization, including principal component analysis. These findings formed the basis for downstream biological interpretation by identifying microbial taxa enriched in CRC or healthy controls. [26]

Results

Alpha Diversity Analysis

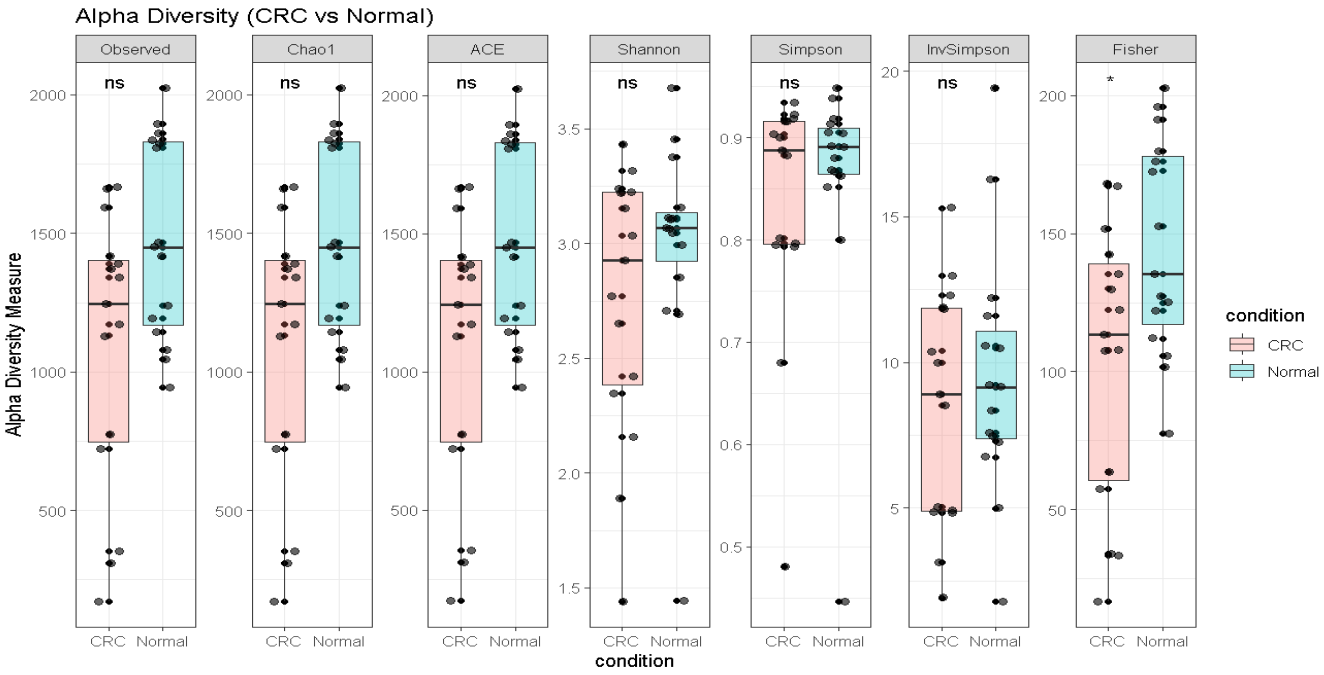

Figure 1: Alpha diversity measures comparing CRC and normal samples.

Multiple richness and evenness estimators were used to quantify alpha diversity like the Chao1, ACE, Shannon, Simpson and Inverse Simpson indices. No statistically significant differences were seen between colorectal cancer and normal samples across these parameters suggesting that the total microbial richness and community evenness of both the groups are still comparable. Although changes in community composition may still occur and be investigated in later analyses, this pattern implies that colorectal cancer does not decrease overall within sample variety.

In contrast, the Fisher diversity index showed a significant reduction in Colorectal cancer samples relative to healthy controls, reflecting a modest loss of low abundance taxa and a less complex community structure. Such a pattern is consistent with reports that CRC associated dysbiosis often involves depletion of specific commensals alongside enrichment of disease associated taxa rather than a broad collapse of diversity.

Beta Diversity and Community Structure

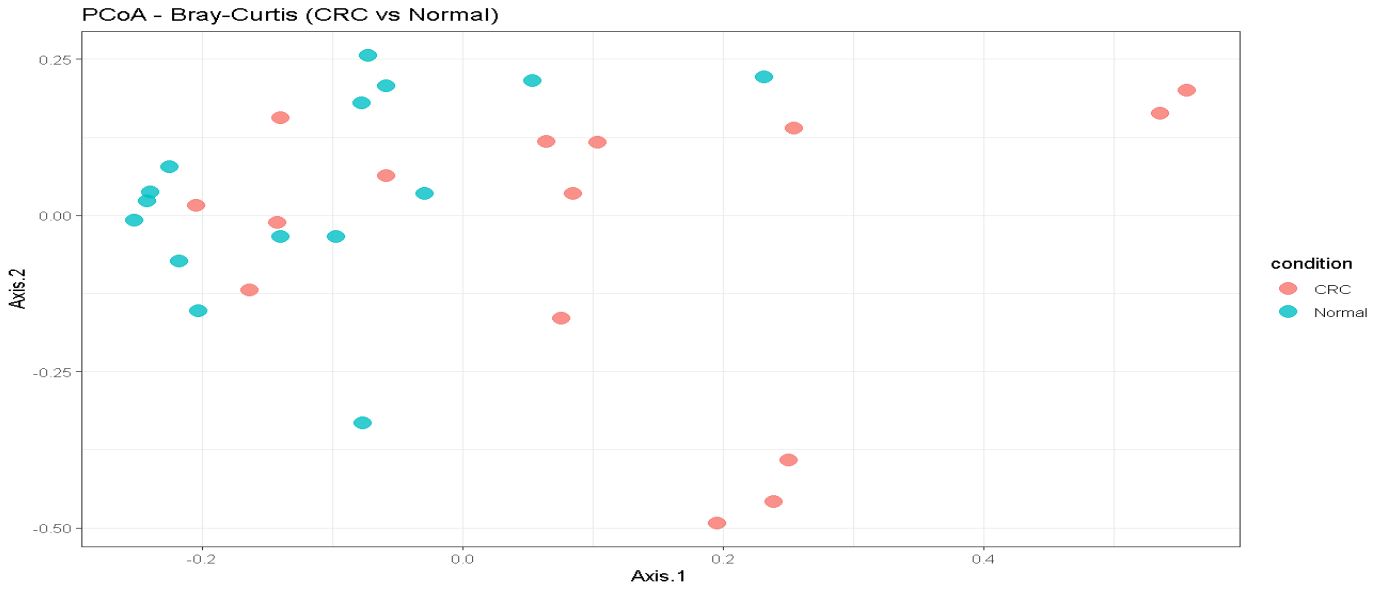

Figure 2: Principal Coordinated Analysis (PCoA) based on Bray Curtis dissimilarity.

In order to compare the microbial community composition of colorectal cancer (CRC) samples with healthy control samples, beta diversity was measured using Bray Curtis dissimilarity. A partial grouping of samples according to clinical condition was shown by the Principal Coordinates Analysis (PCoA) of the Bray Curtis matrix, with several normal samples clustering toward the left and numerous CRC samples separating toward the right of the ordination space. Although there was still some overlap between the groups, the overall distribution revealed quantitative compositional variations associated with CRC.

PERMANOVA confirmed these visual patterns, showing a statistically significant difference in community structure between the two groups ($p = 0.004$), with the condition accounting for about 7.6% of the observed variance. This degree of variation is in line with human gut microbiome datasets, which generally show considerable inter-individual variation. All of these findings show that, even while within-sample richness is mostly maintained, CRC is linked to a discernible change in community makeup.

Differential Abundance Analysis

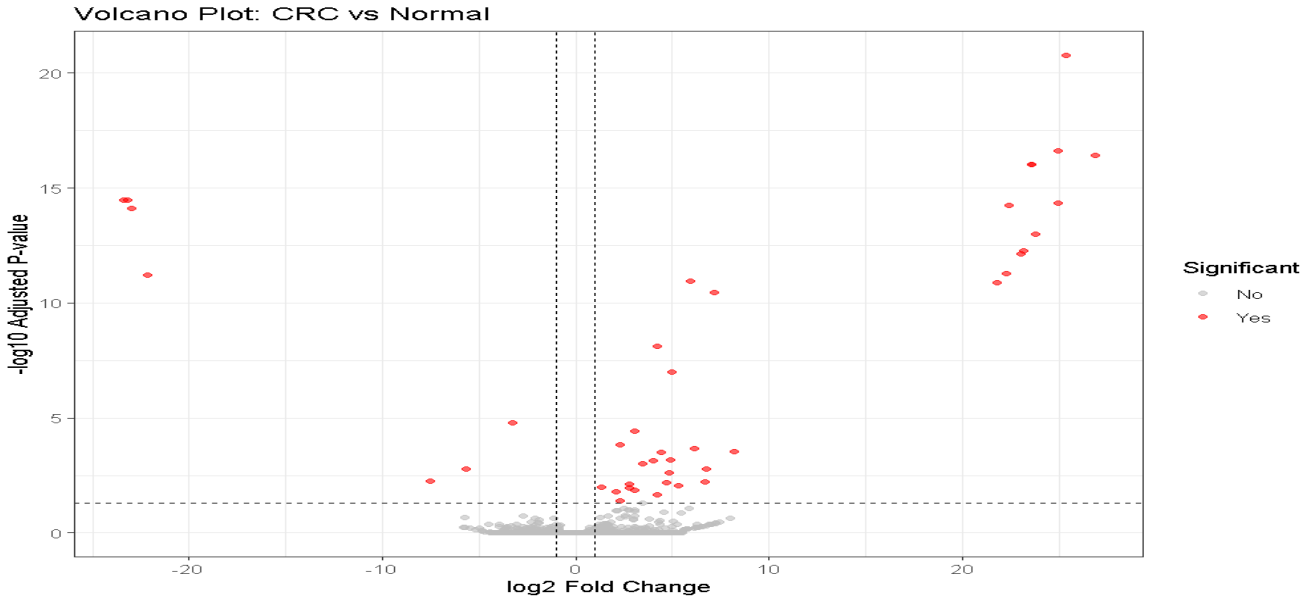

Figure 3: Volcano plot of differentially abundant taxa between CRC and normal samples.

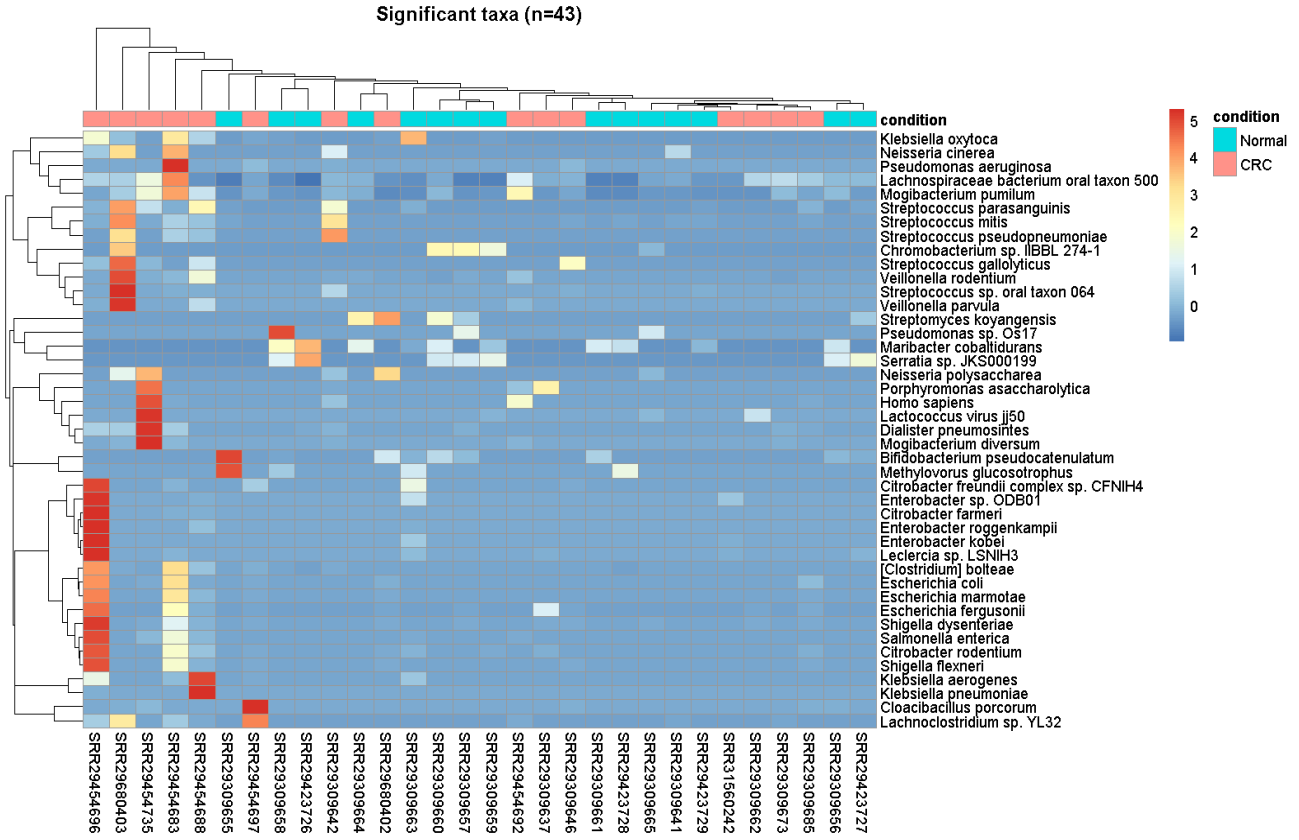

Figure 4: Heatmap of significantly differentially abundant taxa (CRC vs Normal).

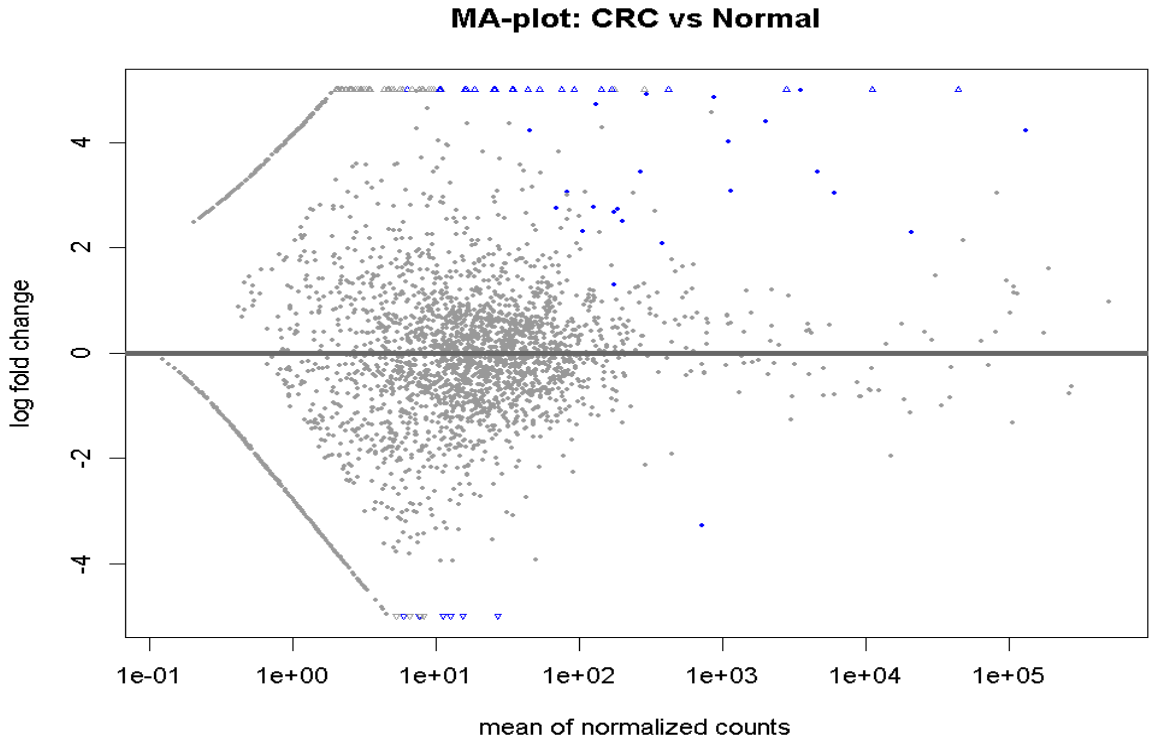

Figure 5: MA plot of log2 fold changes versus mean normalized abundance (CRC vs normal).

Differential abundance analysis using DESeq2 identified a total of 43 microbial taxa that were significantly altered between colorectal cancer and healthy control samples after multiple-testing correction with adjusted P-value < 0.05. 36 taxa were found to be enriched in CRC samples whereas 7 taxa were found to be depleted in CRC samples relative to the controls indicating a composition difference associated with disease status.

The taxa enriched in CRC samples included Enterobacter roggenkampii, Streptococcus gallolyticus, Enterobacter kobei, Klebsiella pneumoniae, Klebsiella oxytoca, Escherichia coli and Shigella species among other taxa belonging to Enterobacteriaceae and Streptococcaceae families. The large positive log2 fold increases and very significant adjusted P-values of many of these species indicated strong enrichment in CRC samples. In addition, certain oral-associated like Streptococcus mitis and opportunist taxa such as Neisseria; Porphyromonas asaccharolytica; Streptococcus parasanguinis were enriched in CRC cases indicating ectopic colonization and community remodeling in the intestinal environment of patients with CRC. Together, these taxa have been reported to flourish with inflammation and hypoxia and are able to generate virulence factors, endotoxins and genotoxic metabolites. Their enrichment contributes to a pro-inflammatory, epithelial-disruptive microbiome in CRC and can also lead to tumour initiation and progression via chronic immune stimulation and DNA damage. [27,28]

On the other hand, a smaller group of taxa demonstrated notable depletion in CRC samples. These included Serratia sp., Bifidobacterium pseudocatenulatum and a number of low-abundance or environmental species that showed negative log2 fold changes. The loss of protective commensals usually described in CRC-associated dysbiosis is consistent with the decrease of Bifidobacterium pseudocatenulatum which is a species commonly associated with advantageous metabolic and immunomodulatory functions in the gut. Bifidobacterium species are known to be producing short-chain fatty acids and are also involved in maintaining the integrity of epithelial barrier, regulation of mucosal immune responses and suppressing inflammation. Their depletion may reduce colonization resistance and weaken host microbiome homeostasis thereby creating conditions that favour the expansion of opportunistic and pro-inflammatory taxa in CRC. [29]

Condition-wise Abundance Profiles of Significant Taxa

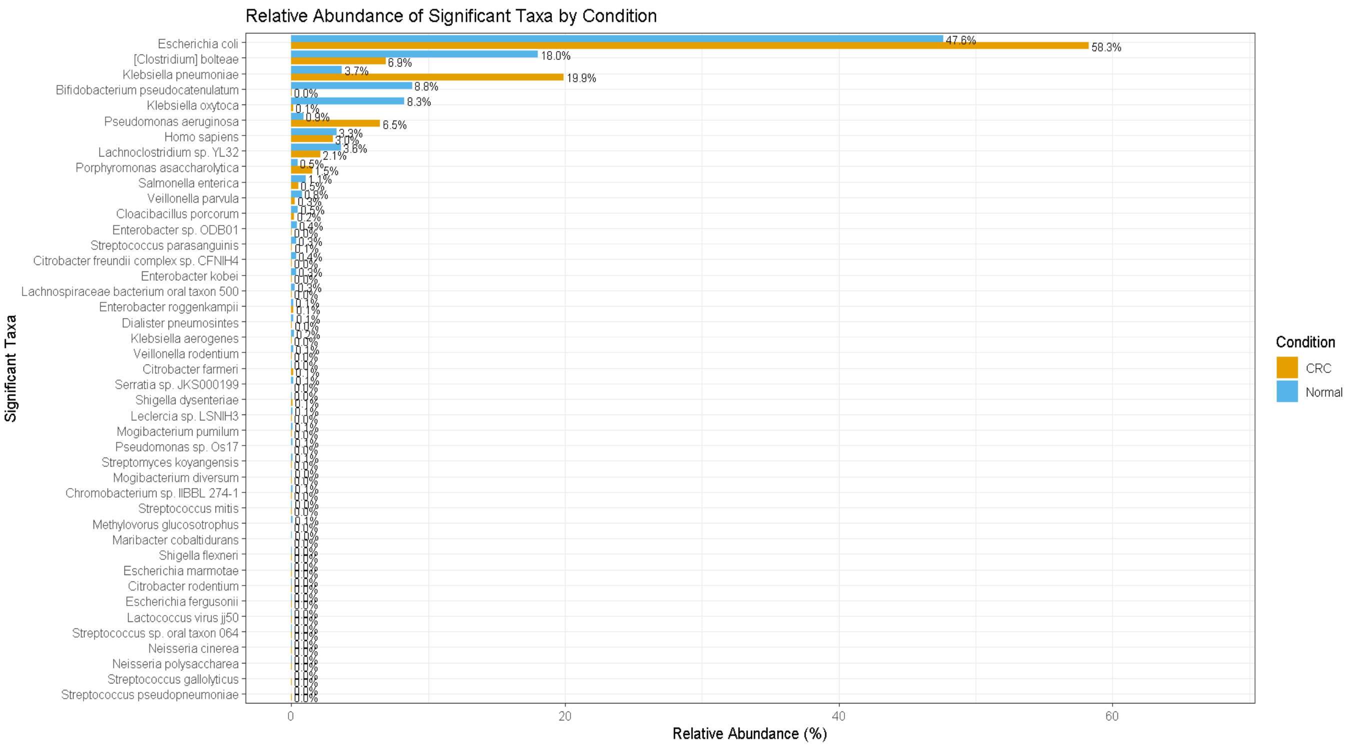

Figure 6: Relative abundance of significant taxa in CRC and normal samples.

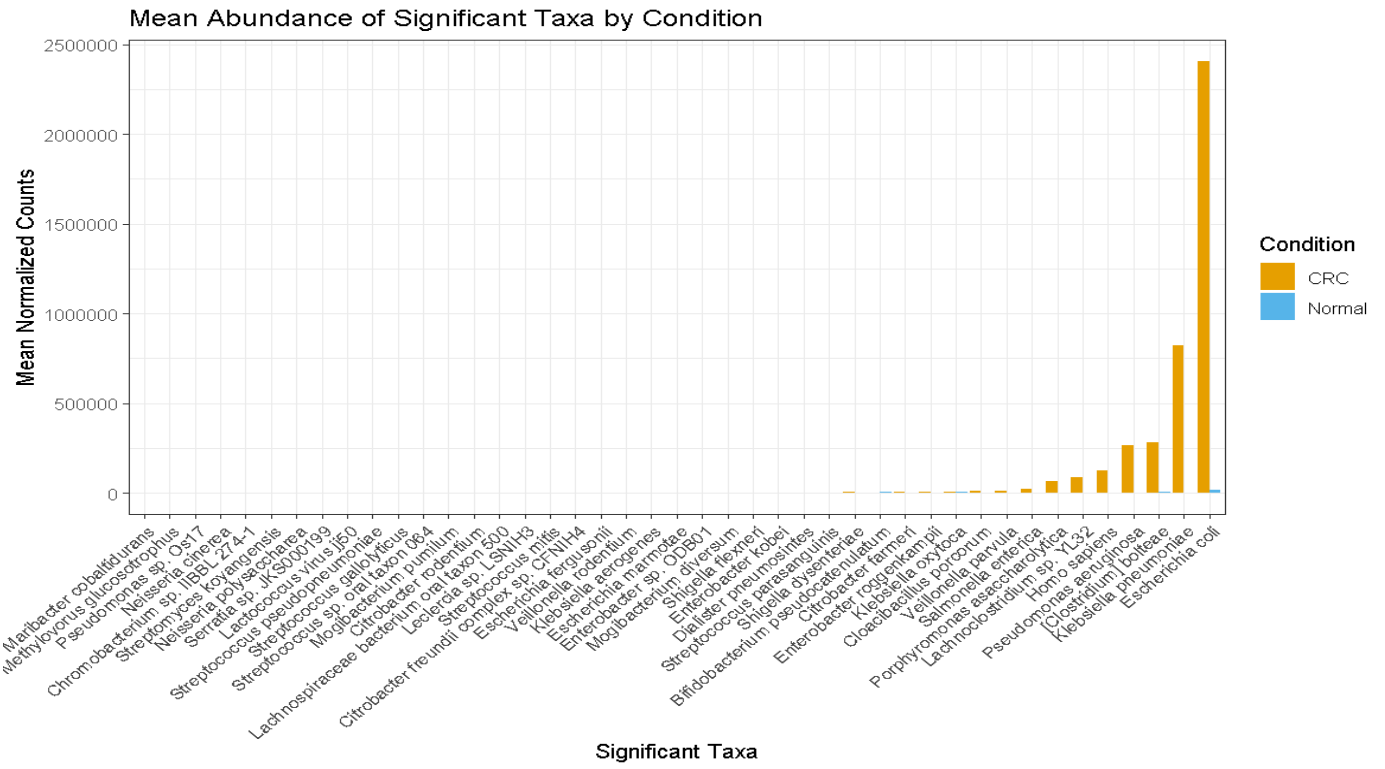

Figure 7: Mean normalized abundance of significant taxa by condition (CRC vs Normal).

Condition wise abundance profiling of the significantly altered taxa revealed marked differences in their relative and absolute contributions between colorectal cancer and healthy control samples. The abundance of a minority of CRC-enriched taxa represented a significantly larger proportion of overall bacterial composition amongst significant taxa in the CRC group with these same taxa observed at much lower abundances within the control group.

Specifically, important genera such as Escherichia coli, Klebsiella pneumoniae, Klebsiella oxytoca and Clostridium bolteae had a mean normalized count that were highly different in relative abundance percentages between CRC cases which demonstrated the dominance of only few taxa by number in diseased gut microbiome. However, healthy control samples exhibited a more balanced distribution of abundant taxa with relatively high abundance levels of commensal species e.g. Bifidobacterium pseudocatenulatum.

Relative abundance analysis further demonstrated that CRC samples were characterized by the expansion of a few highly abundant taxa rather than uniform increases across all significant species. This pattern highlights a restructuring of microbial community dominance in CRC complementing the differential abundance results and underscoring condition-specific shifts in microbial composition.

Discussion

Instead of a homogenous decrease of microbial diversity, the current metagenomic research shows a disease-associated reorganization of the gut microbiome in colorectal cancer (CRC) characterized by coordinated taxonomic alterations. CRC is linked to qualitative alterations in microbial community structure, as evidenced by the considerable compositional variations shown by β-diversity analysis and differential abundance testing, even while α-diversity metrics were generally similar between CRC and control samples.

The enrichment of facultative anaerobic and opportunistic taxa such as members of the Enterobacteriaceae and Streptococcaceae families is a notable characteristic of the CRC-associated microbiome seen in this study. These taxa's expansion has been repeatedly documented in CRC populations and is frequently associated with altered oxygen gradients in the tumour microenvironment impaired epithelial barriers and inflammatory gut environments [4, 27]. These organisms may take advantage of host-derived nutrients generated after epithelium injury and are well suited to inflammatory situations.

The increase in oral-associated and mucosa-adherent bacteria indicates a disruption of colonization resistance

and altered niche availability in the CRC gut. Many taxa have been associated with immune regulation, mucin degradation and epithelial interaction mechanisms that could support microbial survival in tumour-related environments [30]. These ecological shifts are thought to benefit bacteria capable of thriving in low-oxygen and high-nitrate settings commonly found in inflammatory colonic tissue

In parallel, depletion of beneficial commensal taxa particularly short-chain fatty acid producing bacteria points toward a functional reprogramming of the gut microbiome. Reduced abundance of taxa associated with saccharolytic metabolism may lead to diminished production of butyrate and related metabolites that normally support epithelial integrity and exert anti-inflammatory effects. Loss of these functions has been linked to impaired barrier maintenance and increased susceptibility to microbial translocation. [31]

All of these compositional alterations point to a transition from fermentation powered by carbohydrates to anaerobic and proteolytic metabolic activities. Ammonia, hydrogen sulfide, and secondary bile acids all of which have been linked to genotoxicity and epithelial stress—can accumulate as a result of this metabolic reorientation. The intestinal mucosa may experience oxidative stress, DNA damage, and persistent immunological activation as a result of increased exposure to these chemicals. [32]

When considered collectively, the results provide credence to a theory that links colorectal cancer (CRC) to a pro-tumorigenic microbiome configuration that weakens epithelial defense, increases inflammation and modifies metabolic outputs. The idea that microbial community composition rather than diversity alone plays a crucial role in disease-associated dysbiosis is reinforced by the fact that CRC appears to involve selective enrichment and depletion of functionally relevant taxa rather than a global collapse of microbial diversity.

Conclusion

This study provides a metagenomic investigation of gut microbiota alterations associated with colorectal cancer using publicly available shotgun sequencing data. The microbial community's composition was clearly altered in colorectal cancer samples, even if the alpha diversity was largely unaltered. This shift was brought about by an increase in opportunistic, pro-inflammatory, and perhaps genotoxic taxa and a decrease in beneficial commensal bacteria. These findings imply that the characteristic of dysbiosis associated with colorectal cancer is compositional reorganization rather than a general decrease of microbial diversity.

A microbial environment favourable to oxidative stress, epithelial dysfunction, and tumour promoting inflammation is supported by the observed enrichment of mucosa-invasive and inflammation-associated microorganisms as well as the depletion of taxa associated with the production of short-chain fatty acids and the maintenance of the epithelial barrier. All of these findings support the idea that a shift toward a pro-tumorigenic gut microbiota condition occurs concurrently with colorectal cancer.

This work emphasizes the significance of metagenomic profiling for comprehending disease associated microbiome modifications and identifies microbial signatures that may contribute to colorectal cancer pathophysiology by combining strong bioinformatics processing with differential abundance analysis. Future validation studies investigating microbiome-based biomarkers or treatment approaches for colorectal cancer may consider the found taxa.

Author Contributions

All authors contributed equally to this research. All authors read and approved the final manuscript.

Acknowledgement

The authors are thankful to the Department of Life Sciences, Garden City University, Bengaluru, Karnataka, India for supporting to carry out this research.

Conflict of Interest

The authors declare no conflict of interest.

Funding

The authors did not receive any fundings from any private or government sources.

Ethical Approvals

This study does not involve experiments on animals or human subjects.

References

1. Roshandel G, Ghasemi-Kebria F, Malekzadeh R. Colorectal cancer: epidemiology, risk factors, and prevention. Cancers. 2024 Apr 17;16(8):1530.
2. Morgan E, Arnold M, Gini A, Lorenzoni V, Cabasag CJ, Laversanne M, Vignat J, Ferlay J, Murphy N, Bray F. Global burden of colorectal

cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. Gut. 2023 Feb 1;72(2):338-44.
3. Rawla P, Sunkara T, Barsouk A. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. Gastroenterology Review/Przegląd Gastroenterologiczny. 2019 Jan 6;14(2):89-103.
4. Montalban-Arques A, Scharl M. Intestinal microbiota and colorectal carcinoma: Implications for pathogenesis, diagnosis, and therapy. EBioMedicine. 2019 Oct 1;48:648-55.
5. Fan X, Jin Y, Chen G, Ma X, Zhang L. Gut microbiota dysbiosis drives the development of colorectal cancer. Digestion. 2021 Jun 29;102(4):508-15.
6. Pandey H, Tang DW, Wong SH, Lal D. Gut microbiota in colorectal cancer: biological role and therapeutic opportunities. Cancers. 2023 Jan 30;15(3):866.
7. Den Besten G, Van Eunen K, Groen AK, Venema K, Reijngoud DJ, Bakker BM. The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. Journal of lipid research. 2013 Sep 1;54(9):2325-40.
8. Gagnière J, Raisch J, Veziant J, Barnich N, Bonnet R, Buc E, Bringer MA, Pezet D, Bonnet M. Gut microbiota imbalance and colorectal cancer. World journal of gastroenterology. 2016 Jan 14;22(2):501.
9. Yu AI, Chen GY. The gut microbiome and colorectal cancer. InInflammation, Infection, and Microbiome in Cancers: Evidence, Mechanisms, and Implications 2021 Apr 21 (pp. 63-96). Cham: Springer International Publishing.
10. Vogtmann E, Hua X, Zeller G, Sunagawa S, Voigt AY, Hercog R, Goedert JJ, Shi J, Bork P, Sinha R. Colorectal cancer and the human gut microbiome: reproducibility with whole-genome shotgun sequencing. PloS one. 2016 May 12;11(5):e0155362.
11. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, Clancy TE, Chung DC, Lochhead P, Hold GL, El-Omar EM. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. Cell host & microbe. 2013 Aug 14;14(2):207-15.
12. Torshizi Esfahani A, Zafarjafarzadeh N, Vakili F, Bizhanpour A, Mashaollahi A, Karimi Kordestani B, Baratinamin M, Mohammadpour S. Gut microbiome in colorectal cancer: metagenomics from bench to bedside. JNCI Cancer Spectrum. 2025 Jun;9(3):pkaf026.
13. Gamage BD, Ranasinghe D, Sahankumari AG, Malavige GN. Metagenomic analysis of colonic tissue and stool microbiome in patients with colorectal cancer in a South Asian population. BMC cancer. 2024 Sep 10;24(1):1124.
14. Singh G, Chaudhry Z, Boyadzhyan A, Sasaninia K, Rai V. Dysbiosis and colorectal cancer: conducive factors, biological and molecular role, and therapeutic prospectives. Exploration of Targeted Anti-tumor Therapy. 2025 Jun 27;6:1002329.
15. National Center for Biotechnology Information (NCBI). BioProject PRJNA1116523: Human gut metagenome sequencing dataset. Bethesda (MD): NCBI; 2024. Available from: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1116523
16. Kulecka, M., Czarnowski, P., Bałabas, A., Turkot, M., Kruczkowska-Tarantowicz, K., Żeber-Lubecka, N., Dąbrowska, M., Paszkiewicz-Kozik, E., Walewski, J., Ługowska, I. and Koseła-Paterczyk, H., 2024. Microbial and metabolic gut profiling across seven malignancies identifies fecal Faecalibacillus intestinalis and formic acid as commonly altered in Cancer patients. International Journal of Molecular Sciences, 25(15), p.8026.
17. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34: i884–i890 [Internet]. 2018
18. Andrews S. FastQC: A quality control tool for high throughput sequence data [Internet]. Babraham Institute; 2010 [cited 2025 Jan 10]. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
19. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012 Apr;9(4):357-9.
20. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome biology. 2019 Nov 28;20(1):257.
21. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species abundance in metagenomics data. PeerJ Computer Science. 2017 Jan 2;3:e104.
22. McKinney W. Data structures for statistical computing in Python. scipy. 2010 Jun 28;445(1):51-6.
23. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PloS one. 2013 Apr 22;8(4):e61217.
24. Dixon P. VEGAN, a package of R functions for community ecology. Journal of vegetation science. 2003 Dec;14(6):927-30.
25. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-

seq data with DESeq2. Genome biology. 2014 Dec 5;15(12):550.
26. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological). 1995 Jan;57(1):289-300.
27. Garrett WS. Cancer and the microbiota. Science. 2015 Apr 3;348(6230):80-6.
28. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, Amiot A, Böhm J, Brunetti F, Habermann N, Hercog R. Potential of fecal microbiota for early-stage detection of colorectal cancer. Molecular systems biology. 2014 Nov;10(11):766.
29. O'callaghan A, Van Sinderen D. Bifidobacteria and their role as members of the human gut microbiota. Frontiers in microbiology. 2016 Jun 15;7:925.
30. Flemer B, Lynch DB, Brown JM, Jeffery IB, Ryan FJ, Claesson MJ, O'Riordain M, Shanahan F, O'Toole PW. Tumour-associated and non-tumour-associated microbiota in colorectal cancer. Gut. 2017 Apr 1;66(4):633-43.
31. Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. Nature reviews microbiology. 2014 Oct;12(10):661-72.
32. O'keefe SJ. Diet, microorganisms and their metabolites, and colon cancer. Nature reviews Gastroenterology & hepatology. 2016 Dec;13(12):691-706.