A Comparative Study of Gut Microbiota in the Healthy Elderly and the Elderly with Colorectal Cancer

Miss Nutta Iadsee

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Medical Microbiology
Medical Microbiology,Interdisciplinary Program
GRADUATE SCHOOL
Chulalongkorn University
Academic Year 2020

การศึกษาเปรียบเทียบไมโครไบโอตาในลำไส้ของผู้สูงวัยสุขภาพดีและผู้สูงวัยที่เป็นมะเร็งลำไส้ใหญ่และทวารหนัก

น.ส.ณัฏฐา เอียดสี

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาจุลชีววิทยาทางการแพทย์ สหสาขาวิชาจุลชีววิทยาทางการแพทย์
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2563

Thesis Title	A Comparative Study of Gut Microbiota in the Healthy Elderly and the Elderly with Colorectal Cancer
By	Miss Nutta Iadsee
Field of Study	Medical Microbiology
Thesis Advisor	Associate Professor KANITHA PATARAKUL, M.D. Ph.D.
Thesis Co Advisor	Associate Professor NARAPORN SOMBOONNA, Ph.D.

Accepted by the GRADUATE SCHOOL, Chulalongkorn University in Partial Fulfillment of the Requirement for the Master of Science

..... Dean of the GRADUATE SCHOOL
(Associate Professor THUMNOON NHUJAK, Ph.D.)

THESIS COMMITTEE

..... Chairman
(Assistant Professor DIREKRIT CHIEWCHENGCHOL, M.D. Ph.D.)

..... Thesis Advisor
(Associate Professor KANITHA PATARAKUL, M.D. Ph.D.)

..... Thesis Co-Advisor
(Associate Professor NARAPORN SOMBOONNA, Ph.D.)

..... Examiner
(KANNIKA KHANTASUP, Ph.D.)

..... External Examiner
(Assistant Professor Iyarit Thaipisuttikul, M.D.,Ph.D.)

ณัฏฐา เอียดสี : การศึกษาเปรียบเทียบไมโครไบโอตาในลำไส้ของผู้สูงวัยสุขภาพดีและผู้สูงวัยที่เป็นมะเร็งลำไส้ใหญ่และทวารหนัก. (A Comparative Study of Gut Microbiota in the Healthy Elderly and the Elderly with Colorectal Cancer) อ.ที่ปรึกษาหลัก : รศ. ดร. พญ.กนิษฐา ภัทรกุล, อ.ที่ปรึกษาร่วม : รศ. ดร.นราพร สมบูรณ์นะ

มะเร็งลำไส้ใหญ่และทวารหนักจัดเป็นหนึ่งในมะเร็งที่พบได้บ่อยเป็นอันดับสามของประชากรทั่วโลก การเสียสมดุลของจุลินทรีย์ในลำไส้มีความสัมพันธ์กับการเกิดมะเร็งลำไส้ใหญ่และทวารหนัก การศึกษานี้มีวัตถุประสงค์เพื่อเปรียบเทียบจุลินทรีย์ในลำไส้ของคนไทยอายุ 50 ปีขึ้นไป จำนวน 80 คน ระหว่างกลุ่มโรคมะเร็งลำไส้ใหญ่และทวารหนัก จำนวน 25 คน, กลุ่มมีติ่งเนื้อลำไส้ใหญ่ จำนวน 33 คน และกลุ่มควบคุม จำนวน 22 คน โดยทำการวิเคราะห์ลำดับนิวคลีโอไทด์บริเวณยีน 16S rRNA ทั้งในตัวอย่างชิ้นเนื้อและอุจจาระ และเพื่อหาปริมาณสุทธิของแบคทีเรียที่เกี่ยวข้องกับมะเร็งลำไส้ใหญ่ จำนวน 6 สายพันธุ์ ในตัวอย่างทั้งสองชนิด ได้แก่ *Fusobacterium nucleatum*, *Parvimonas micra*, *Streptococcus gallolyticus*, *Blautia* spp., *Fusicatenibacter saccharivorans* และแบคทีเรียที่มียีนสร้างสารพิษ colibactin ด้วยเทคนิคปฏิกิริยาลูกโซ่พอลีเมอเรสแบบเรียลไทม์ รวมทั้งแบคทีเรียเหล่านี้ยังถูกประเมินความสามารถในตรวจหาโรคมะเร็งลำไส้ใหญ่และติ่งเนื้อลำไส้ใหญ่ ผลการศึกษาพบว่าจุลินทรีย์ในอุจจาระสามารถแสดงให้เห็นถึงจุลินทรีย์ในลำไส้ที่อาศัยอยู่บนชั้นเยื่อเมือกได้เพียงบางส่วน และพบความแตกต่างอย่างมีนัยสำคัญของจุลินทรีย์บนเยื่อเมือกระหว่างกลุ่มมะเร็งลำไส้ใหญ่และกลุ่มควบคุม แต่ไม่พบความแตกต่างในกลุ่มติ่งเนื้อลำไส้ใหญ่เมื่อเทียบกับกลุ่มควบคุม ซึ่งพบการเพิ่มขึ้นอย่างเรียงตามลำดับจากติ่งเนื้อลำไส้ใหญ่และมะเร็งลำไส้ใหญ่ของแบคทีเรียจีนัส *Bacteroides* และ *Parabacteroides* ในขณะที่แบคทีเรียกลุ่มที่สร้างสาร butyrate อย่างจีนัส *Faecalibacterium* ลดลงอย่างมีนัยสำคัญในกลุ่มมะเร็งลำไส้ใหญ่ นอกจากนี้ขนาดเอฟเฟกต์การวิเคราะห์จำแนกเชิงเส้น (LEfSe) ระบุการเพิ่มขึ้นอย่างมีนัยสำคัญของแบคทีเรียก่อโรคแบบฉวยโอกาสอย่าง *Erysipelatoclostridium ramosum* ทั้งในตัวอย่างทั้งสองชนิดของกลุ่มมะเร็งลำไส้ใหญ่ แสดงให้เห็นว่าการเปลี่ยนแปลงของจุลินทรีย์ในลำไส้อาจมีส่วนเกี่ยวข้องกับการเกิดมะเร็งลำไส้ใหญ่และทวารหนัก นอกจากนี้ผลการศึกษาปริมาณของแบคทีเรียพบว่า *F. nucleatum* และ *P. micra* เพิ่มขึ้นอย่างมีนัยสำคัญในตัวอย่างทั้งสองชนิดของกลุ่มมะเร็งลำไส้ใหญ่ อีกทั้งการทดสอบแบคทีเรียทั้งสองสายพันธุ์ในอุจจาระร่วมกับการตรวจหาเลือดแฝงในอุจจาระ (FIT) ในการตรวจหามะเร็งลำไส้ใหญ่มีความไวร้อยละ 93.8 และความจำเพาะร้อยละ 95.2 ในขณะที่การทดสอบแบคทีเรียห้าสายพันธุ์ (ยกเว้น *S. gallolyticus*) ร่วมกับ FIT สามารถตรวจหาติ่งเนื้อลำไส้ใหญ่ด้วยความไวร้อยละ 83.3 และความจำเพาะร้อยละ 64.7 การศึกษานี้ชี้ให้เห็นว่า *E. ramosum* น่าจะเป็นตัวแทนของตัวบ่งชี้ทางชีวภาพที่จำเพาะต่อคนไทยและปริมาณของแบคทีเรียที่เกี่ยวข้องกับมะเร็งลำไส้ใหญ่ในอุจจาระอาจจะสามารถเสริมประสิทธิภาพในการตรวจหามะเร็งลำไส้ใหญ่และติ่งเนื้อลำไส้ใหญ่ให้กับชุดตรวจคัดกรองโรคในปัจจุบัน อย่างไรก็ตามตัวแทนตัวบ่งชี้ทางชีวภาพเหล่านี้ยังคงต้องการการทดสอบเพิ่มเติมในกลุ่มประชากรที่ใหญ่ขึ้น

สาขาวิชา	จุลชีววิทยาทางการแพทย์	ลายมือชื่อนิสิต .....
ปีการศึกษา	2563	ลายมือชื่อ อ.ที่ปรึกษาหลัก .....
		ลายมือชื่อ อ.ที่ปรึกษาร่วม .....

6187128920 : MAJOR MEDICAL MICROBIOLOGY
KEYWORD: Gut microbiota, Colorectal cancer

Nutta Iadsee : A Comparative Study of Gut Microbiota in the Healthy Elderly and the Elderly with Colorectal Cancer. Advisor: Assoc. Prof. KANITHA PATARAKUL, M.D. Ph.D. Co-advisor: Assoc. Prof. NARAPORN SOMBOONNA, Ph.D.

Colorectal cancer is one of the third most common cancer worldwide. Dysbiosis of the human gut microbiota has been linked to sporadic colorectal cancer (CRC). This study aimed to compare the gut microbiota profile of a total of 80 Thai volunteers, who were above 50 years old, among 25 CRC patients, 33 adenoma patients, and 22 healthy controls (HC). The 16S rRNA sequencing was utilized to characterize the gut microbiome in both mucosal tissue and stool samples. Moreover, absolute quantitative PCR (qPCR) assay was conducted to quantify six CRC-associated bacteria including *Fusobacterium nucleatum* (FN), *Parvimonas micra* (PM), colibactin positive strains (EC), *Streptococcus gallolyticus* (SG), *Blautia* spp. (Bla) and *Fusicatenibacter saccharivorans* (FS), in both sample types, and these bacteria were evaluated the performance in CRC and adenoma detections. The results suggested that the fecal microbiota only partially reflected gut microbiota on the mucus layer. The mucosal microbiota of the CRC patients and HC group differed significantly but no difference between adenoma and HC groups was observed. The stepwise increase of *Bacteroides* and *Parabacteroides* according to adenomas-carcinomas sequence were found whereas the butyrate-producing genus *Faecalibacterium* was significantly less abundant in CRC patients. Linear discriminant analysis effect size (LEfSe) showed a higher level of *Erysipelatoclostridium ramosum*, an opportunistic pathogen, in both sample types of CRC patients. The findings indicated the imbalance of gut microorganisms might be involved in CRC tumorigenesis. In addition, the qPCR assays revealed FN and PM were significantly overrepresented in both sample types of CRC subjects. The combined test of fecal FN and PM with qualitative fecal immunochemical test (FIT) could predict CRC with a sensitivity of 93.8% and a specificity of 95.2%. In a combined test of five fecal bacteria without SG together with the FIT, adenoma was detected with a sensitivity of 83.3% and a specificity of 64.7%. These results indicated *E. ramosum* may serve as a population-specific biomarker for CRC screening and the quantity of fecal bacteria could complement the current FIT in CRC and adenoma screening. Larger sample size is required for the validation of these candidate biomarkers in CRC and adenoma detections.

Field of Study:	Medical Microbiology	Student's Signature .....
Academic Year:	2020	Advisor's Signature .....
		Co-advisor's Signature .....

ACKNOWLEDGEMENTS

First and foremost, I cannot express enough my special thanks of gratitude to my kindly advisor Assoc. Prof. Kanitha Patarakul, M.D., Ph.D., Department of Microbiology, Faculty of Medicine, for her continued encouragement, knowledge, and precious advice. She was who gave me the learning opportunity to do this valuable project on the topic of gut microbiome study on colorectal cancer, which also challenged me to overcome a lot of struggles. I came to know about so many new things from this thesis and it can be adapted in my real life. My completion of this thesis could not have been accomplished if there was not for her dedication.

Secondly, I would also like to express my sincere gratitude to my co-advisor Assoc. Prof. Naraporn Somboonna, Ph.D., Department of Microbiology, Faculty of Science, for giving me invaluable guidance throughout my thesis.

I am extremely grateful to Natthaya Chuaypen, Ph.D., Department of Biochemistry, Faculty of Medicine, her immense teaching, and support helped me to complete this dissertation.

Besides, I would like to express my special thanks of gratitude to Assoc. Prof. Tanisa Patcharatrakul, M.D., Department of Gastroenterology, Faculty of Medicine, and Songphol Malakorn, M.D., Colorectal Division, Department of Surgery, Faculty of Medicine. My rare sample collection step would not be successful without their valuable support.

I also would like to acknowledge the thesis committee Assist. Prof. Direkrit Chiewchengchol, M.D., Ph.D., Kannika Khantasup, Ph.D., and Assist. Prof. Iyarit Thaipisuttikul, M.D., Ph.D., for their suggestion to complete this thesis.

I would like to extend my sincere thanks to Teerasit Techawiwattanaboon, Ph.D. who provided me with encouragement, dedication, knowledge, and patience throughout the duration of this thesis. I am really grateful for his kind support and everything.

My sincere thanks also go to all fellows and residents from the Department of Gastroenterology and Colorectal Diversion as well as GI nurses, nurses from Sor Tor 4

building, nurses from all patient wards, OR nurses, all medical staff, and all of 80 volunteers. Their support and kindness fulfilled my willpower and this project to go through.

I am overwhelmed with special thanks to Patcharin Prakobwat, Pratomporn Krangvichian, Suwitra Satheaeanankun, Theerapat Phoka, Asamaporn Klinkajorn, and all my colleagues. Thank you for being an important part of my master's degree.

The Scholarship from the Graduate School, Chulalongkorn University to commemorate the 72nd anniversary of his Majesty King Bhumibol Aduladej is gratefully acknowledged.

Lastly, I would like to be grateful to my mother and father who are behind the scenes in this finished thesis.

Nutta Iadsee

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

ABBREVIATION

ACG	American College of Gastroenterology
AUC	Area under the receiver operating characteristic curve
BFT	*Bacteroides fragilis* enterotoxin
Bla	*Blautia* spp.
CAGs	Co-abundance groups
CEC	Colonic epithelial cells
CIN	Chromosomal instability
CN	Copy number
CRC	Colorectal cancer
Ct	Cycle threshold
DADA2	Divisive Amplicon Denoising Algorithm
DC	Distal colon
DLP	Dyslipidemia
DM	Diabetes mellitus
DNA	Deoxyribonucleic acid
EC	Colibactin positive bacteria
ECM	Extracellular matrix components
EHEC	Enterohemorrhagic *Escherichia coli*
ETBF	Enterotoxigenic *Bacteroides fragilis*
FDR	False discovery rate
FIT	Fecal Immunochemical Test
FN	*Fusobacterium nucleatum*
FOBT	Fecal Occult Blood Test
FS	*Fusicatenibacter saccharivorans*
GF	Germ free
HC	Healthy controls
HT	Hypertension
IBD	Inflammatory bowel disease
Ig	Immunoglobulin
IL	Interleukin

L	Lesion
LB	Luria-Bertani
LDA	Linear discriminant analysis
LEfSe	Linear discriminant analysis (LDA) effect size
LPS	Lipopolysaccharide
NADH	Nonalcoholic steatohepatitis
NGS	Next generation sequencing
NL	Non-lesion
NOD2	Nucleotide-binding oligomerization domain 2
NPV	Negative predictive value
OD	Optical Density
OTUs	Operational Taxonomic Units
PC	Proximal colon
PCoA	Principal coordinate analysis
pks	Polyketide synthase
PL	Peri-lesion
PM	*Parvimonas micra*
PPV	positive predictive value
PSA	Polysaccharide A
qPCR	Quantitative polymerase chain reaction
ROC	Receiver operating characteristic
SCFA	Short chain fatty acid
SG	*Streptococcus gallolyticus*
SOC	Super Optimal broth with Catabolite repression
SPF	Specific pathogen free
STAT3	Signal transducer and activator of transcription 3
TLR2	Toll-like receptor 2
TNM	Tumor, Node, Metastasis
UC	Ulcerative colitis

CHAPTER I

INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer and the third leading cause of death worldwide (1). Based on the National Cancer Institute of Thailand, the incidence of CRC was the third of all Thai cancer patients. CRC was ranked the second in male and the second in female cancer patients (2). Moreover, the incidence of CRC tends to increase continually leading to one of the major health problems in Thailand. CRC is associated with several risk factors including age, genetic factors, lifestyle, and environmental factors, such as high red meat or high-fat diet consumption, high alcohol consumption, smoking, inflammatory bowel disease (IBD), and intestinal polyps (3). Adenomatous polyp which is generally considered to be CRC precursors is necessarily identified for CRC prevention (4). Moreover, people who are above 50 years old and high-risk subjects have been recommended to undergo the CRC screening by colonoscopy (5).

Gut microbiota is a complex and dynamic microbial ecosystem harboring in human intestine. These microbes are known to play a crucial role in maintenance of gut homeostasis. There is growing evidence suggesting that the imbalance of gut microbiota, also called microbial dysbiosis, is related to many diseases including gastrointestinal cancers (6). CRC patients have different composition of gut microbiota from healthy people in term of diversity and richness, i.e., decrease of beneficial bacteria and increase of pathogenic bacteria (7). In previous *in vivo* study, the transfer of commensal pathogens from tumor-bearing mice to recipient mice showed the promoting effects of gut microbiota on tumorigenesis (8). The CRC-related molecular mechanisms of certain intestinal microorganisms in the experimental models have been documented such as DNA damage (9), chronic inflammation (10), and production of carcinogenic metabolites (11). However, in human study, there was no consensus of dysbiotic microbiota composition in CRC patients (12, 13). The inconsistency was caused by different individual parameters and determining methodological used, such as participant criteria, sampling differences, age of

subjects, a number of samples, geographical location, and molecular approach (12, 13). Furthermore, microbiota at the colon mucosa (mucosal microbiota) and in the feces (luminal microbiota) are different because fecal and mucosal microbiota profiling were shown to be partially correlated (14). Hence, studying CRC-associated dysbiosis of both sample types is essential to provide comprehensive information on the colon bacterial community. Several previous studies reported the feasibility of applying CRC-associated microbes in feces as biological markers for CRC screening tools.

So far there has been no study of microbiota changes or microorganism diversity associated with CRC in the Thai population, especially in the elderly who are the high-risk group. Therefore, the major objective of this study is to compare the gut microbiota in both tissue and fecal samples among the healthy elderly, elderly with adenomatous polyp, and elderly with colorectal cancer. The minor objective is to validate the CRC-associated bacteria in both sample types of the same population via quantitative PCR (qPCR) technique and to identify the putative candidate biomarkers for CRC and adenomas detection in Thai population.

CHAPTER II

OBJECTIVE

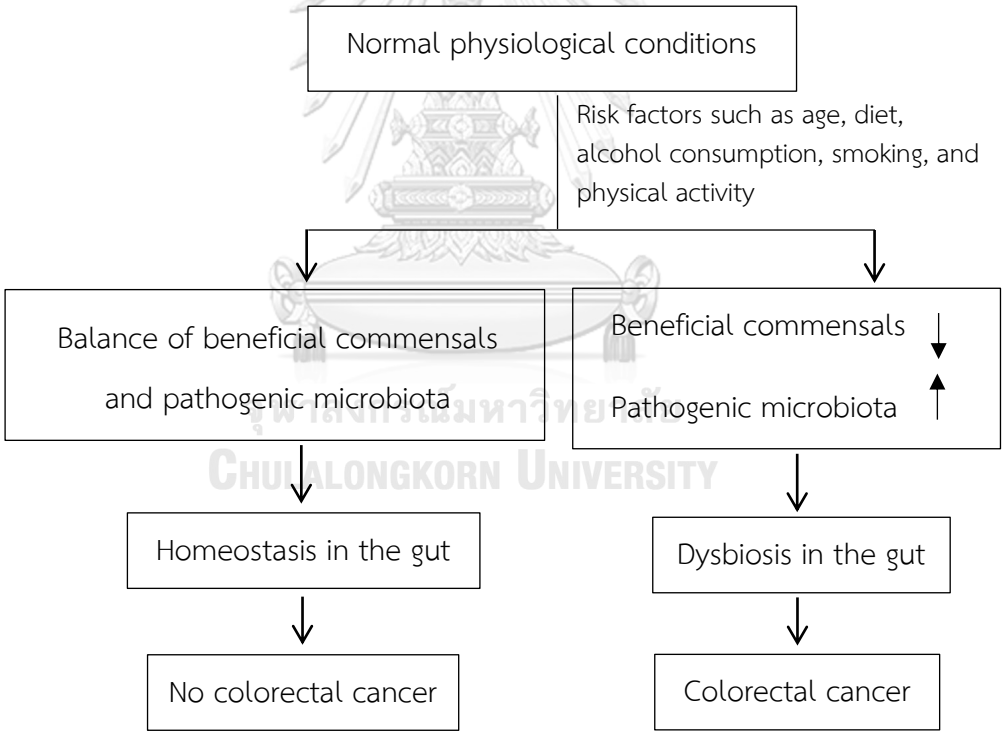
2.1. Hypothesis

The gut microbiota of the Thai healthy elderly is different from the Thai elderly with colorectal cancer.

2.2. Objective

To compare the gut microbiota between the Thai healthy elderly and the Thai elderly with colorectal cancer

2.3. Conceptual Framework



CHAPTER III

REVIEW OF RELATED LITERATURE

3.1. Colorectal Cancer (CRC)

Colorectal cancer (CRC) is the third most common malignancy and the second leading mortality worldwide based on the global cancer burden estimated by the International Agency for Research on Cancer. Approximately 1.9 million new cases and 935,000 deaths of CRC were reported in 2020 (1). In Thailand, the incidence of CRC was ranked the third in male (11.4% of total cancer cases) and the second in female cancer patients (10.7% of total cancer cases) (Figure 1) (2). Moreover, CRC patients tend to increase continually leading to one of the major public health issues in Thailand.

CRC occurs in the colon or rectum. Anatomy of the colon, a part of the digestive system, consists of two major parts: proximal colon and distal colon. The proximal colon contains cecum, ascending, hepatic flexure, and transverse colon. The distal colon includes splenic flexure, descending colon, followed by sigmoid colon and rectum. The function of large intestine is to remove water and nutrients from food materials. The undigested solid materials, referred as stool, move through the bowel, stored in the rectum, and eliminated from the body through the anus (1).

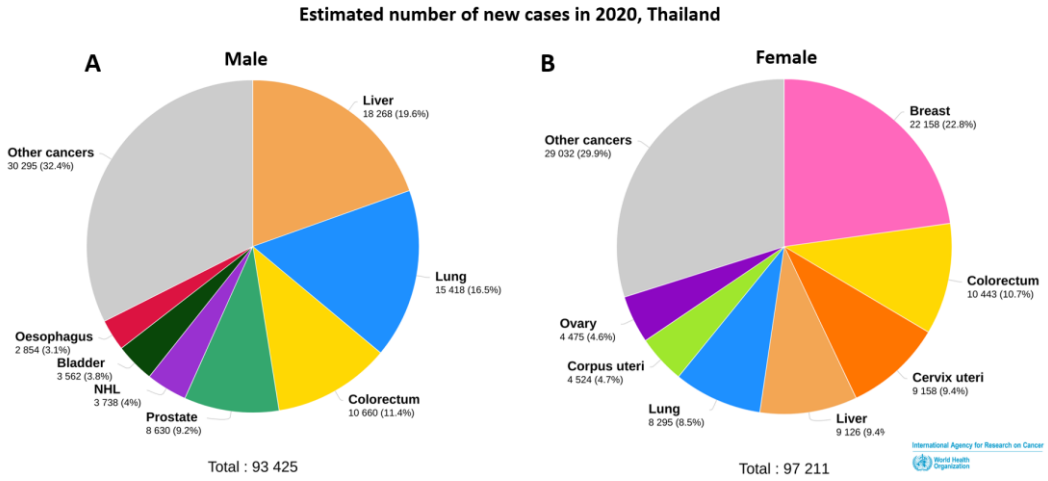


Figure 1 The estimated incidence of all Thai cancer patients in 2020

A) the incidence rate of cancer in male; B) the incidence rate of cancer in female. The data was published by GLOBOCAN 2020 (2)

Colonic adenoma are widely considered to be precancerous lesions in almost all sporadic CRCs, and found in up to 50% of persons above 50 years of age undergoing colonoscopy (15). The tenet of adenoma-carcinoma sequence begins from normal colonic mucosa to abnormal cell proliferation leading to the development of adenomatous polyp, and those with more advanced histopathology features to eventually adenocarcinoma (Figure 2) (7, 15). Colorectal polyps are nonmalignant growth that occurs in the colorectal mucosa. Although not all polyps are adenomas and fewer than 10% of polyps slowly progress to cancer, early CRC screening and removal of polyps by colonoscopy are recommended, especially in the elderly individuals (16).

Approximately 75% of sporadic CRCs, which occur spontaneously without a family history of CRC, are the major proportion of CRC cases. The remaining 25% have a family history of CRC, and probably are a result of the combination of genetic and environmental factors. However, only 5-6% of hereditary CRC cases are due to inherited mutation in major CRC genes (17, 18). This information reflects the considerable influences of lifestyle and environmental factors in CRC carcinogenesis.

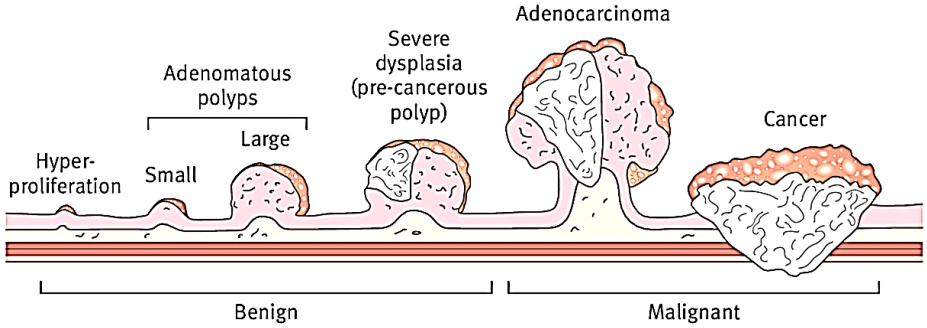


Figure 2 Adenoma-carcinoma sequence (3).

Apart from hereditary factors, several risk factors associated with CRC have been implicated including age; sex; ethnicity; medical history such as Inflammatory

Bowel Disease (IBD) and type 2 diabetes; unhealthy dietary habits such as eating red meat, processed meat, and fat; obesity; low physical activity; tobacco smoking; and alcohol consumption (2, 3) (Figure 3). In addition, accumulating studies reported the role of colonic bacteria in the CRC development (Table1).

3.2. CRC screening tests

Owing to its slow advancement from asymptomatic precancerous lesion i.e. adenomatous polyp to be a malignant tumor, the CRC screening strategies for early detection have been established to reduce the CRC mortality or the burden of the disease (19). The ideal screening test should be efficient with high performance, safe, convenient, accessible, and cost-effective (20). Current CRC screening approaches include invasive tests and non-invasive tests.

Invasive tests consist of flexible sigmoidoscopy and colonoscopy, which are accepted to be the gold standard methods for CRC screening procedure. Two invasive tools allow direct visualization and removal of the premalignant polyp to obtain definite pathological result (21). However, these methods require general anesthesia and prior bowel preparation step and may cause post-procedural complications (22).

Noninvasive tests are divided into blood-based test, stool-based tests, and radiographic examinations. The newly emerged blood test (Epi proColon ®) for cancer screening is used for detection of mutated methylated septin9 DNA from patient whole blood specimens using qualitative polymerase chain reaction (PCR) (19). Methylated SEPT9 has been linked to the incidence of CRC (23). In terms of stool-based tests, the concept of current available tools is to detect occult blood or shredded cell debris from the lesion (24). Those tests compose of the guaiac-based fecal occult blood test (gFOBT), fecal immunochemical test (FIT), and the fecal DNA testing (Cologuard®). Moreover, the radiologic tests including double contrast barium enema, capsule endoscopy, and computed tomographic colonography (CTC) play roles in radiographic visualization and identification of the colonic dysplasia (19).

However, CTC approach has some disadvantages such as high cost, absence of standardized method, requirement of bowel preparation, poor performance of finding flat or tiny polyp, and the inaccessibility to biopsy (25).

3.3. Gut microbiota in health

The term gut microbiome is used for the collective bacterial genomes, outnumbering the human genome by 150-fold (26). Although individuals have different compositions of gut microbiota acting like a fingerprint, three phyla are predominant including *Firmicutes* (30-50%), *Bacteroidetes* (20-40%), and *Actinobacteria* (1-10%) (27). The major proportion of bacterial population is strict anaerobes including *Bacteroides, Eubacterium, Bifidobacterium*, *Fusobacterium*, *Peptostreptococcus,* and the minority is facultative anaerobes constituting *Enterobacter, Escherichia,* and *Lactobacillus* (28). Various and complex microbial communities naturally inhabiting in human large intestine consist of approximately 100 trillion bacterial cells, which are estimated to be tenfold more than human cells, along with fungi, archaea, and viruses. Intestinal microorganisms exist and co-evolve as a mutualistic relationship in humans. Intestinal bacteria residing in a colonic niche play a crucial role in maintaining gut homeostasis by uptake of indigestible carbohydrates, production of vitamin B and K, beneficial fermentation end products such as short chain fatty acids (SCFAs), maturation of intestinal immune system, and maintaining the mucosal barrier function (29). Either extrinsic factors such as dietary intake, feeding habits, antibiotic treatment, and the maternal microbiota, or intrinsic factors can affect the intestinal ecosystem in term of species richness and evenness. Moreover, increasing reports showed that the disruptive changes in the intestinal microbiota are associated with many diseases including CRC (26, 29, 30).

Table 1 The evidence of a relationship between intestinal bacteria and colorectal carcinogenesis

Observation	Ref.
Germ-free (GF) or genetic deficient animal models reduced the colonic tumor burden compared with conventional or Specific Pathogen-Free (SPF) conditions.	(31-34)
Conventionalized GF mice with intestinal microorganisms from tumor-bearing mice significantly enriched the colonic tumors compared with those received from control mice.	(8)
Antibiotics treated mice prior to chemical induced treatment had lower number of tumors compared with those of untreated mice.	(8, 35)

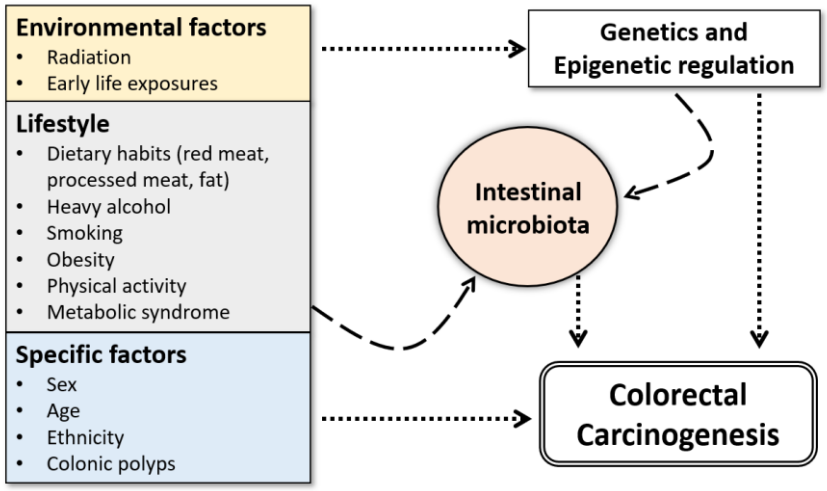


Figure 3 The risk factors related to the development of CRC (modified from a previous study by Nistal et al. 2015 (6))

3.4. Gut microbiota in CRC

The host physiology, intestinal microbiota, and dietary factor are crucial for the balanced microbial ecosystem. A shift in any of these factors might turn a state of homeostasis into microbial dysbiosis, which is associated with CRC tumorigenesis

(36). Numerous studies reported the alteration of microbial composition in CRC patients and healthy control (Table 2). Several studies showed lower bacterial diversity in fecal samples and biopsy tissues of CRC patients compared with normal individuals (37, 38). In addition, microbial diversity and richness was higher in adenoma mucosa than that in normal mucosa (39, 40). There is a dynamic change of intestinal microbiota across different stages of tumor. However, most studies were cross-sectional, i.e., specimens were collected at a single time point, therefore the microbial changes only indicate disease association but cannot firmly be established as the etiology of CRC carcinogenesis.

3.4.1. Intestinal microorganisms in fecal samples of CRC patients

Four major phyla of fecal bacteria consisting of Firmicutes, Bacteroidetes, Actinobacteria, and Proteobacteria were recognized in both CRC patients and non-CRC volunteers. Up to 90% of bacterial population were dominated by Firmicutes and Bacteroidetes (41, 42).

Under the phylum Firmicutes, a reduction of *Roseburia*, *Faecalibacterium*, and *Eubacterium*, and outnumber of *Streptococcus* and *Enterococcus* were shown in fecal specimens of cancer patients compared with the noncancer group (14, 43). At species level, the abundance of *Lachnospira pectinoschiza*, *Lachnospira bovis*, *Pseudobutyrivibrio ruminus*, *Ruminococcus obeum* and *Ruminococcus albus* was lower in the stool samples of cancer patients (42). Due to the decline of bacterial members belonging to Firmicutes, which has the capability of SCFA production, fecal butyrate in CRC stool samples was lower than that in healthy individuals (42). Furthermore, the genus *Lactobacillus* did not alter in fecal samples of the cancer group (43-45).

For the phylum Bacteroidetes, the genus *Prevotella* and *Porphyromonas* in fecal samples of the cancer group were higher than those in healthy donors (42, 43). However, the inconsistent results were found in *Bacteroides*. Some studies reported higher level of *Bacteroides* in the cancer group (41, 44), whereas another reported

lower level (43). Moreover, at the species level of *Bacteroides*, *B. fragilis* was significantly higher in fecal samples of the CRC group, but *B. uniformis* and *B. vulgatus* were lower in the cancer group compared with controls (43, 45).

Among other minor phyla, *Bifidobacterium*, one of *Actinobacteria*, revealed inconsistent results. Some studies showed increased level of the bacteria in fecal samples of the CRC group (43), whereas other studies showed no alteration between groups (44, 45). The increased levels of genus *Fusobacterium* belonging to *Fusobacteria* and *Escherichia* belonging to Proteobacteria were shown in CRC stool samples.

3.4.2. Intestinal microorganisms in tissue samples of CRC patients

To study the mucosal-associated microbiome, a comparison between lesion tissues from CRC patients and adjacent normal tissues from the same cancer patients or normal tissues from healthy volunteers was performed. A previous study showed that the mucosal microbiome was predominated with approximately 60% of phylum *Proteobacterium* in non-cancer individuals (46). Another study reported the major percentage of phyla *Firmicutes, Bacteroidetes* and *Proteobacteria* in the healthy group (47). In contrast to mucosal-associated microbiota, these results indicated that fecal microbiota incompletely represents the ecosystem of intestinal bacteria on the mucosa layer.

Although the genera *Fusobacterium* and *Escherichia* were the minor phyla of fecal microbiota (41, 42), mucosal microbiota in both genera were significantly higher in lesional tissues than non-lesional tissues (46-48).

Mucosal microbiota belonging to Firmicutes and Actinobacteria was low in CRC groups compared with healthy donors (46, 48). The reduction of the family *Ruminococcaceae* and *Lachnospiraceae* was found in the CRC groups (43, 47). No difference in the genus *Lactobacillus* was observed between CRC and control groups (44).

Under phylum Bacteroidetes, some studies reported that the CRC group had increased level of genus *Bacteroides* (14, 46), but another showed low level in CRC patients compared with healthy volunteers (47).

Taken together, CRC-associated microbial dysbiosis in stool and tissue samples reported dissimilar results (Table 2). Moreover, the gut microbiota alteration in CRC patients has not yet been elucidated and remains to be explored.

Table 2 Bacterial enrichment in CRC patients in different geographical location

Study	Location	Sample type	Bacterial enrichment in CRC patients
Zeller et al., 2014 (49)	France/ Denmark / Spain/ Germany	Tissue	*Fusobacterium nucleatum, Bacteroides. fragilis*
		Feces	*F. nucleatum, Peptostreptococcus stomatis, Porphyromonas asaccharolytica, B. fragilis Clostridium symbiosum*
Zackular et al., 2014 (50)	USA/ Canada	Feces	*Fusobacterium, Porphyromonas, Lachnospiraceae, Enterobacteriaceae*
Wu et al., 2013 (41)	China	Feces	*Bacteroides, Campylobacter, Fusobacterium*
Weir et al., 2013 (42)	USA	Feces	*Acidaminobacter, Phascolarctobacterium, Citrobacter farmer, Akkermansia muciniphila*
Ahn et al., 2013 (37)	USA	Feces	*Fusobacterium*, *Porphyromonas, Atobium*
Wang et al., 2012 (43)	China	Feces	*Peptostreptococcus, Enterococcus, Streptococcus, Escherichia/Shigella, Klebsiella*

Study	Location	Sample type	Bacterial enrichment in CRC patients
Flemer et al., 2017 (14)	Ireland	Tissue	*Bacteroides, Roseburia, Ruminococcus, Oscillibacter, Porphyromonas, Peptostreptococcus, Parvimonas, Fusobacterium*
		Feces	*Parvimonas, Anaerococcus, Streptococcus, Fusobacterium*
Baxter et al, 2016 (51)	USA/ Canada	Feces	*F. nucleatum, P. asaccharolytica, P. stomatis, P. micra*
Mira-Pascual et al, 2016 (52)	Spain	Tissue	*Escherichia-Shigella, Streptococcus*
Nakatsu et al., 2015 (53)	HongKong	Tissue	*B. fragilis, Gemella, Parvimonas, Peptostreptococcus, Granulicatella*
Kasai et al., 2015 (45)	Japan	Feces	*Actinomyces odontolyticus, Clostridium nexile, Veillonella dispar, Haemophilus parainfluenzae, F. varium, P. stercorea, S. gordonii B. fragilis*
Sobhani et al., 2011 (44)	France	Feces	*Bacteroides/Prevotella*
Gao Z. et al., 2015 (46)	China	Tissue	*Lactococcus, Fusobacterium*
Burns et al., 2015 (47)	USA	Tissue	*Fusobacterium, Providencia*

Study	Location	Sample type	Bacterial enrichment in CRC patients
Gao R. et al., 2017 (48)	China	Tissue	*Fusobacterium, Prevotella, Alloprevotella, Porphyromonas, Peptostreptococcus, Parvimonas*

3.5. Potential role of gut microbiota in CRC tumorigenesis

Intestinal microbiota has been studied in the experimental model to find out the possible underlying mechanisms related to CRC initiation and progression. Several of those empirical mechanisms include (i) attachment, invasion, and translocation. (ii) the induction of chronic inflammation, and (iii) production of carcinogenic metabolites such as extracellular superoxide, and genotoxins. The present reports on the relationship of these CRC-associated bacteria are described as follows.

Fusobacterium nucleatum

F. nucleatum is a gram-negative, rod shaped, and strict anaerobe. In healthy state, it colonizes in oral cavity and intestinal tract. Besides CRC, it also involves in periodontal disease and pregnancy complications (54). Various studies reported *F. nucleatum* was overabundant in both tissue and fecal samples of CRC patients (14, 37, 46-51). The experimental study by Rubinstein et al. (2016) revealed that its key virulence factor is the cell surface protein, FadA. The molecule is known to adhere to extracellular domain of E-cadherin and invade the host mucosa. This event promotes pro-oncogenic and inflammatory pathways via Wnt/β-catenin signaling (55). Furthermore, *F. nucleatum* can inhibit host mucosal immunity that favor the tumor growth within the colon mucosa (56-58). However, CRC progression as a consequence of the bacterium is still controversial because of the conflicting evidence from experimental studies. The pathogen does not always stimulate cancer

formation *in vivo* (59), but highly induces cell proliferation in all cancer cell lines tested *in vitro* (55).

Enterotoxigenic Bacteroides fragilis

B. fragilis is an obligate anaerobe, gram-negative, rod-shaped bacterium. This bacterium generally resides in human gastrointestinal tract as a commensal organism and predominates in gut microbiota. A nontoxigenic *B. fragilis* conferred tumor inhibitory effects by production of polysaccharide A (PSA) that mediated Toll-like receptor 2 (TLR2) signaling in a mouse model of colitis-associated CRC (60). However, enterotoxigenic *Bacteroides fragilis* strain (ETBF) was shown to be related to CRC (61). The ETBF produces a metalloprotease toxin, named Fragilysin, or *B. fragilis* enterotoxin (BFT) to cleave E-cadherin, the extracellular matrix of the adherens junctions, contributing to morphologic alteration in colonic epithelial cells (CEC) (11). This cleavage results in the higher permeability of gut barrier and the activation of Wnt/β-catenin signaling pathway leading to hyperproliferation of CEC (62). Moreover, the toxin can trigger inflammatory cascades i.e., Interleukin17 (IL-17), NF-**K**B signaling, and signal transducer and activator of transcription3 (STAT3), leading to mucosal inflammation and tumor initiation and progression of CEC (63). Nevertheless, the pathogenicity of *B. fragilis* relies on the certain expression of virulence factors of pathogenicity islands (64). These observations support the association between ETBF and colorectal carcinogenesis.

Colibactin producing *Escherichia coli*

The association of *E .coli* strains and CRC are found in strains that harbor a polyketide synthase (*pks*) island (65). The *pks* positive *E. coli* are accounted for approximately 34% of all *E. coli* isolates. The bacteria can produce a genotoxin named Colibactin, which is capable to induce DNA double-strand breaks *in vivo* and chromosomal instability (CIN) in mammalian cells (9). These effects may promote the initiation and progression of CRC. In addition, the study in human showed the high

prevalence of colibactin-producing *E. coli* in colon cancer patients (66). However, *E. coli* is normally regarded as normal flora in the intestine tract, and further studies are therefore needed to investigate the differences between *E. coli* isolates collected from CRC patients and healthy controls (36).

Streptococcus gallolyticus

Streptococcus gallolyticus subsps. *gallolyticus* (*S. gallolyticus*), formerly known as *S. bovis* biotype I, is an opportunistic pathogen in humans and mostly found in case of bacteremia and endocarditis. Strikingly, *S. gallolyticus* infection is involved in CRC (36). The bacterium was able to adhere to extracellular matrix components (ECM), i.e., laminin, collagen, and fibronectin, as well as human colon cancer epithelial cells (Caco-2) (67). After colonization, the bacterium had the ability to translocate paracellularly through differentiated Caco-2 cells and produced biofilms on the cell surface to facilitate the innate immune evasion (68). Furthermore, Kumar and colleagues (69) investigated the increase of tumor cell proliferation when co-culture of clinical *S. gallolyticus* strains with human colon cancer cells and bacterium-gavage mouse models. The elevation of NF-**K**B and IL-8 expression in tissue, a proinflammatory state, was detected in CRC and adenoma patients who had *S. gallolyticus* immunoglobulin G sero-positive (70). However, the cancer cases induced by *S. gallolyticus* exposure were the minority of all cancer cases (71). Therefore, the contribution of *S. gallolyticus* to promote tumor growth might be limited to a subset of individuals (72).

3.6. Gut microbiome analysis

These definitions and related words are used for gut microbiome analysis (Table 3).

Table 3 Definitions related to the intestinal microbiome study

Terminology	Denotation	Reference
Microbiota	The microbial taxa in a variety of environment	(73, 74)
Microbiome	The collective genome of microbes in a particular ecosystem	
16S ribosomal genes	A housekeeping gene region with hypervariable sequenced which universally conserved in prokaryotes. This gene region can be used to identify the microbial communities within samples.	(12, 75)
Operational Taxonomic Units (OTUs)	An analysis of 16S rRNA sequencing data based on sequence similarity (typically >97% similarity). An OTU is used to classify statistical clusters of highly related bacteria.	(73, 75)
Diversity	A measure of the variability of species which rely on the diversity indices	
Alpha diversity	Diversity within each sample	(75)
Beta diversity	Diversity between different samples	
Species richness	A measure of the total number of different species occurs in a defined area	(76)
Species evenness	A measure of the relative abundance of species within a community	

3.7. Approaches used for gut microbiome study

Gut microbiome studies were conducted on two main types of samples, feces, and biopsy tissue, to provide the information of microbial ecosystem structure. Feces is the most commonly used sample that represents luminal associated

microbiota. Fecal collection is a non-invasive procedure, easy to handle, and more practical to identify microbial biomarkers. However, stool samples only partially reflect the gut microbiota throughout the colon. Moreover, the complexity of stool specimens may include many unrelated components that could interfere a disease screening. A utilization of tissue samples is likely to give more relevant data to illustrate the involvement of colonizing microbial community on colon mucosa (12-14, 77). However, a collection of biopsy tissue or surgical tissue, which gives the result of mucosa-associated microbiota, is more difficult to carry out especially in healthy volunteers, since it is an invasive technique that requires colonoscopy or surgery, and also expensive.

Besides sample type, other biological parameters; for example, lifestyle, dietary habits, geography of the studied cohort, colon location of tissue sampling (e.g., proximal, or distal colon), age of volunteers, stage of the tumor (e.g., TNM classification), influence the variety of the intestinal microbiota community contributing to the lack of consensus in term of microbial dysbiosis in CRC patients (12, 13, 54). Moreover, the technical parameters including sample size, molecular procedure performed (e.g., whole-genome sequencing, 16S rRNA gene sequencing, qPCR), selection of a hypervariable region of 16S rRNA gene (e.g., V1-V2, V3-V4, V4, V3-V5, V6) for Next Generation Sequencing (NGS), level of taxonomy determined, and selection of dissimilar databases may also affect the results of gut microbiome.

3.8. Translational application of gut microbiome study

- Gut microbiota-related biomarkers for CRC screening test

Understanding of CRC-associated gut microbiota based on NGS innovation is beneficial for selection of good biomarker candidates for CRC screening. Although the exact microbial consensus related to CRC and adenomatous polyp has not been precisely determined, various studies have shown that fecal microbial alteration may give new potential biological markers for CRC detection and prognosis, especially for early stages of CRC (13). Moreover, high specificity and sensitivity of noninvasive

screening tools are needed to reduce the incidence and mortality of CRC. A number of studies have applied a relative abundance of each bacterium or bacterial co-abundance groups (CAGs) to distinguish cancer patients from non-cancer volunteers (49-51). Two studies were conducted on fecal samples from 3 groups, CRC patients, adenomas, and healthy control subjects, to create a classification model using 22 and 34 microbial markers with areas under the receiver operating characteristic (ROC) curves (AUC) of 0.84 and 0.85 (49, 51), respectively, indicating their potential as a promising screening tool. Moreover, the combination of bacterial markers with Fecal Immunochemical Test (FIT) or Fecal Occult Blood Test (FOBT) could provide superior performance in detecting CRC and advanced adenoma than using the latter test alone (49, 78). However, the development of reproducible procedures for human intestinal microbiota study is desired for more comparable results among populations (79).

CHAPTER IV

Materials and Methods

4.1. Ethics statement

This study was carried out with approval from the Institutional Review Board (IRB) of the Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand (approval number 182/62). Written informed consent was obtained from each participant for the sample collection and clinical data collection.

4.2. Volunteer recruitment

This study enrolled eighty volunteers who visited King Chulalongkorn Memorial Hospital, Bangkok, Thailand from June 2019 to December 2020. The participants were selected based on the inclusion and exclusion criteria listed below.

4.2.1. Inclusion criteria

Eligible participants were individuals above 50 years old who were able to provide informed consent.

4.2.2. Exclusion criteria

Exclusion criteria were as follows: (i) antibiotics consumption within 3 months before sampling; (ii) probiotics usage in any form within 1 week before enrollment; (iii) history of inflammatory bowel disease, including ulcerative colitis and Crohn's disease; (iv) active bowel inflammation or infection within one month before participation; (v) current immunosuppressive drug usage; (vi) previous chemotherapy or radiotherapy. (vii) colonoscopy within a month prior to participation.

All volunteers were divided into three groups: colorectal cancer (CRC), colorectal polyp, and healthy control. Twenty-five newly diagnosed CRC patients were recruited at the Colorectal Surgery Unit. Thirty-three patients with colorectal polyp and twenty-two healthy controls who had no colorectal polyp or CRC were recruited from subjects undergoing screening colonoscopy at the Division of

Gastroenterology. The group classification of all participants was confirmed by the pathological results. Besides, general information including health status, lifestyle, dietary habits, and anthropometric measurements (height and body weight) were obtained from participants and recorded.

4.3. Sample collection

The workflow and sample collections of this study were conducted following the schematic diagram below (Figure 4).

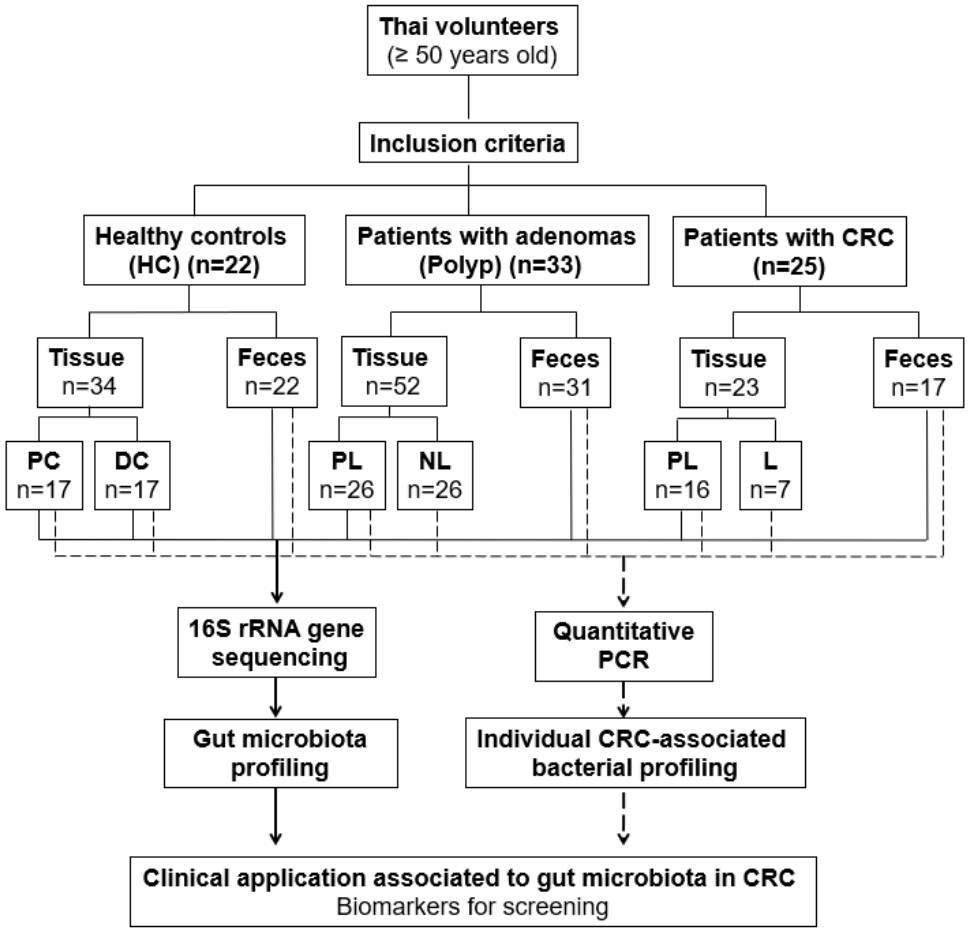


Figure 4 Schematic diagram of experimental design and total specimen collection in this study.

Abbreviation: PC, proximal colon; DC, distal colon; PL, peri-lesion; NL, non-lesion; L, lesion.

4.3.1. Stool collection

The fecal collection kits along with detailed printed instructions were given to the participants before undergo colonoscopy or surgery. To minimize the change of gut microbiota, the volunteers were requested to collect their feces on the day before consuming polyethylene glycol-electrolyte solution for bowel preparation, or before receiving chemotherapy or radiotherapy. Stool samples were collected as described in the following instruction. After urination, the participants were asked to defecate on a provided absorbent pad. The stool was collected using a spatula and collected up to the equal amount of buffer in the tubes containing 3 ml of DNA preservative buffer (Monarch DNA/RNA Protection Reagent, NEB, England), and 3 ml of 50% sterile glycerol for fecal immunochemical test (FIT) and bacterial culture. After mixing, the sample tubes were immediately placed into the kit bag containing an iced pack and kept in the freezer. The fecal specimens were carried to the laboratory on the doctor appointment day. The stool samples were kept at -80 °C for long-term storage until DNA extraction.

4.3.2 Tissue collection

Approximately one cubic centimeter of mucosal tissues was collected from CRC patients during surgery. One tissue sample from the tumor (lesion, L) and one normal tissue sample adjacent to the lesion (peri-lesion, PL) were obtained from each patient. One cubic millimeter of biopsy tissues were collected from adenoma subjects and healthy control during colonoscopy. For adenoma subjects, one tissue sample closed to the polyp (peri-lesion, PL) was collected because the main tissue was reserved for pathological examination to avoid misdiagnosis of malignancy. In addition, one normal tissue sample at the opposite side of colon (non-lesion, NL) was biopsied. For healthy controls, one biopsy tissue was collected at each location in the proximal (PC) and distal (DC) colon. All samples were placed in DNA preservative buffer (Monarch DNA/RNA Protection Reagent, NEB, England) and

immediately kept at 4 ℃. The samples were transferred to the laboratory and stored at -80 ℃ for long-term storage until use.

4.4. DNA extraction from stool samples

Fecal genomic DNA was extracted using the QIAamp PowerFecal Pro DNA Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The stool sample was centrifuged at 12,000 xg for 10 min to discard the DNA preservative buffer, and approximately 250 mg of feces was added to the PowerBead Pro Tube containing 800 µl of Solution CD1. Then, the sample was homogenized using the TissueLyser LT (Qiagen) at a maximum speed for 10 min. The fecal mixture was centrifuged at 13,000 xg for 3 min, and the supernatant was transferred to a clean 2 ml tube. Two hundred-microliter Solution CD2 was added into the tube and mixed. After centrifugation, up to 700 µl of the supernatant was transferred to a clean tube followed by mixing with 600 µl of Solution CD3. The lysate was loaded into an MB Spin Column. The column was centrifuged and the flow-through was discarded. Then, the column was placed into a clean collection tube followed by two washing steps with 500 µl of Solution EA and 500 µl of Solution C5, respectively. After drying the column, 50-100 µl of Solution C6 was loaded into the column and DNA was eluted into a new 1.5 ml elution tube. DNA quality and quantity was determined by measuring absorbance at the wavelength of 260 and 280 nm using a NanoDrop2000 spectrophotometer (Thermo Fisher Scientific, MA, USA), and checked the integrity of DNA by 1.5% (w/v) agarose gel electrophoresis. All DNA samples were stored at -20 ℃ until use.

4.5. DNA extraction from tissue samples

DNA was extracted from tissue samples using the QIAamp Fast DNA Tissue Kit following the manufacturer's instruction (Qiagen). One biopsy sample or 10 mg of surgical tissue was placed in the Tissue Disruption Tube containing master mix (200 µl of AVE, 40 µl of VXL, 1 µl of DX Reagent, 20 µl of 600 mAU/ml proteinase K, and 4 µl

of 100 mg/ml RNase A). The tissue samples were homogenized using TissueLyser LT (Qiagen) at 45 Hz for 2 min and incubated at 56 ℃ for 10 min before mixing with 265 µl of Buffer MVL. Then, the mixture was transferred to the QIAamp Mini Spin Column. After centrifugation at 13,000 xg for 3 min, the column was placed into a new collection tube. Then, the column was washed using 500 µl of Buffer AW1 followed by 500 µl of Buffer AW2. The column was dried in a new 1.5 ml microcentrifuge tube by centrifugation. DNA was eluted from the column by 50-100 µl of ATE buffer. The amount of DNA was quantitated by measuring absorbance at the wavelength of 260 and 280 nm using the NanoDrop2000 spectrophotometer (Thermo Fisher Scientific). Integrity and size of DNA was measured by 1.5% (w/v) agarose gel electrophoresis. The DNA was kept at -20 ℃ until processing.

4.6. 16S rRNA gene sequencing and bioinformatics analysis

Paired-end sequencing was conducted using the Illumina MiSeq 250 bp platform (Illumina, San Diego, CA, USA) at Génome Québec Innovation Centre (Montréal, QC, Canada). The V1-V2 hypervariable regions of the 16S rRNA gene were targeted using the forward primers: 27bF (5'- AGRGTTTGATCMTGGCTCAG-3') and the reverse primers: 338R (5'-TGCTGCCTCCCGTAGGAGT-3'). The raw 16S rRNA amplicon sequences were preprocessed in the bioinformatics pipelines by the team of the Canadian Centre for Computational Genomics (C3G) (McGill University, Canada). Briefly, these data were mainly preprocessed following divisive amplicon denoising algorithm 2 (DADA2) pipelines (80) including quality control of sequencing reads and clustering of reads into amplicon sequence variants (ASVs). Then taxonomy was annotated with the silva reference database (81).

4.7. Microbiome data analysis

Comparative analysis of microbiota abundance was performed using the MicrobiomeAnalyst web-based platform (https://www.microbiomeanalyst.ca/) (82). Gene abundance data were analyzed by Marker Data Profiling (MDP). Data were

filtered removing features with less count of 4 and less than 20% of prevalence, as a minimum, and a low variance filter of 20%, based on inter-quartile range. Alpha-diversity profiling, describing the within-community diversity of bacteria within a sample, was calculated based on total numbers of ASV analyzed using the non-parametric tests. Moreover, the parameter of beta-diversity analysis, describing the microbiota diversity among samples, was calculated using Bray Curtis distance and permutational multivariate analysis of variance (PERMANOVA). Heat tree analysis was generated for pairwise comparisons of microbial communities. Reingold-Tilfold graph layout was performed and Log2 fold change of relative abundance was displayed. Additionally, the different taxonomy abundance among groups was identified with classical univariate statistical comparison. Values were considered statistically significant when *p*-value <0.05. While robust biomarkers of CRC were also identified using the linear discriminant analysis (LDA) effect size (LEfSe) approach (83) with *p*-value and adjusted *p*-value cut-off of 0.05.

4.8. Detection of CRC-associated bacteria in stool and tissue samples by real-time qPCR

4.8.1. Control strains and control bacterial DNAs

All bacterial strains listed in Table 4 were obtained from the Bacteriology Unit, the Department of Microbiology, Faculty of Medicine, Chulalongkorn University, and kindly provided by Assoc. Prof. Somying Tumwasorn at the Department of Microbiology, Faculty of Medicine, Chulalongkorn University. To be used as control DNA, genomic DNA of all bacteria was extracted by GeneJet Genomic DNA Purification Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. All bacterial DNAs were stored at -20 °C until use. Control DNAs of *Parvimonas micra* (PM), *Blautia* spp. (Bla) and *Fusicatenibacter saccharivorans* (FS), and genomic DNA of selected fecal samples in this study were directly amplified using specific primers in Table 4.

Table 4 Bacterial strains used in this study

Bacterial names	Strains	Sources
Fusobacterium nucleatum (FN)	Laboratory strain	Bacteriology Unit, Department of Microbiology, Faculty of Medicine
Escherichia coli (EC)	ATCC29212	
Streptococcus gallolyticus (SG)	ATCC9809	
DNA of *Parvimonas micra* (PM)		Selected fecal sample in this study
DNA of *Blautia* spp. (Bla)		
DNA of *Fusicatenibacter saccharivorans* (FS)		

4.8.2. Primers

Primer pairs targeting specific genes of each bacterial genus or species were listed in Table 5. The primers were selected from previously published literatures, and their specificity was confirmed using Primer-BLAST (84).

Table 5 Oligonucleotide primers specific to each bacterium in this study

Taxa	Std. strain	Primer sequences (5' to 3')	Gene	Anneal-ing temp	Size (bp)	Ref.
FN	FN	**F-**CAACCATTACTTTAACTCT ACCATGTTCA	*nusG*	55 °C	112	(85)
		R-GTTGACTTTACAGAAGGAG ATTATGTAAAAATC				
EC	EC	**F**-GCGCATCCTCAAGAGTAAATA	*clbB (pks)*	55 °C	283	(86)
		R-GCGCTCTATGCTCATCAACC				

Taxa	Std. strain	Primer sequences (5' to 3')	Gene	Anneal-ing temp	Size (bp)	Ref.
SG	SG	**F**-CAATGACAATTCACCATGA	*sodA*	55 °C	408	(87)
		R-TTGGTGCTTTTCCTTGTG				
PM		**F**- GTCACTACGGAAGAATTTGTC	*rpoB*	55 °C	200	(78)
		R- GGCTTGAGCGATAATAACTTC				
FS	Fecal sample	**F**-CTGCATTGGAAACTGTCTGG	16S rRNA	55 °C	389	(88)
		R-CGTTACGGGCCGGTCATC				
Bla		**F**-GTGAAGGAAGAAGTATCTCGG	16S rRNA	55 °C	559	
		R- TTGGTAAGGTTCTTCGCGTT				

Abbreviation: std, standard.

4.8.3. Insertion fragment preparation and ligation

The gene specific to each bacterium was amplified by conventional Polymerase Chain Reaction (PCR) using the primers listed in Table 2. The reaction mixture for each PCR reaction consisted of 2.5 µl of 10X Taq Buffer with KCl, 1.5 µl of 25 mM $MgCl_2$, 0.5 µl of 10µM dNTP, 0.25 µl of Taq DNA polymerase (Thermo Fisher Scientific), and 16.25 µl of nuclease-free water. The amplification condition using Thermocycler (ProFlex™, Thermo Fisher Scientific) was as follows: 95 °C for 5 min, followed by 30 cycles of 95 °C for 30 sec, 55 °C for 30 sec and 72 °C for 30 sec and a final extension at 72 °C for 7 min. The PCR products were checked by 1% (w/v) agarose gel electrophoresis. The specific bands were cut and purified from the gel by GeneJET Gel Extraction Kit (Thermo Fisher Scientific) following the manufacturer's instructions. Based on TA cloning, the 3´A-tailed PCR products were individually used as the insertion fragment to ligate into the pGEM®-T Easy Vector Systems (Promega, WI, USA) according to the manufacturer's protocol. The ligation mixes included 10X buffer of T4 DNA ligase with 10 mM ATP (NEB, England), T4 DNA ligase (NEB, England),

pGEM®-T Easy Vector, PCR product, and nuclease-free water. The mixture was subsequently incubated at 4 °C overnight before transformation step.

4.8.4. Competent cell preparation

A single colony of *Escherichia coli* DH5α (Novagen, Darmstadt, Germany) was inoculated into 5 ml of Luria-Bertani (LB) broth and incubated at 37 °C, 200 rpm for 16-18 hours. After 100-fold dilution of the bacterial culture, 100 µl of diluted culture was added into 10 ml LB broth followed by incubation at 37 °C and shaking at 200 rpm for 1.5-2 hours to reach the log phase (OD600 ≈ 0.2-0.4). The bacterial broth was transferred on ice for 10 min and then centrifuged at 1600 xg at 4 °C for 10 min. After discarding the supernatant, the pellet was resuspended in one-half of their original volume with sterile cold TB solution (10 mM PIPES, 55 mM $MnCl_2$, 15 mM $CaCl_2$, and 250 mM KCl) (89) and incubated on ice for 25 min. Afterwards, the cells were centrifuged at 1600 xg at 4 °C for 10 min and resuspended with sterile cold TB solution by one-tenth ratio of their original volume. Each 100 µl of mixture was transferred into a sterile cold 1.5 ml microcentrifuge tube. The bacterial cells were on ice until use and kept in sterile cold TB solution containing 15% (v/v) glycerol for long-term preservation.

4.8.5. Bacterial transformation by heat shock procedure

After ligation, 100 µl of the DH5α cells were mixed gently with 5 µl of the ligation mix and incubated on ice for 30 min. Then, the mixture tube was incubated in a circulating water bath at 42 °C for 45 sec. The tube was transferred rapidly on ice for 2 min and 900 µl of Super Optimal broth with Catabolite repression (SOC) medium was added. Subsequently, the tube was incubated at 37 °C and shaking at 200 rpm for 2 hours. The bacterial culture was plated on LB agar containing 100 µg/ml ampicillin using the spread plate technique, followed by incubation at 37 °C for 24 hours. The colonies of transformants were picked up to check the inserted plasmid via colony PCR using specific primers. The nucleotide sequences of the

constructed plasmid were confirmed by DNA sequencing using M13 primers (M13F: **5**'-GTAAAACGACGGCCAGT-3' and M13R: 5'- GCGGATAACAATTTCACACAGG-3') (Macrogen Inc., Seoul, Republic of Korea).

4.8.6. DNA standard curve

The colonies containing the inserted plasmid were inoculated into LB broth and incubated at 37 °C with shaking overnight, and then the bacterial culture was extracted plasmid using HiYield™ Plasmid Mini Kit (RBCBioscience, Taiwan) according to the manufacturer's protocols. The quality and quantity of plasmid DNA was measured by the absorbance at the wavelength of 260 and 280 nm using NanoDrop2000 spectrophotometer (Thermo Fisher Scientific). The plasmid copy number of each gene was calculated based on the length of the PCR product and the DNA concentration using the following formula:

Gene copy number = (amount*6.022x1023)/(length*1x109*650)

Each plasmid was 10-fold serially diluted from 2×10^8 to 2×10^0 copies/µl to be used as a standard curve to quantitate the copy number of individual strains per gram of each sample.

4.8.7. Quantitative real-time PCR (qPCR)

To quantify bacterial load in feces and tissue samples, the absolute quantitative real-time PCR was performed using the same primers used for the conventional PCR (Table 2). The standard curves were constructed using serially diluted plasmid DNA containing the gene specific to the relevant positive control bacterium. The experiments were performed in duplicate using QuantStudio 6 Flex Real-Time PCR systems (Applied Biosystem, Thermo Fisher Scientific) and the Luna Universal qPCR Master Mix (NEB, England). The reaction was performed in a total volume of 20 µl; the component of the master mix was as follows: 10 µl of Luna Universal qPCR Master Mix, 1 µl of 10 µM primers, 4 µl of nuclease-free water, and 5 µl of DNA template. The qPCR condition consisted of an initial denaturation at 95 °C

for 5 min; 40 cycles of denaturation at 94 °C for 60 sec, annealing at 60 °C for 30 sec, and extension at 72 °C for 30 sec; a final extension cycle at 72 °C for 8 min. The samples, standard curve, and negative control were all simultaneously assayed in duplicate. After amplification, the specificity of PCR product was conducted by the melting curve analysis. The cycle threshold (Ct) of each sample was compared with Ct of the standard curve to calculate the bacterial quantity. The data were normalized to total weight of extracted samples and represented as a copy number of bacteria per gram weight.

4.9. Fecal Immunochemical Test

To detect human fecal hemoglobin in the fecal samples, two approaches of the fecal immunochemical test (FIT) were conducted composing of the qualitative FIT, using OC-Light™ S FIT test strip (Eiken Chemical, Japan), and the quantitative FIT, using OC-Auto Sampling Bottle 3 (Eiken Chemical) with automated OC-SENSOR io series (Eiken Chemical). Both methods were performed according to the manufacturer's instructions. For both FITs, the sample probe from the sampling bottle was dipped in the fecal sample tube containing 50% glycerol. The test strips and sampling bottle containing stool samples were incubated at 20-30 °C. The sampling bottle was shaken vigorously. In the case of the qualitative FIT, the OC-Light S FIT test strip was removed from the canister. Then, the sample end of the test strip was dropped into the sampling bottle. After incubation for 5 min, the result was read from the strip following the interpretation manual. In part of the quantitative FIT, the sampling bottles were applied together with the automated analyzer. After that, the instrument printed out the quantity of fecal human hemoglobin. The test was analyzed one at a time and was reported positive at a cut-off value of 50 ng of hemoglobin per milliliter (ng/ml).

4.10. Statistical analysis

Statistical analysis was conducted using IBM SPSS Statistics version 22 (SPSS Inc.) and GraphPad Prism 8.0 (GraphPad Software Inc.). Chi-square test was used to compare categorical variables. The nonparametric Kruskal-Wallis test with Dunn's post hoc was used to compare the differences in continuous variables among three clinical groups (i.e., healthy control, polyp group and CRC group) and the nonparametric Mann-Whitney test was used to compare the differences in continuous variables between two groups of tissues (i.e., lesion vs peri-lesion or proximal colon vs distal colon). Spearman's rank correlation coefficient was used to determine the associations between continuous variables. The independent variables related to CRC or adenomas diagnosis were estimated using binary logistic regression model. The area under the receiver operating characteristic (ROC) curve (AUC) was used to evaluate the diagnostic value of bacterial candidates in discriminating CRC patients and adenomas groups versus healthy controls. Youden's index (J=Sensitivity-Specificity-1) was used to identify the best cut-off value that maximize sensitivity and specificity in the disease detection. *P*-value <0.05 was considered statistically significant.

CHAPTER V

RESULTS

5.1. Clinicopathological characterization

In this study, a total of 80 Thai participants (range, 51-85 years old) including 29 males and 51 females, were recruited between June 2019 and December 2020. Among these, 25 and 33 patients were diagnosed with CRC (CRC group) and adenomatous polyp (polyp group), respectively, according to the pathological results. The remaining participants were 22 healthy control subjects (HC group) whose colonoscopy showed no pathological finding. The mean age of volunteers was 64.8 years old, while the mean body mass index (BMI) was 23.1 kg/m^2. No statistical differences in age, BMI, and gender were observed among the three groups (Table 6 and Figure 5). Nonetheless, a significant difference in patients with diabetes mellitus (DM) was found among the three groups (p<0.05) (Table 6). Hypertension (HT) and dyslipidemia (DLP) did not significantly differ among all groups. In addition, more than 90% of adenoma patients and control subjects had the negative result of both fecal immunochemical tests (FIT and qFIT), whereas 16 out of 17 CRC cases were positive (p<0.001) (Table 6). Regarding the specimen data, approximately 61% (20/33) of adenomas were located in the proximal colon, and most of the adenomas were tubular adenoma (67%, 22/33) (Table 7). The pathological results of all CRC cases were adenocarcinomas and most of the malignant tumors were moderately differentiated (60%, 15/25) (Table 7). The majority of CRC was located at the distal colon and was found in stage III of the TNM staging system (60%, 15/25) (Table 7). The detailed demographic features of all participants are presented in Table 6 and 7.

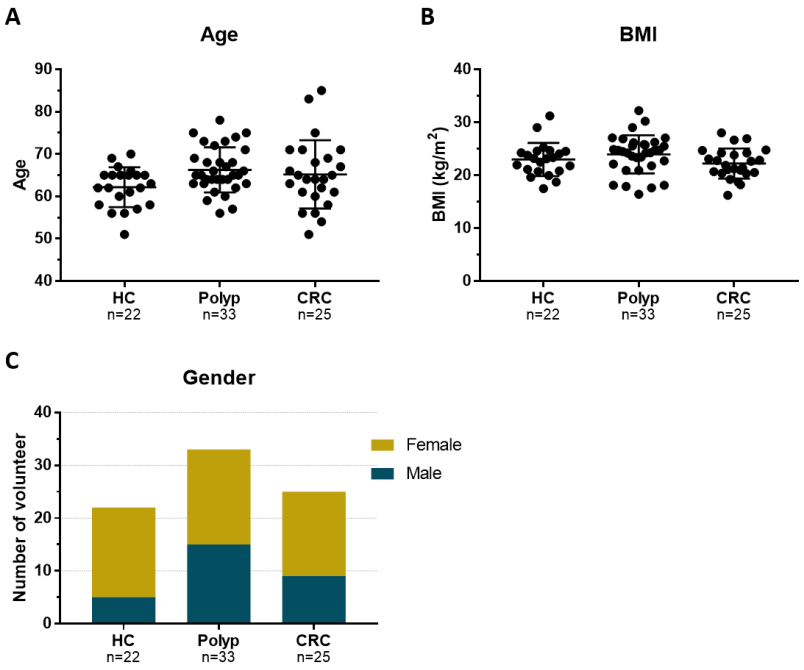


Figure 5 The demographic data of the participants in the study

A)age distribution; B) BMI distribution; C) volunteer quantity of each gender. Data in A and B are displayed as means ± SD.

Table 6 Clinical characteristics of participants in this study

Variables	Group			Total	*p*-value
	HC	Polyp	CRC		
No. of volunteers	22	33	25	80	
Age (Mean±SD)	62.2±4.7	66.3±5.3	65.2±8.1	64.8±6.3	0.068
BMI (Mean±SD)	23.0±3.1	24.0±3.6	22.2±2.8	23.1±3.3	0.055
Gender (n, (%))					
Male	5 (22.7)	15 (45.5)	9 (36)	29 (36.25)	0.229
Female	17 (77.3)	18 (54.5)	16 (64	51 (63.75)	
Diabetes Mellitus, DM (n, (%))					
Yes	0 (0)	4 (12.1)	7 (28)	11 (13.75)	0.020*
No	22 (100)	29 (87.9)	18 (72)	69 (86.25)	

Variables	Group			Total	*p*-value
Hypertension, HT (n, (%))					
Yes	8 (36.4)	15 (45.5)	10 (40)	33 (41.25)	0.789
No	14 (63.6)	18 (54.5)	15 (60)	47 (58.75)	
Dyslipidemia, DLP (n, (%))					
Yes	11 (50)	17 (51.5)	10 (40)	38 (47.5)	0.660
No	11 (50)	16 (48.5)	15 (60)	42 (52.5)	
Qualitative Fecal Immunochemical Test, FIT (n, (%))					
Positive	2 (9)	2 (6)	16 (64)	20 (25)	<0.001
Negative	20 (91)	29 (88)	1 (4)	50 (63)	***
No test	0 (0)	2 (6)	8 (32)	10 (13)	
Quantitative Fecal Immunochemical Test, qFIT (n, (%))					
Positive	1 (5)	3 (9)	16 (64)	20 (25)	<0.001
Negative	21 (95)	28 (85)	1 (4)	50 (63)	***
No test	0 (0)	2 (6)	8 (32)	10 (13)	

Note: no test, no fecal specimens for testing

Table 7 Clinicopathological characteristics of the samples in this study

Variables	Group			Total
	HC	Polyp	CRC	
No. of volunteers	22	33	25	80
Tumor location				n=58
A. Proximal colon (n, (%))	-	20 (60.6)	4 (16)	24 (41.4)
Cecum	-	4 (12.1)	1 (4)	5 (8.6)
Ascending colon	-	13 (39.4)	0 (0)	13 (22.4)
Hepatic flexure	-	1 (3.0)	0 (0)	1 (1.7)
Transverse colon	-	2 (6.1)	3 (12)	5 (8.6)
B. Distal colon (n, (%))	-	13 (39.4)	21 (84)	34 (58.6)

Splenic flexure	-	0 (0.)	2 (8)	2 (3.4)
Descending colon	-	3 (9.1)	5 (20)	8 (13.8)
Sigmoid colon	-	6 (18.2)	4 (16)	10 (17.2)
Rectosigmoid junction	-	2 (6.1)	1 (4)	3 (5.2)
Rectum	-	2 (6.1)	9 (36)	11 (19)
TNM classification				
Tumor stage (T) (n, (%))				
T1	-	-	2 (8)	
T2	-	-	5 (20)	
T3	-	-	15 (60)	
T4	-	-	3 (12)	
Node stage (N) (n, (%))				
N0	-	-	15 (60)	
N1	-	-	6 (24)	
N2	-	-	4 (16)	
Metastasis stage (M) (n, (%))				
Mx	-	-	10 (40)	
M0	-	-	14 (56)	
M1	-	-	1 (4)	
Pathological result (n, (%))				
I. Tubular adenoma	-	22 (67)	-	
II. Tubulovillous adenoma	-	2 (6)	-	
III. Sessile serrated adenoma	-	1 (3)	-	
IV. Traditional serrated adenoma	-	1 (3)	-	
V. Tubular adenoma and tubulovillous adenoma	-	1 (3)	-	
VI. Tubulovillous and	-	1 (3)	-	

	sessile serrated adenoma			
VII.	Tubular adenoma and hyperplastic polyp	-	3 (9)	-
VIII.	Tubular adenoma, hyperplastic polyp, and inflammatory polyp	-	2 (6)	-
IX.	Adenocarcinoma	-	-	25 (100)
Tumor differentiation (n, (%))				
	Well	-	-	6 (24)
	Moderate	-	-	15 (60)
	Poor	-	-	0 (0)
	No report	-	-	4 (16)

5.2. Sample collection

A total of 179 samples consisting of 70 stool samples and 109 tissue samples were collected from all participants and are summarized in Table 8.

Table 8 Summary information of sample collection

	HC			**Polyp**			**CRC**			**Total**
No. of subjects	22			33			25			80
Sample types	Stool	Tissue		Stool	Tissue		Stool	Tissue		
Sampling location	22	PC	DC	31	PL	NL	16	PL	L	
No. of sample		17	17		26	26		16	7	
Total	22	34		31	52		17	23		179

Abbreviations: PC, proximal colon; DC, distal colon; PL, peri-lesion; NL, non-lesion; L, lesion.

5.3. Characteristics of sequencing results

Sequencing of 16S rRNA gene amplicons of DNA extracted from 146 samples (70 stool samples: 22 HC, 31 polyp and 17 CRC, and 76 tissue samples: 34 HC, 26 polyp and 16 CRC) retrieved an overall number of 5,813,182 reads, with an average of 39,816 reads per samples. In total, 1,567 amplicon sequence variants (ASVs) were delineated at 97% similarity threshold and a total of 1,418 low abundance features and 30 low variance features were removed based on prevalence and the inter-quartile range, respectively. After removal of ASVs with unmet quality, 119 ASVs were used for further analyses. Besides, the Good's coverage value of each group was >99% (data not shown).

5.4. Comparison of bacterial microbiota between fecal and mucosa tissue samples

5.4.1. Alpha-diversity and beta-diversity analyses

The Chao1's index was used to evaluate taxa richness while the Shannon and Simpson diversity indexes were applied to estimate both richness and evenness of fecal and mucosa tissue samples. All alpha-diversity measures (Chao1, Shannon and Simpson's index) of fecal samples (n=70) were significantly high compared with mucosa tissue samples (n=76) (*p*-value <0.01, 0.001, and 0.01, respectively) (Figure 6A-6C). In terms of beta diversity, a principal coordinate analysis (PCoA) plot under a Bray-Curtis distance was performed to compare the overall structure of gut microbiota between sample types. The PCoA plots showed significantly separated clusters between two sample types (*p*-value <0.001) (Figure 6D, 6E).

5.4.2. Relative abundance and composition of microbiota

At the bacterial phylum level, the bacterial pattern revealed that Bacteroidetes was the most predominant phylum, contributing 59.2% and 52.5% of the fecal and mucosal tissue samples, respectively. The second and the third most abundant phylum in fecal samples were Firmicutes (24.6%) and Proteobacteria

(11.4%), respectively. On the other hand, Proteobacteria (25.4%) and Firmicutes (11.1%) were found as the second and the third bacterial abundance, respectively, in tissue samples (Figure 7A, 7B). Moreover, the relative abundance of the dominant bacterial genera was shown in Figure 7C. The top three dominant bacterial genera were *Bacteroides*, unspecified genera, and *Faecalibacterium* in fecal samples but were *Bacteroides*, *Escherichia_Shigella*, and *Fusobacterium* in tissue samples. These results indicated that the gut microbiota of the fecal samples referred to as lumen-associated microbiota were different from the mucosa-associated microbiota.

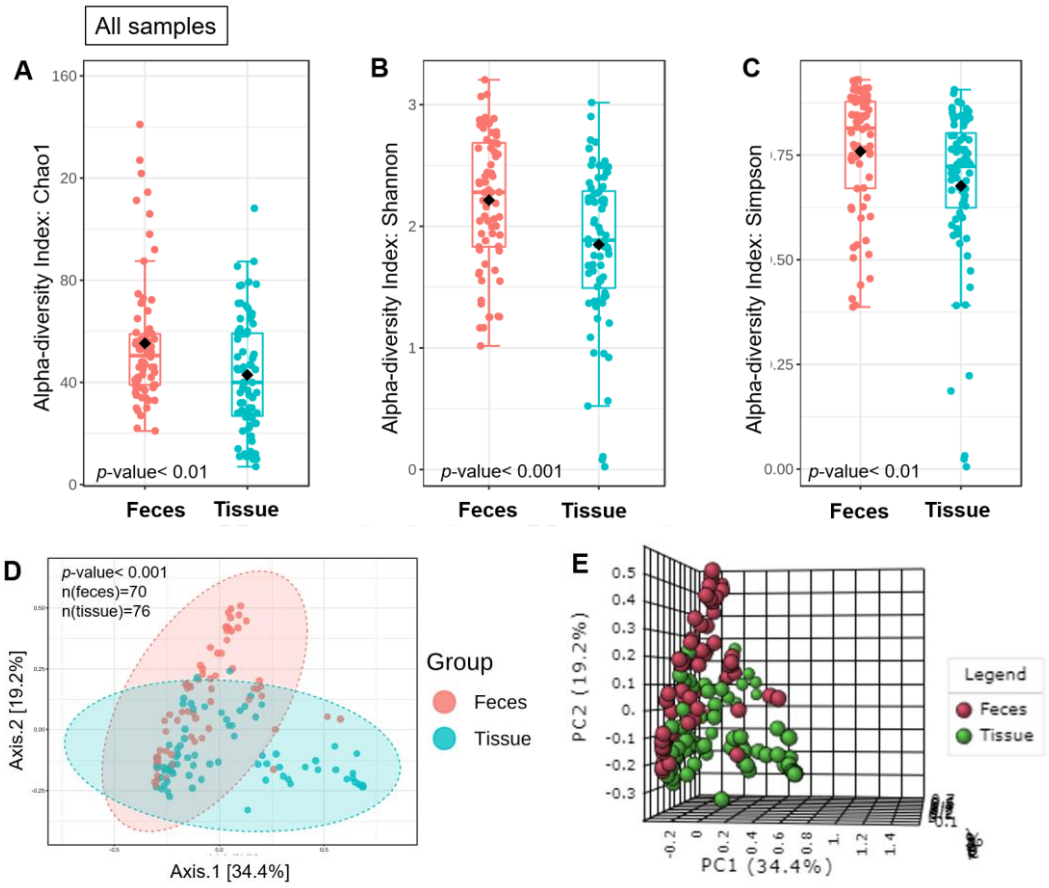


Figure 6 Alpha-diversity and beta-diversity analyses between stool samples and mucosal tissue samples.

A) Chao1's index, B) Shannon's index, C) Simpson's index, D) principal coordinate analysis (PCoA) based on Bray-Curtis distance in 2-Dimension, E) PCoA analysis based on Bray-Curtis distance in 3-Dimension.

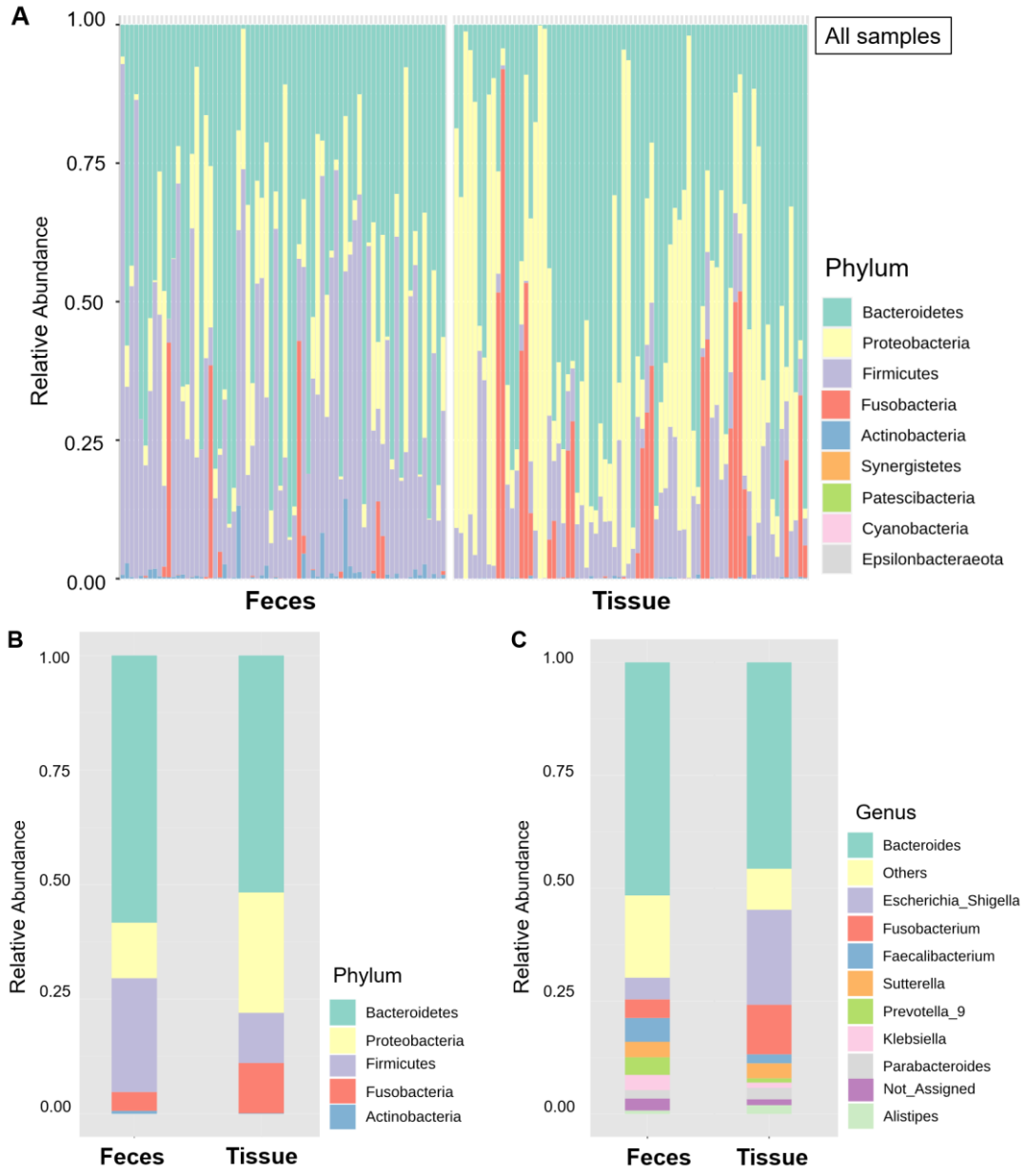


Figure 7 The differences in microbial abundance profiling between stool samples and mucosal tissue samples.

A) Taxonomic composition of each sample type at the phylum level, B) microbiota composition of merged samples at the phylum level, C) microbiota composition of merged samples at the genus level. Data are shown as relative abundance.

5.5. Comparison of mucosa-associated microbiota among adenocarcinoma, adenoma, and HC subjects

5.5.1. Alpha-diversity and beta-diversity analyses

No difference in alpha diversity was observed among patients with CRC (n=16), patients with adenomas (n=26), and the HC group (n=34) (Figure 8A-8C). As for beta diversity, the PCoA plot of mucosa microbiota revealed statistically significant ($p<0.05$) among the three groups (Figure 8D, 8E). The significant difference was probably between the CRC and HC groups ($p<0.01$) (Figure A1).

5.5.2. Relative abundance and composition of mucosa-associated microbiota

Overall microbial compositions of the polyp and the CRC groups were shifted compared to that of controls as a baseline. At the phylum level, the most prevalent phylum in the polyp and the CRC groups was Bacteroidetes (57.1% and 52.1%) while in the HC group was Proteobacteria (40.6%) (Figure 9A, 9B). At the genus level, a stepwise increase of *Bacteroides* and *Parabacteroides* in the CRC group (55.2% and 3.5%) was observed when compared with the polyp (50.4% and 3.4%) and the HC (36.0% and 1.2%) groups. Furthermore, *Escherichia_Shigella* and *Faecalibacterium* in the polyp (8.8% and 2.4%) and the CRC (16.4% and 0.7%) groups were decreased compared with the HC (35% and 4.0%) group. Additionally, *Fusobacterium* was over presented in the polyp group (17.9%) when compared with other groups (8.1% for CRC and 7.9% for HC) (Figure 9C).

Owing to the pattern differences of mucosa microbial composition among groups, a heat tree was additionally generated to illustrate the group-wise relative abundance of significant bacterial genera (*p*-value <0.05). The taxonomic tree also

showed the distinct bacterial abundance between groups by the color gradient. For HC versus CRC, the *Escherichia_Shigella* (Gamma-proteobacteria), *Faecalibacterium,* and a member from the family *Lachnospiraceae* were present significantly more in the HC group (yellow color) than the CRC group (green color) (Figure 10A). In contrast, *Bacteroides, Parabacteroides*, and *Butyricimonas* which belong to the order Bacteroidales as well as *Collinsella*, *Erysipelatoclostridium*, and those genera from the family *Ruminococcaceae* had a greater proportion in CRC patients compared with those of HC (Figure 10A). However, only *Flavobacterium* was significantly higher in the polyp group compared with that of controls (Figure 10B).

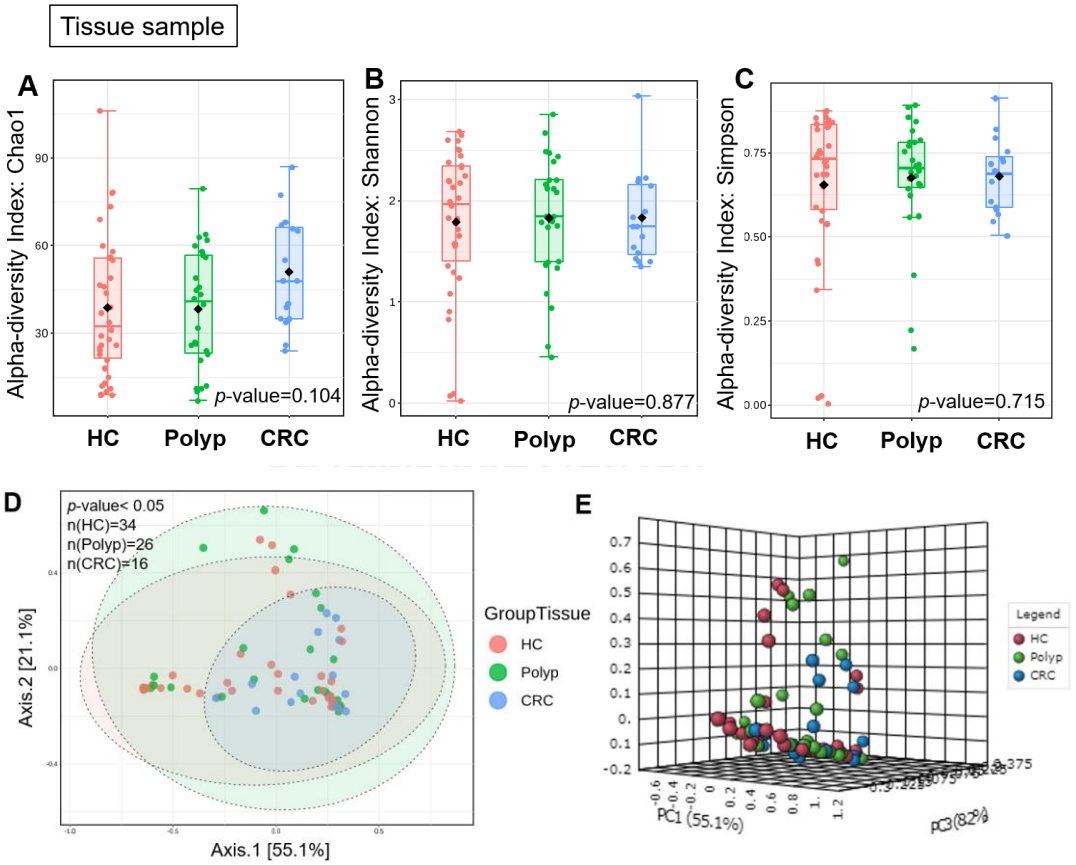


Figure 8 Alpha-diversity and beta-diversity analyses in mucosal tissue samples among HC subjects, adenomas subjects, and CRC subjects.

A) Chao1's index, B) Shannon's index, C) Simpson's index, D) PCoA analysis based on Bray-Curtis distance in 2-Dimension, E) PCoA analysis based on Bray-Curtis distance in 3-Dimension.

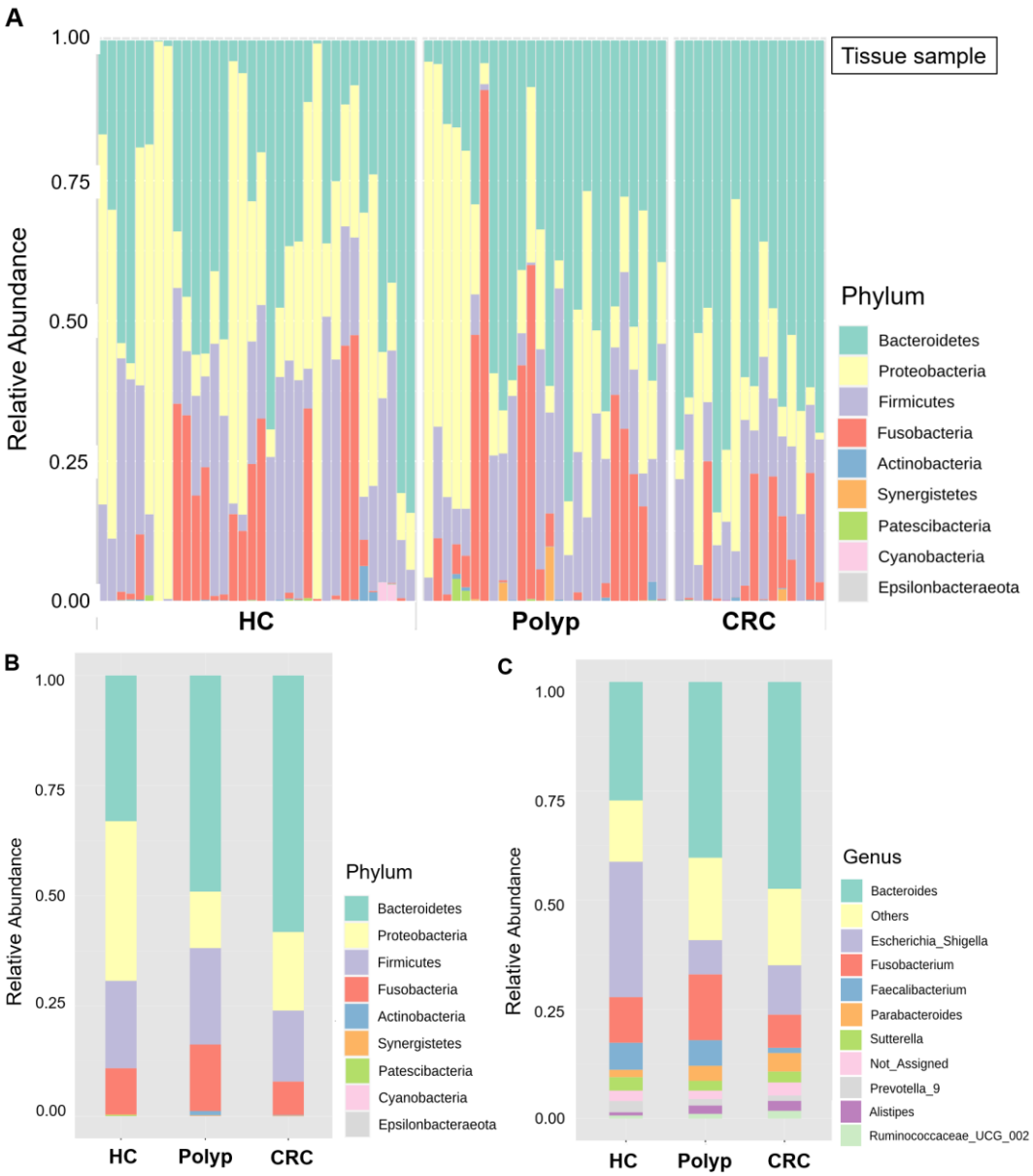


Figure 9 The differences in microbial abundance profiling of mucosa tissue samples among HC subjects, adenomas subjects, and CRC subjects.

A) Taxonomic composition of each sample type at the phylum level, B) microbiota composition of merged samples at the phylum level, C) microbiota composition of merged samples at the genus level. Data are shown as relative abundance.

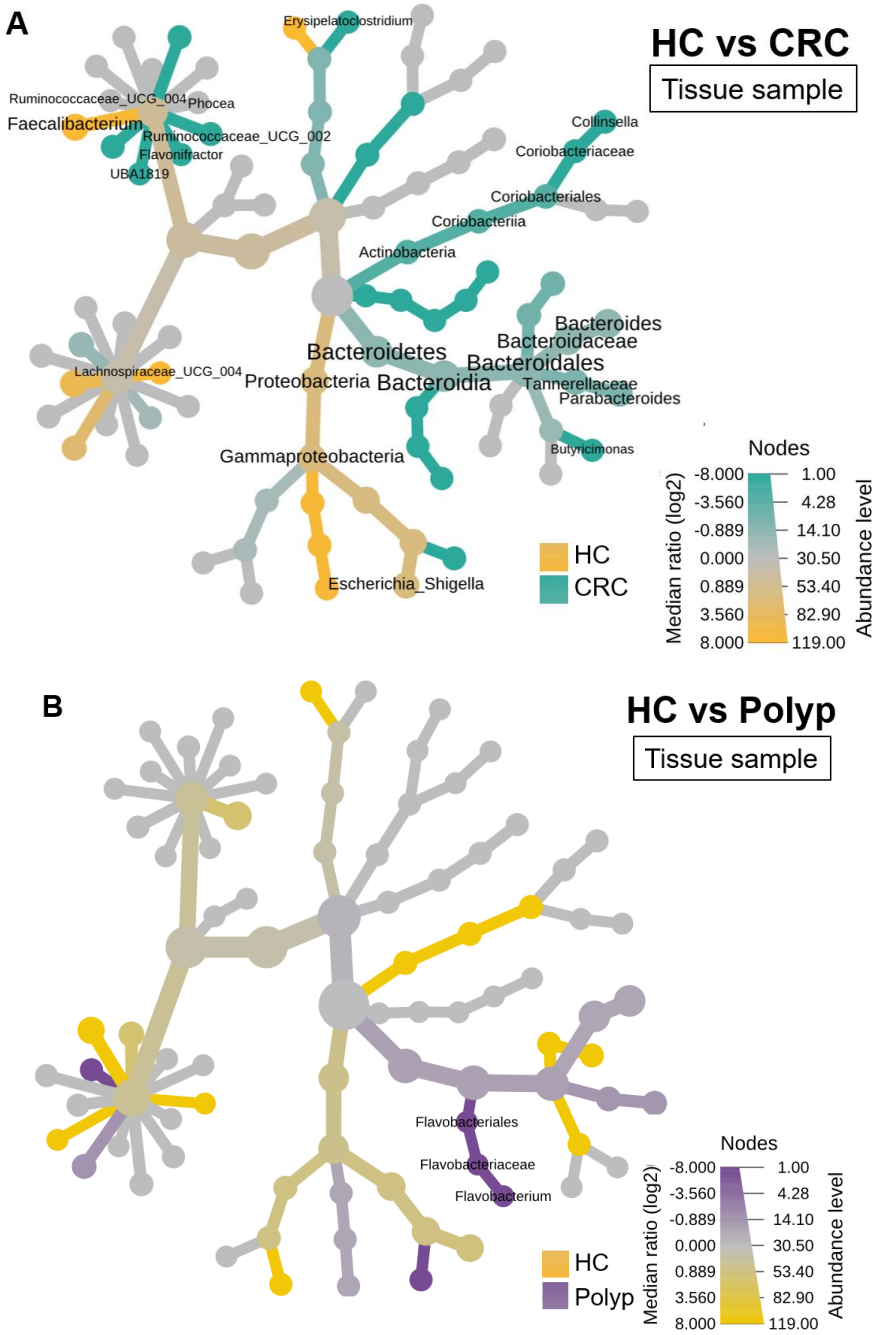


Figure 10 Taxonomical differential analysis of the mucosa-associated microbiota.

The heat tree shows only the genera or higher classification that was significantly different between two groups. A) HC versus CRC, B) HC versus polyp. Abundances of each taxon are given by the node size and color gradients mean statistically significant differences in taxa abundance assessed by a non-parametric Wilcoxon test (green and purple mean higher abundance in the CRC and adenoma, respectively, and yellow means higher abundance in the HC group).

5.5.3. Significant differential abundance of mucosa-associated bacterial species

Regarding the significant difference in the bacterial abundance at the species level, a total of 7 differential species were estimated using the classical univariate comparison as shown in Figure 11. Notably, *Erysipelatoclostridium ramosum* (ER), *Bacteroides thetaiotaomicron* (BT), *Flavonifractor plautii*, *Parabacteroides merdae* and *P. distasonis* were escalated in patients with CRC compared with other groups (all, *p*-value cut-off <0.05, except ER and BT, FDR-adjusted *p*-value<0.05), while the abundances of *Escherichia_shigella_coli* and Not_assigned bacteria declined (*p*-value<0.05).

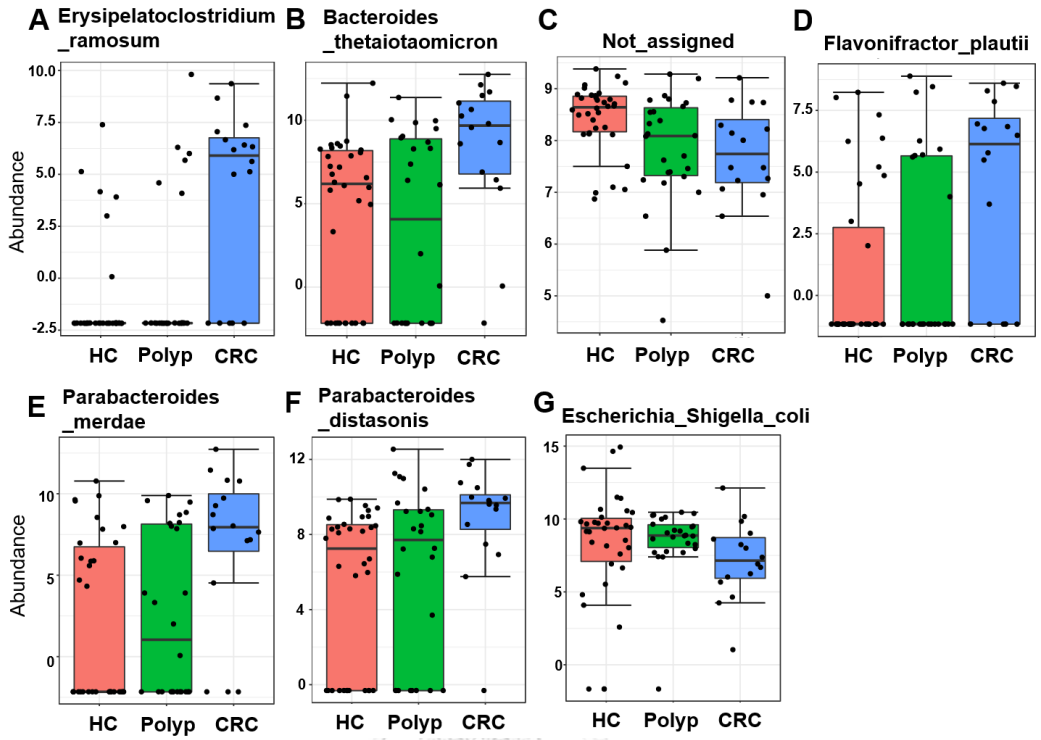


Figure 11 The significant difference of individual bacterial abundance in mucosal tissue samples at the species level among HC subjects, adenomas subjects, and CRC subjects.

A) *Erysipelatoclostridium ramosum*, B) *Bacteroides thetaiotaomicron*, C) Not_assigned, D) *Flavonifractor plautii*, E) *Parabacteroides merdae*, F) *Parabacteroides distasonis*, G) *Escherichia_Shigella_coli*. Data are shown as log-transformed count.

5.6. Comparison of lumen-associated microbiota among adenocarcinoma, adenoma, and HC subjects

5.6.1. Alpha-diversity and beta-diversity analyses

Chao1, Shannon and, Simpson's metrics were used to evaluate the richness and diversity aspects of luminal-associated microbiota among groups, but no statistically significant difference in these indexes were found (Figure 12A-12C). Concerning beta diversity, PCoA analysis based on Bray-Curtis dissimilarity did not show a separated trend among the three groups (Figure 12D, 12E).

5.6.2. Relative abundance and composition of lumen-associated microbiota

Of these 9 phyla, the top 3 phyla comprising Bacteroidetes, Firmicutes, and Proteobacteria were observed in all groups (Figure 13A, 13B). In addition, Proteobacteria and Fusobacterium had higher proportions in the polyp (12.8% and 8.3%) and the CRC (15.0% and 2.9%) groups compared with the HC group (8.5% and 1.2%) (Figure 13A, 13B). At the genus level, as shown in Figure 13C, *Bacteroides* was less abundant in fecal samples of the polyp (46.6%) and the CRC (50.4%) groups compared with the HC (64.9%), while *Faecalibacterium* had a lower proportion in CRC (2.8%) subjects compared with other groups (8.3% for the polyp group and 5.4% for the HC group). Conversely, the proportion of *Escherichia_Shigella*, *Fusobacterium*, *Klebsiella, Sutterella, Dorea,* and *Parabacteroides* were ascended in the polyp, or the CRC group compared with the HC group.

In terms of heat tree analysis (Figure 14), *Agathobacter* and CAG_56 bacteria (family *Lachnospiraceae*), *Faecalibacterium* (family Rumiococcaceae) and a member of family *Erysipelatotrichaceae* were more significantly abundant in the HC group than in the CRC group (*p*-value <0.05) (Figure 14A). On the contrary, the genera *Parabacteroides* and *Butyricimonas* had lower fold change in the HC when compared with the CRC group (*p*-value <0.05) (Figure 14A). Regarding the HC group versus the CRC group, the genus *Oscillibacter* was more present in the HC compared with the polyp group (*p*-value <0.05) (Figure 14B). Besides, *Fusicatenibacter* within the family *Lachnospiraceae* was more specific to the polyp group than the HC group (*p*-value <0.05) (Figure 14B).

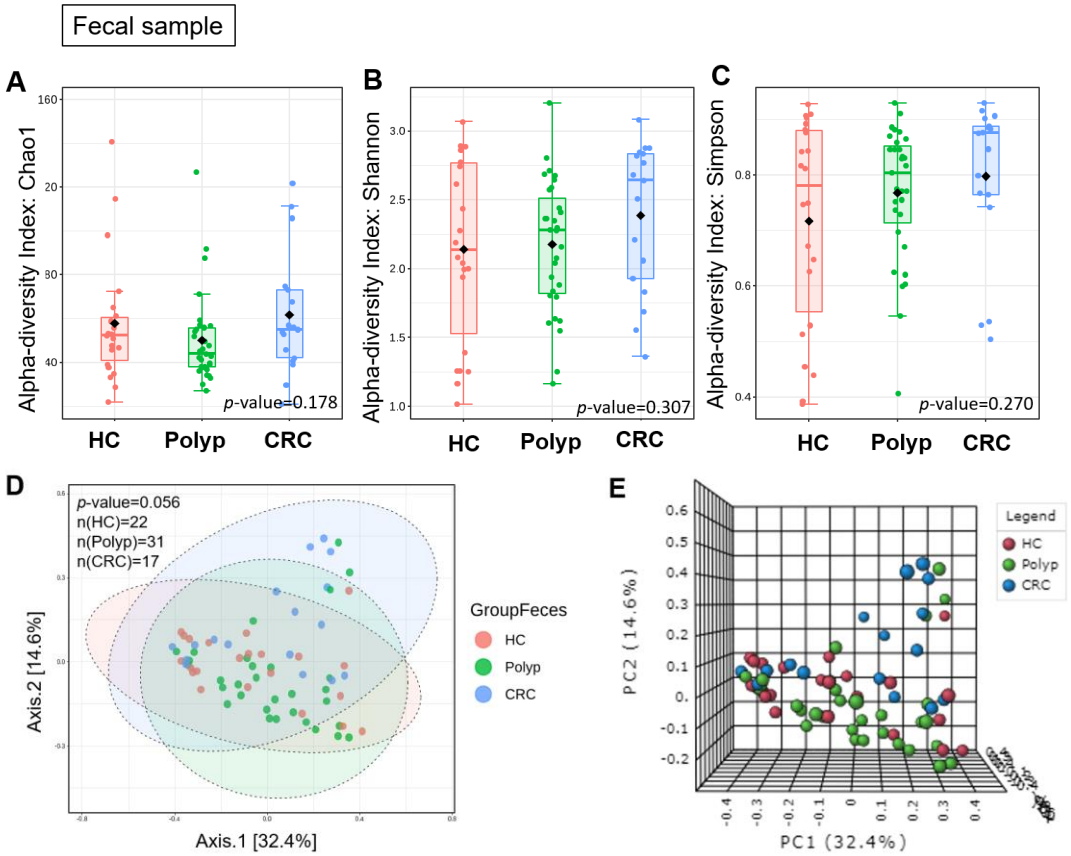


Figure 12 Alpha-diversity and beta-diversity analyses in fecal samples among HC subjects, adenoma subjects, and CRC subjects.

A) Chao1's index, B) Shannon's index, C) Simpson's index, D) PCoA analysis based on Bray-Curtis distance in 2-Dimension, E) PCoA analysis based on Bray-Curtis distance in 3-Dimension.

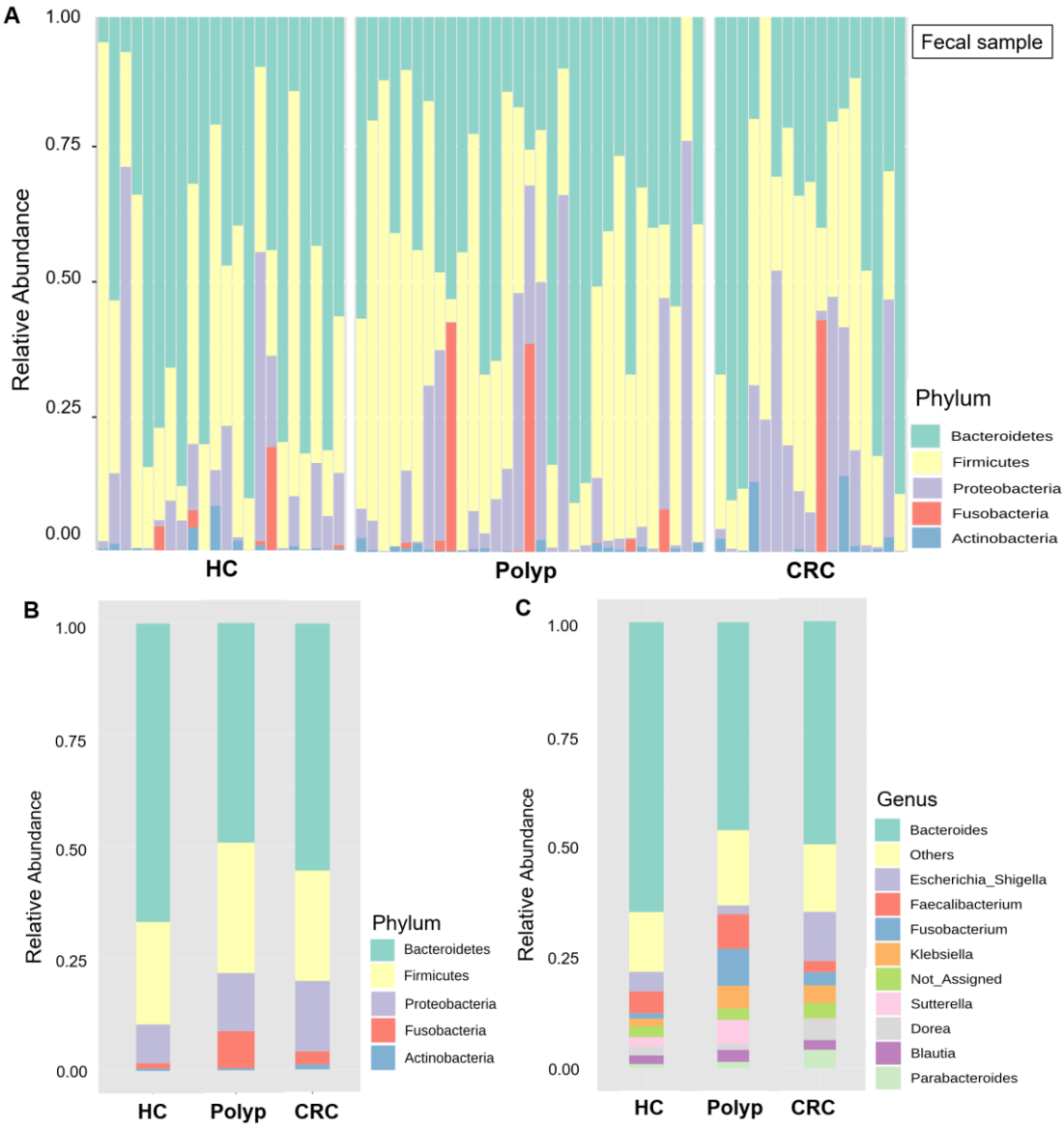


Figure 13 The differences in microbial abundance profiling of fecal samples among HC subjects, adenoma subjects, and CRC subjects.

A) Taxonomic composition of each sample type at the phylum level, B) microbiota composition of merged samples at the phylum level, C) microbiota composition of merged samples at the genus level. Data are shown as relative abundance.

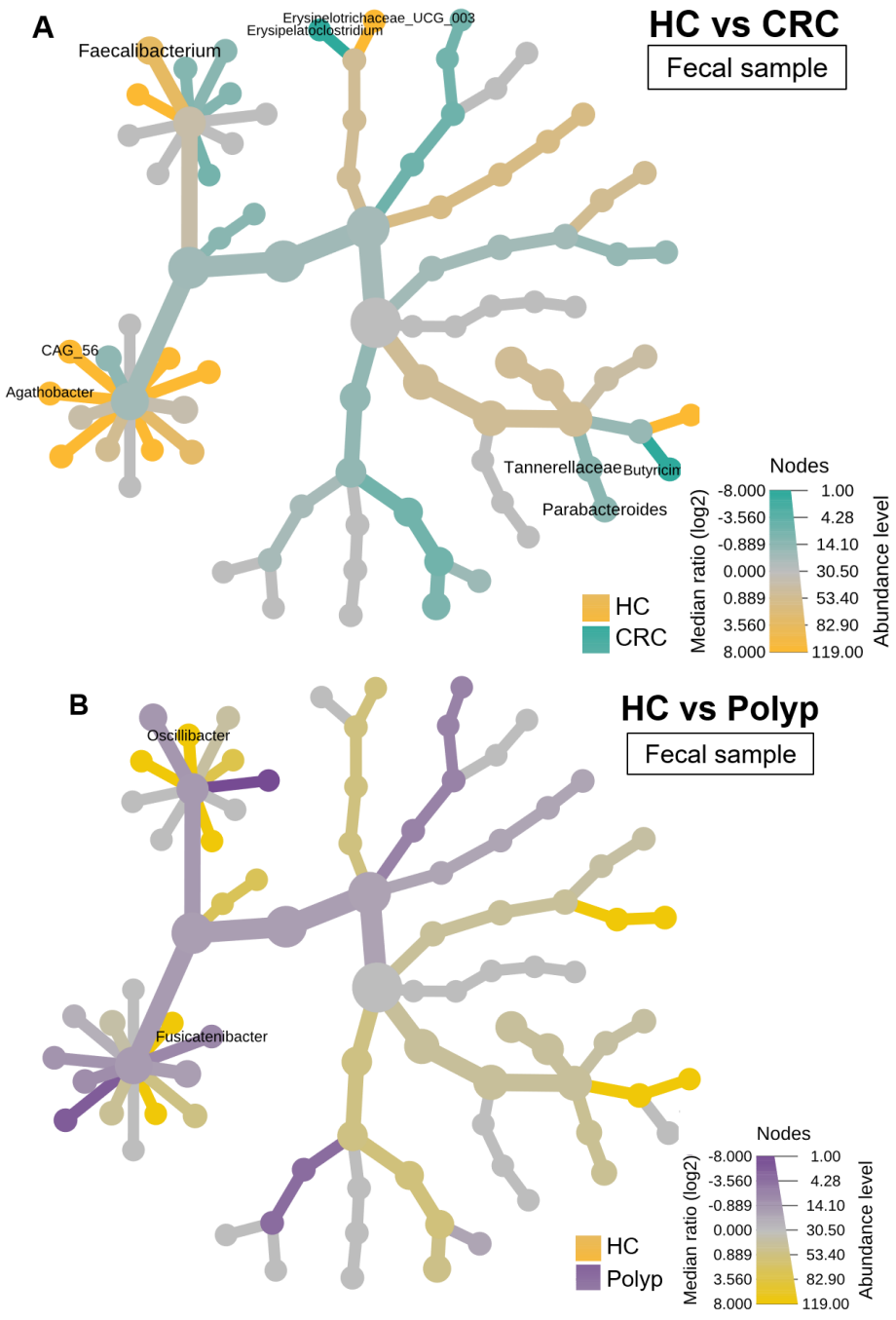


Figure 14 Taxonomical differential analysis of the lumen-associated microbiota.

The heat tree shows only the genera or higher classification that was significantly different between two groups.

A) HC versus CRC, B) HC versus polyp. Abundances of each taxon are given by the node size and color gradients mean statistically significant differences in taxa

abundance assessed by a non-parametric Wilcoxon test (green and purple mean higher abundance in the CRC and adenoma, respectively, and yellow means higher abundance in the HC group).

5.6.3. Significant differential abundance of lumen-associated bacterial species

To observe the significant difference of relative abundance at the species level, *E. ramosum* and *Eggerthella lenta* in CRC patients had higher proportion than those of HC subjects (FDR-adjusted *p*-value<0.05 and *p*-value<0.05) (Figure 15A, 15C). However, *B. vulgatus* was higher in HC subjects (*p*-value<0.05) (Figure 15B).

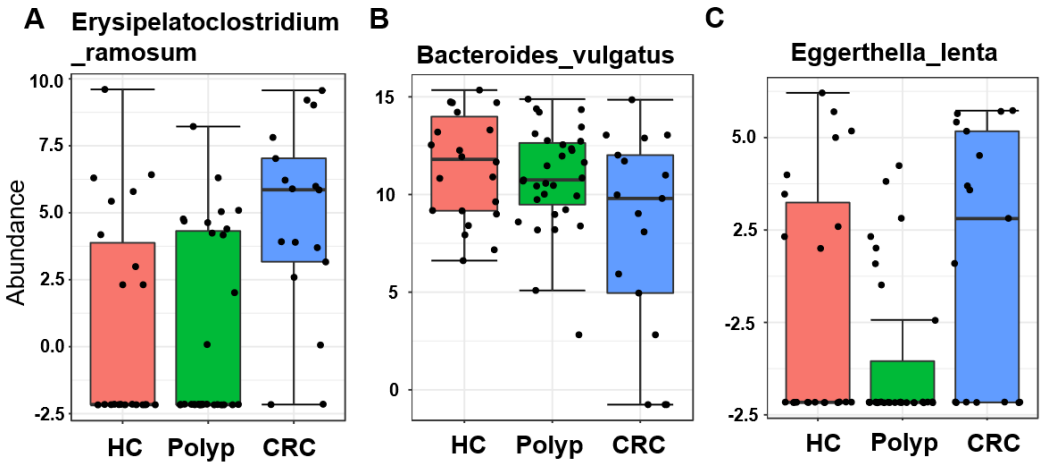


Figure 15 The significant difference of bacterial abundance in fecal samples at the species level among HC subjects, adenoma subjects, and CRC subjects.
A) *Erysipelatoclostridium ramosum*, B) *Bacteroides vulgatus*, C) *Eggerthella lenta*. Data are shown as log-transformed count.

5.7. Identification of putative biomarkers for CRC/adenoma

To further evaluate the bacteria in clinical samples as biomarkers for CRC, linear discriminant analysis (LDA) coupled with effect size (LEfSe) algorithm was used at the LDA cut-off of ± 3. Six out of 7 bacterial taxa from tissue samples had a LDA score of more than 4 (Figure 16A). Five out of 7 taxa comprising of *B.*

thetaiotaomicron, P. merdae, *P. distasonis*, *E. ramosum*, and *F.plautii* were over-represented and 2 out of 7 taxa including unspecified bacteria and *Escherichia_Shigella_coli* were under-represented in the CRC group compared with other groups (all, *p*-value<0.05, except ER, FDR-adjusted *p*-value< 0.05) (Figure 16A). As for fecal samples, the LDA scores of *E. ramosum* and *B. vulgatus* had more than 4 (Figure 16A). *E. ramosum* and *E. lenta* were enriched in the CRC group, while *B. vulgatus* was predominant in the HC group (all, *p*-value<0.05, except ER, FDR-adjusted *p*-value< 0.05) (Figure 16A). Additionally, only *E. ramosum* was found specifically in the CRC patients in both fecal and tissue samples (Figure 16B). As for adenoma, putative biomarkers which can differentiate adenomas patients from the HC group were not found in this analysis.

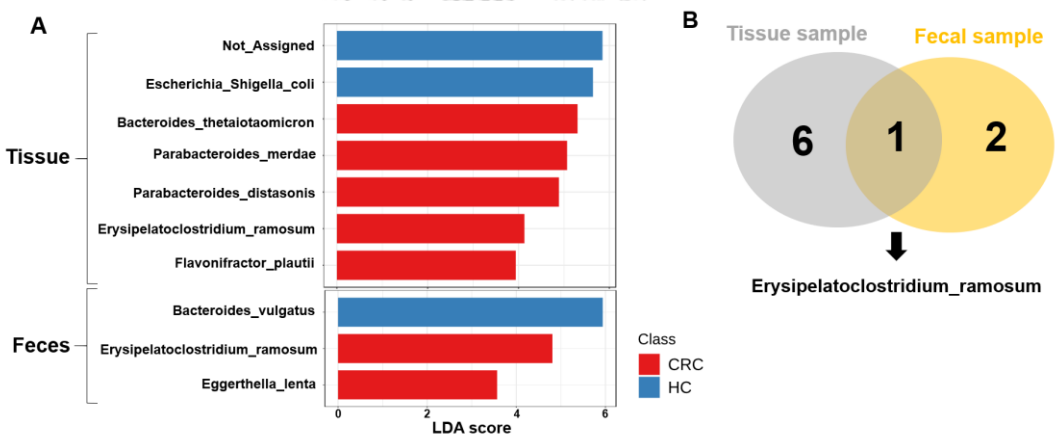


Figure 16 LEfSe analysis of mucosal tissue and fecal microbiota among CRC and HC subjects.

A) Histogram of the LDA scores for significantly abundant species, B) Venn diagram represented the number of unique and overlapping significantly abundant species.

5.8. Bacterial quantification in the clinical samples

To quantitate CRC-associated bacteria in the clinical samples, we conducted the absolute quantification of six interesting bacteria in CRC, consisting of *Fusobacterium nucleatum* (FN), colibactin positive strains (EC), *Parvimonas micra*

(PM), *Blautia* spp. (Bla), *Streptococcus gallolyticus* (SG), and *Fusicatenibacter saccharivorans* (FS), in both mucosal tissues and feces from three groups by qPCR, using a serial dilution of a plasmid carrying the gene specific to each bacterium as a standard curve. In this study, *S. gallolyticus* and *F. saccharivorans* were only performed in fecal samples. At first, the quantity of SG and FS in tissue samples was also investigated, however the tissue burden of SG and FS could not be accurately determined because of poor specificity of primers used (non-specific amplification with host tissue). Due to possible variation at different locations, the mucosal tissues of CRC included lesional tissue, peri-lesional tissue, non-lesional tissue, whereas normal tissue of the HC group was from both sides (proximal and distal) of the colon.

The pairwise comparisons of individual bacterial quantity between tissue samples from distinct areas of the colon were preliminarily analyzed i.e., proximal colon vs distal colon for HC group, peri-lesional vs non-lesional for the polyp group, and peri-lesional vs lesional for the CRC group. As shown in Figure 17, no significant difference in bacterial colonization was found between two tissue types across entire groups for any of the bacteria in this study. Therefore, the bacterial quantity of both tissue types was combined and used as the total tissue of each group for further comparison.

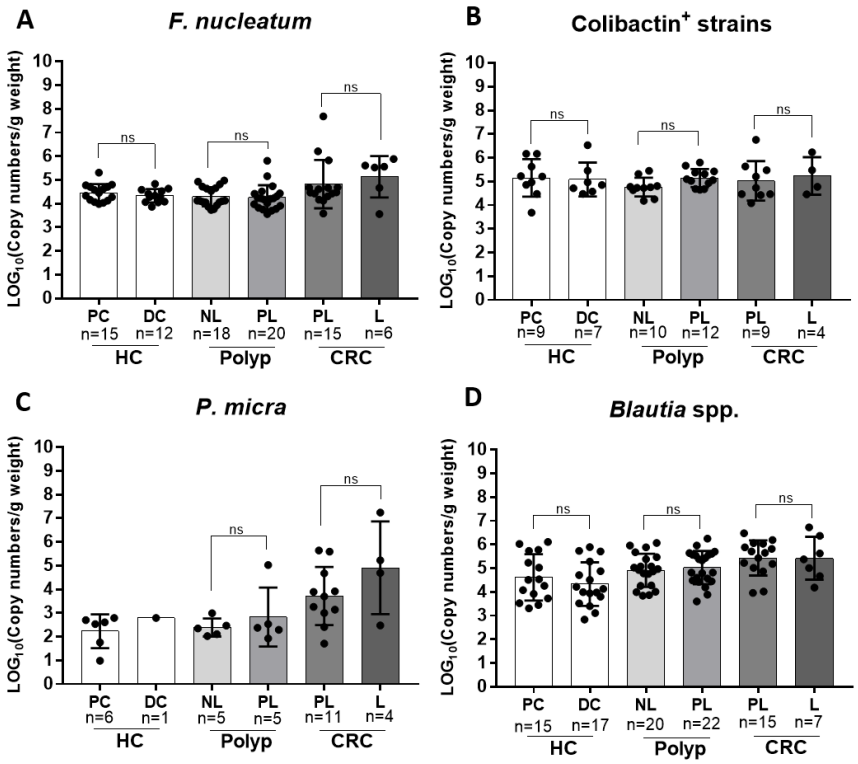


Figure 17 The comparison of absolute quantity of CRC-associated bacteria between two types of tissues in each group.

A) *F. nucleatum*, B) colibactin positive strains, C) *P. micra*, D) *Blautia* spp. Each scatter plot is expressed as log10 copy number per gram weight and data are displayed as means ± SD. Each dot represents one sample. Abbreviations: PC, proximal colon; DC, distal colon, NL, non-lesion; PL, peri-lesion; L, lesion; ns, not significant.

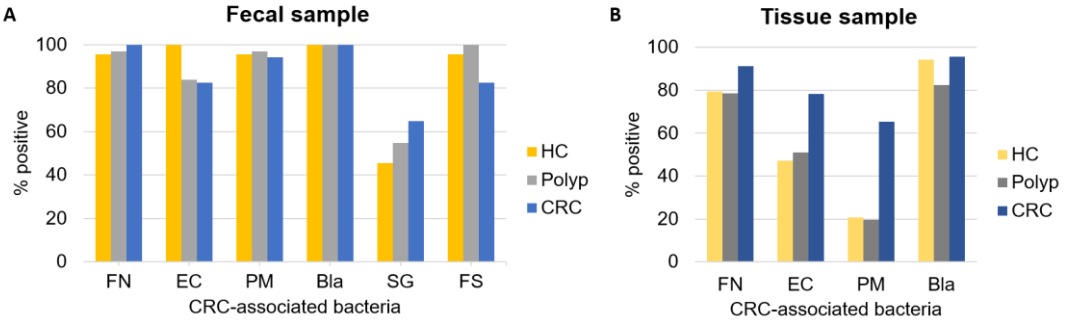


Figure 18 The prevalence of CRC-associated bacteria in the clinical samples.

A) fecal samples; B) tissue samples. Abbreviations: FN, *F.nucleatum*; EC, colibactin positive strains; PM, *P.micra*; Bla, *Blautia* spp.; SG, *S.gallolyticus*; FS, *F.saccharivorans.*

The qPCR assay targeting the *nusG* gene was used to detect *F. nucleatum*. The *nusG* gene was detected in more than 95% of fecal samples from all groups (Figure 18A). The detection frequency of *F. nucleatum* in CRC tissue (91%) was slightly higher than that in the HC (79%) and the polyp groups (78%) (Figure 18B). The absolute abundance of *F. nucleatum* in stool was predominantly higher in patients with CRC compared with the polyp group ($P<0.001$) and HC ($P<0.001$) (Figure 19A). Furthermore, FN in tissue was also significantly higher in CRC cases compared with the polyp group ($P<0.05$) (Figure 19A). No difference of *nusG* level was found between patients with polyp and HC in both sample types (Figure 19A).

The presence of *clbB* gene, a part of the *pks* pathogenicity island encoding a polyketide-peptide genotoxin (colibactin) of *Enterobacteriaceae* mainly in *Escherichia coli*, was used to assess colibactin positive bacteria (EC) in the samples. The $clbB^+$ bacteria in 78% of CRC tissues were higher than those found in tissues of HC (47%) and the polyp groups (51%) (Figure 18B). In contrast, $clbB^+$ bacteria were detected in all fecal samples of the HC group, while they were found in 84% and 82% of the polyp and the CRC groups, respectively (Figure 18A). However, no significant difference in quantity of $clbB^+$ bacteria between groups for both types of samples. Notably, a significant enrichment of $clbB^+$ bacteria was found in patients with stage III CRC when compared with controls ($P<0.05$, Figure 20C).

In this study, *P. micra* was found in 65% of tissue samples from CRC patients but was found only 21% and 20% of tissues from the HC and polyp groups, respectively (Figure 18B). This bacterium was detected in over 94% of all fecal samples (Figure 18A). *P. micra* was significantly enriched in the stool of patients with CRC compared with the polyp group ($P<0.01$) and the HC group ($P<0.01$) (Figure 19C). This is consistent with the results of tissue samples, in which the bacterium was also

more abundant in CRC tissues compared with the polyp group (P<0.05) and the HC group (P<0.05) (Figure 19C).

Of these bacteria, *Blautia* spp. was most commonly found in all sample types, more than 80% and 100% in tissue and feces of all cases, respectively (Figure 18). No difference was found in the amount of *Blautia* in stool, but a significantly higher level of *Blautia* in cancer tissue was found compared with control tissues (P<0.001) (Figure 19D). In addition, the significant positive correlation between *Blautia* in tissue and TNM staging was observed (Spearman r_s = 0.5893, Figure 20G and 20H). The positive detection of *S. gallolyticus* in fecal samples increased stepwise from HC (45%), polyp (55%), and CRC (65%) (Figure 18A). The fecal SG levels seemed to increase in the CRC group. However, no significant difference was found in fecal samples among the three groups (Figure 19E). *F. saccharivorans* was commonly detected in stool samples of all groups (82%-100%) (Figure 18A). The abundance of this bacterium significantly decreased in the stool of CRC patients compared with the polyp group (P<0.05) (Figure 19F).

In addition, bivariate correlation analysis showed that the absolute quantity of FN, PM, and EC was significantly correlated with TNM classification (Spearman r_s = 0.425-0.667, Figure 20A-20F). The significant difference of the bacterial abundance was mostly found in the late stages (stage III, IV) as shown in Figure 20A, 20C and 20F. Thus, these bacteria were further selected for the CRC prediction test.

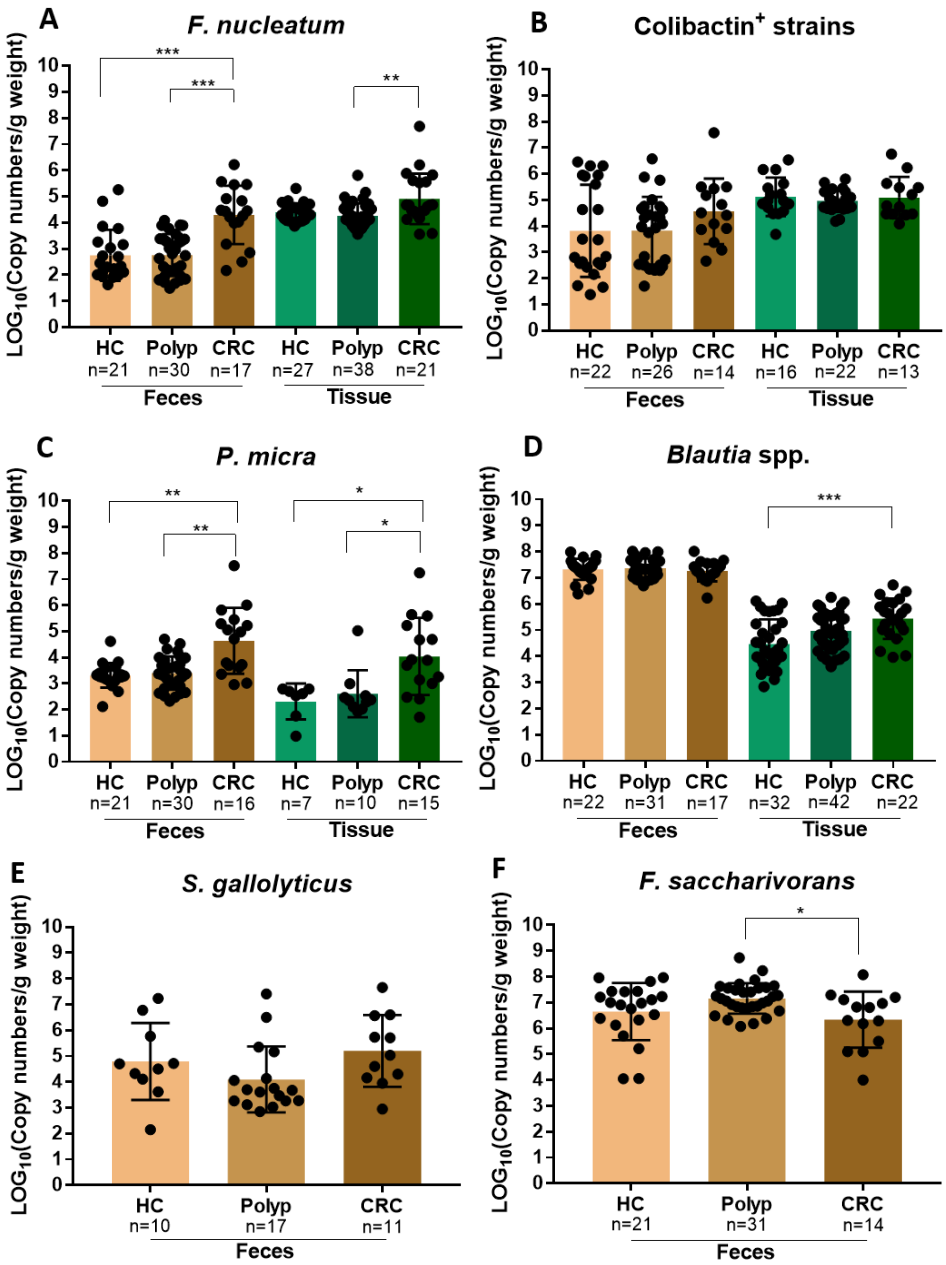


Figure 19 Absolute quantification of CRC-associated bacteria in the clinical samples. A) *F. nucleatum*, B) colibactin positive bacteria, C) *P. micra*, D) *Blautia* spp., E) *S. gallolyticus*, F) *F. saccharivorans*. Each scatter plot is expressed as log10 copy number per gram weight and data are displayed as means ± SD. Each dot represents one sample, and each bar of tissue sample represents as total tissue (both types of tissue). Brown bar, fecal sample; green bar, tissue sample. Abbreviations: *, $P<0.05$; **, $P<0.01$; ***, $P<0.001$; ****, $P<0.0001$.

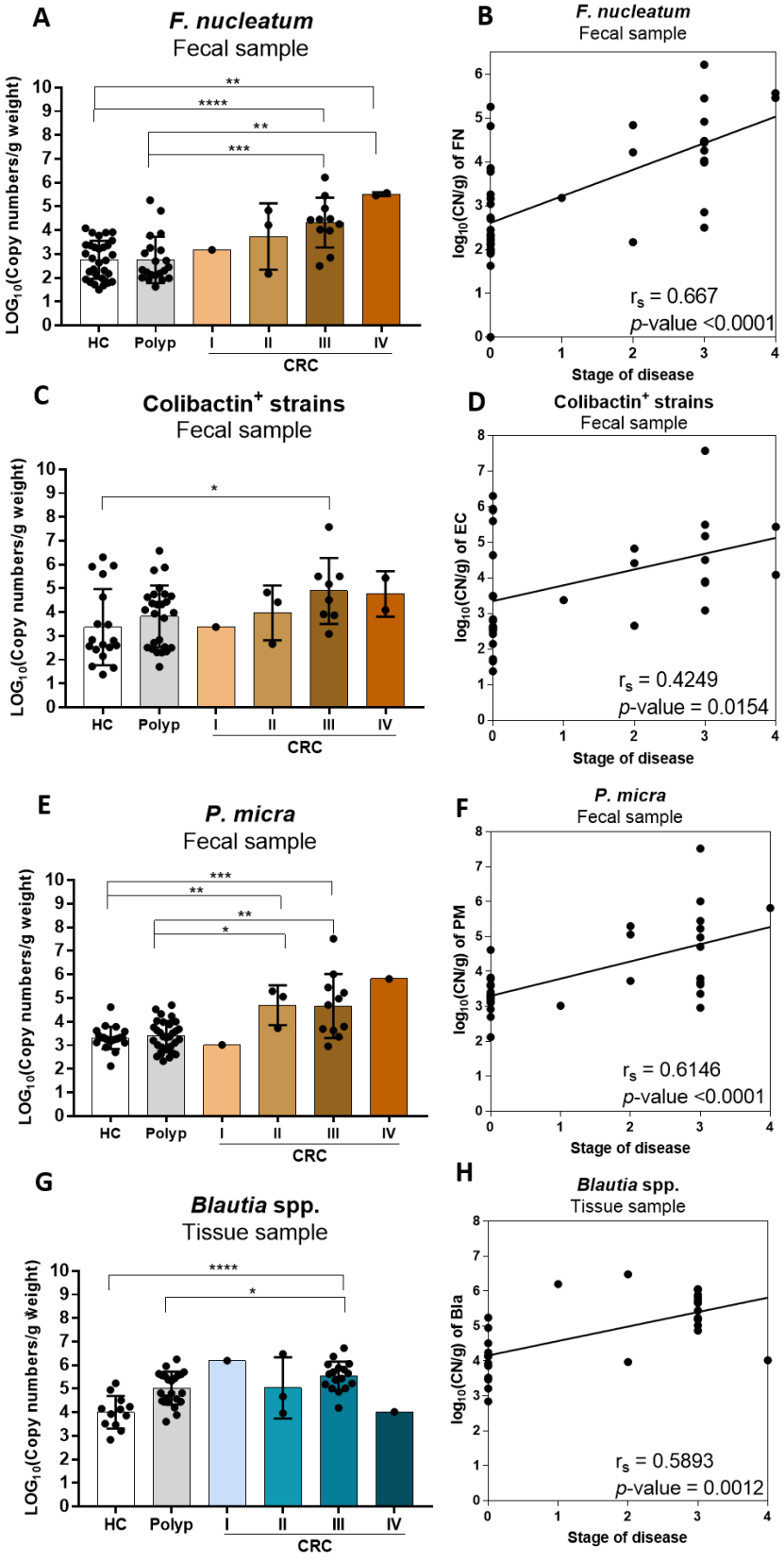

A
F. nucleatum
Fecal sample
LOG10(Copy numbers/g weight)
HC
Polyp
I
II
III
IV
CRC
B
F. nucleatum
Fecal sample
log10(CN/g) of FN
rs = 0.667
p-value <0.0001
Stage of disease
C
Colibactin+ strains
Fecal sample
LOG10(Copy numbers/g weight)
D
Colibactin+ strains
Fecal sample
log10(CN/g) of EC
rs = 0.4249
p-value = 0.0154
Stage of disease
E
P. micra
Fecal sample
LOG10(Copy numbers/g weight)
F
P. micra
Fecal sample
log10(CN/g) of PM
rs = 0.6146
p-value <0.0001
Stage of disease
G
Blautia spp.
Tissue sample
LOG10(Copy numbers/g weight)
H
Blautia spp.
Tissue sample
log10(CN/g) of Bla
rs = 0.5893
p-value = 0.0012
Stage of disease

Figure 20 Spearman's rank correlation coefficient between the absolute abundance of CRC-associated bacteria and stage of disease of HC versus CRC.
A-B) *F. nucleatum* in the fecal sample, C-D) colibactin positive bacteria in the fecal sample, E-F) *P. micra* in the fecal sample, G-H) *Blautia* spp. in the tissue sample. Abbreviations: *, $P<0.05$; **, $P<0.01$; ***, $P<0.001$; ****, $P<0.0001$.

5.9. The performance of single fecal bacterial candidates for CRC/adenoma detection

- **Comparison of healthy volunteers and CRC patients**

Binary logistic regression models were generated using the number of bacteria to differentiate between the HC group and CRC patients. Among all six bacteria, *F. nucleatum* (FN) showed the best performance in distinguishing patients with CRC from the HC group, giving an area under the ROC curve (AUC) of 0.86 (Figure 21A). At the optimal cut-off using the maximum Youden's index, the FN level at above 8.41×10^3 copy numbers per gram weight (CN/g) could detect CRC with sensitivity of 76.47% and specificity of 90.91%, (Figure 21B, Table 9). Secondly, the AUC for CRC detection was 0.84 for *P. micra* (PM) and 0.73 for *clbB*$^+$ bacteria (EC) (Figure 21A). At the best cut-off of 4.22×10^3 CN/g for PM, PM could detect CRC with sensitivity of 81.25% and specificity of 85.71%, and at the selected cut-off of 9.23×10^2 CN/g for EC, EC detected CRC with sensitivity of 92.86% and specificity of 61.11% (Figure 21B, Table 9). These results confirmed that three bacteria could be the potential bacterial biomarkers for discriminating CRC patients from the control group.

- **Comparison of healthy volunteers and adenoma patients**

The logistic regression model was also utilized to differentiate between patients with adenomas and control subjects. The bacterial markers showed poorer performance in adenomas detection than those in cancer detection. The greatest AUC for adenomas detection was 0.69 for *S. gallolyticus* (SG) level (Figure 21C). At the selected cut-off of 1.22×10^4 CN/g, SG could discriminate patients with polyp from

the control group with sensitivity of 70.59% and specificity of 77.78% (Figure 21D, Table 10). Next, fecal *F. saccharovirans* (FS) and fecal EC had an AUC of 0.61 and 0.59, respectively (Figure 21C). At the best cut-off value of FS (5.82×10^6 CN/g) and EC (4.22×10^3 CN/g) could discriminate adenoma patients from the HC group with sensitivity of 80.65% and 61.54%, specificity of 42.86% and 72.22%, respectively (Figure 21D, Table 10).

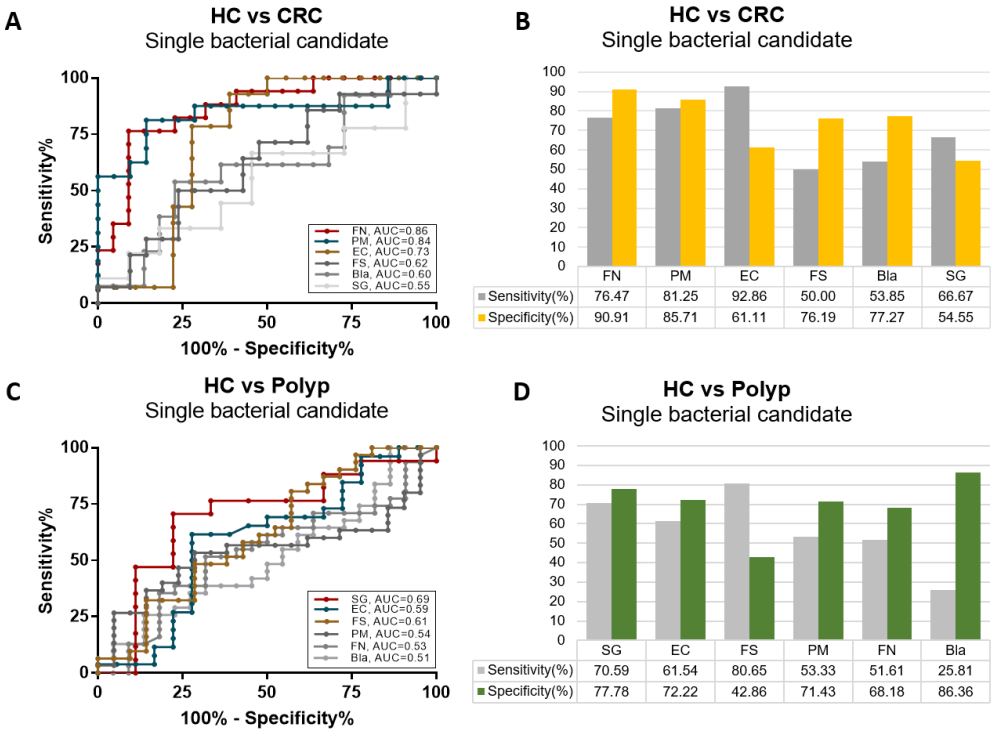


Figure 21 Receiving operating characteristic (ROC) curve displaying the sensitivity and the specificity of single fecal bacterial markers in distinguishing between CRC/adenoma patients and control groups.

A) ROC curve for CRC patients versus control subjects. B) sensitivity and specificity for CRC detection. C) ROC curve for patients with adenomas versus control subjects. D) sensitivity and specificity for adenomas detection. Abbreviations: FN, *F. nucleatum*; SG, *S. gallolyticus*; PM, *P. micra*; EC, colibactin positive strains; Bla, *Blautia* spp.; FS, *F. saccharivorans.*

Table 9 Microbial alteration in stools of patients using the cut-off value for CRC detection

		HC	CRC	Total	*P*-value
***F. nucleatum* (%)**		n=21	n=17	n=38	<0.001***
Cut-off = $8.41x10^3$ CN/g	Low	19 (90)	4 (24)	23 (61)	
	High	2 (10)	13 (76)	15 (39)	
***S. gallolyticus* (%)**		n=9	n=11	n=20	0.343
Cut-off = $7.59x10^4$ CN/g	Low	6 (67)	5 (45)	11 (55)	
	High	3 (33)	6 (55)	9 (45)	
Colibactin positive bacteria (%)		n=18	n=14	n=32	0.002**
Cut-off = $9.23x10^2$ CN/g	Low	11 (61)	1 (7)	12 (38)	
	High	7 (39)	13 (93)	20 (63)	
***P. micra* (%)**		n=21	n=16	n=37	0.000***
Cut-off = $4.22x10^3$ CN/g	Low	18 (86)	3 (19)	21 (57)	
	High	3 (14)	13 (81)	16 (43)	
***F. saccharivorans* (%)**		n=21	n=14	n=35	0.163
Cut-off = $9.44x10^6$ CN/g	Low	10 (48)	10 (71)	20 (48)	
	High	11 (52)	4 (29)	15 (52)	
***Blautia* spp. (%)**		n=22	n=13	n=35	0.061
Cut-off = $1.60x10^7$ CN/g	Low	5 (23)	7 (54)	12 (34)	
	High	17 (77)	6 (46)	23 (66)	

Abbreviations: CN/g, copy number per gram. *, *P*<0.05; **, *P*<0.01; ***, *P*<0.001; ****, *P*<0.0001. Note: The best cut-off values that maximized Youden's J statistic were used.

Table 10 Microbial alteration in stools of patients using the cut-off value for adenoma detection

		HC	Polyp	Total	*P*-value
***F. nucleatum* (%)**		n=21	n=30	n=51	0.158
Cut-off = $5.43x10^2$ CN/g	Low	14 (67)	14 (47)	28 (55%)	
	High	7 (33)	16 (53)	23 (45%)	
***S. gallolyticus* (%)**		n=9	n=17	n=26	0.019*
Cut-off = $1.22x10^4$ CN/g	Low	2 (22)	12 (71)	14 (54)	
	High	7 (78)	5 (29)	12 (46)	
Colibactin positive bacteria (%)		n=18	n=26	n=44	0.027*
Cut-off = $4.22x10^3$ CN/g	Low	13 (72)	10 (38)	23 (52)	
	High	5 (28)	16 (62)	21 (48)	
***P. micra* (%)**		n=21	n=30	n=51	0.079
Cut-off = $2.26x10^3$ CN/g	Low	15 (71)	14 (47)	29 (57)	
	High	6 (29)	16 (53)	22 (43)	
***F. saccharivorans* (%)**		n=21	n=31	n=52	0.066
Cut-off = $5.82x10^6$ CN/g	Low	9 (43)	6 (19)	15 (29)	
	High	12 (57)	25 (81)	37 (71)	
***Blautia* spp. (%)**		n=22	n=31	n=53	0.282
Cut-off = $4.62x10^7$ CN/g	Low	19 (86)	23 (74)	42 (79)	
	High	3 (14)	8 (26)	11 (21)	

Abbreviations: CN/g, copy number per gram. *, *P*<0.05; **, *P*<0.01; ***, *P*<0.001; ****, *P*<0.0001. Note: The best cut-off values that maximized Youden's J statistic were used.

5.10. The combination of fecal microbial markers and qualitative FIT to improve the screening efficacy for CRC detection

The fecal bacterial tests were selected from the top 3 bacterial candidates with the highest AUC values of single bacterium tests in CRC detection. The binary logistic regression models were performed to distinguish between healthy subjects and cancer patients. The combination of FN with PM improved diagnostic performance by increasing AUC to 0.90 as compared with other combinations (all 2-3 markers: AUC ≤0.86, Figure 22A), FN alone (0.86), or PM alone (0.84). At the best cut-off value, this combination could discriminate CRC patients from the HC group with sensitivity of 93.75% and specificity of 71.43% (Figure 22B).

Since qualitative FIT (FIT) is the most common non-invasive screening test for CRC, the addition of FIT was further analyzed whether it could improve the efficacy of fecal microbial markers. The FIT test was performed on the fecal samples of 16 CRC patients and 22 control subjects. The results showed that 94.22% (16/17) of stool samples from cancer cases were FIT positive. The addition of FIT could increase the AUC (0.93-0.97) of all combinations (Figure 22C, 22D) as compared with fecal microbial markers without FIT (0.81-0.90) (Figure 22A) and also enhanced both sensitivity (≥ 92.31%) and specificity (≥ 90.91%) (Figure 22E). Furthermore, the addition of FIT to the tests for PM alone or together with FN could discriminate the cancer group from the HC group with 93.75% sensitivity, 95.2% specificity, 93.8% positive predictive value (PPV), and 95.2% negative predictive value (NPV), while FIT alone gave 94.1% sensitivity, 90.9% specificity, 88.9% PPV, and 95.2% NPV (Figure 22E, Table 6). Additionally, the occurrence rate of PM together with FN was 94% (76%+18%) in CRC stools (Figure 22F). Altogether, these results suggested that the combination of fecal microbial markers and FIT increased the diagnostic performance for non-invasive CRC screening test.

In addition, the quantitative FIT (qFIT) was also performed to differentiate CRC from healthy controls. At a threshold of 50 ng Hb/ml, qFIT detected 16 out of the 17 CRC patients. The qFIT alone gave sensitivity of 94.1%, specificity of 95.5%, PPV of

94.1%, and NPV of 95.5% (Table 11). The quantitative FIT showed strong correlation between the amount of fecal human hemoglobin (hHb) and CRC stages (Spearman r= 0.715, Figure 23B). The level of hemoglobin was significantly elevated in CRC patients with stage II, III and IV when compared with healthy controls (p<0.001, p<0.0001, and p<0.01, respectively) and the adenoma group (p<0.001, p<0.0001, and p<0.01, respectively). However, there was no performance improvement of combining microbial markers with qFIT in CRC detection (Figure A4).

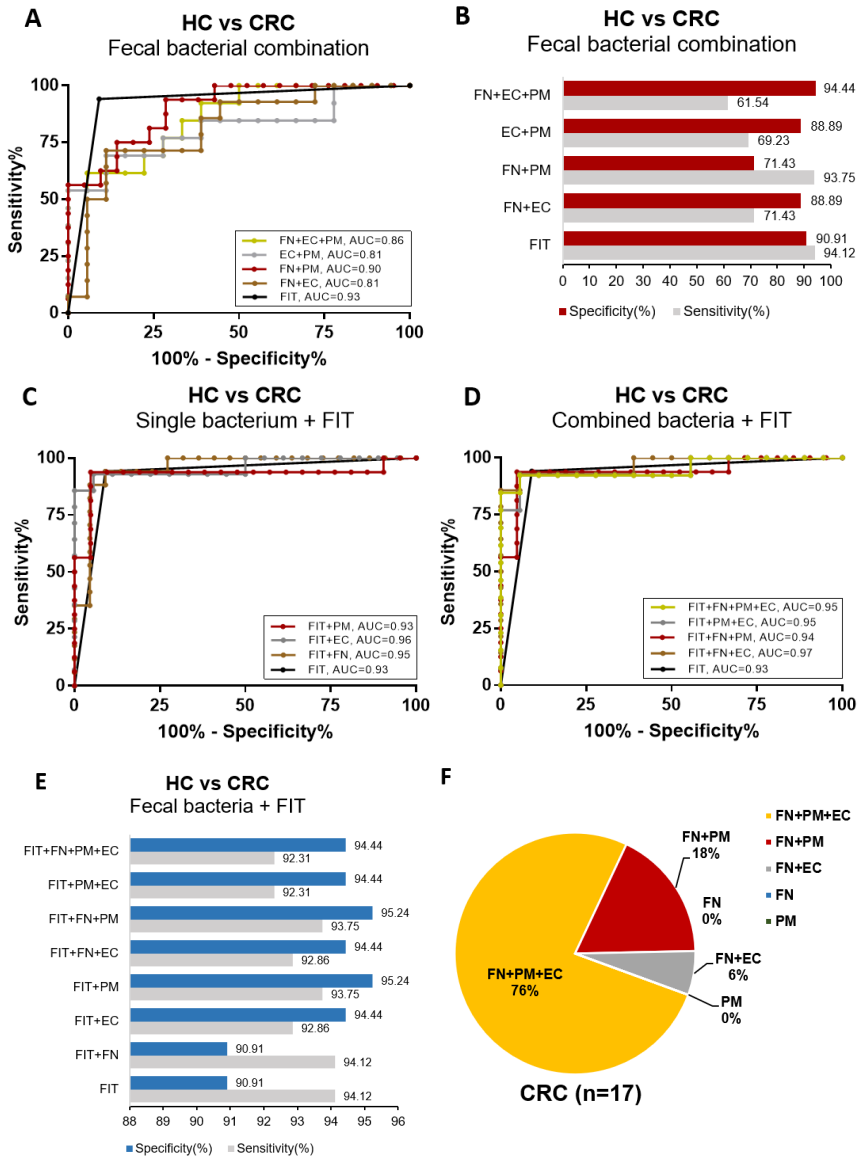


Figure 22 Receiving operating characteristic (ROC) curve displaying the sensitivity and the specificity for the combination of fecal bacterial markers and FIT in distinguishing CRC patients versus control groups.

A) ROC curve of fecal bacterial combination B) sensitivity and specificity of fecal bacterial combination C) ROC curve of single bacterium with FIT D) ROC curve of combined bacteria and FIT E) sensitivity and specificity of combined bacteria and FIT. F) the distribution of three bacterial markers in stool of CRC cases. Abbreviations: FN, F. nucleatum; SG, S. gallolyticus; PM, P. micra; EC, colibactin positive strains; Bla, Blautia spp.; FS, F. saccharivorans.

Table 11 Performance of FIT alone and in combination with selected fecal bacterial markers for CRC screening test

Variables	Qualitative FIT			Quantitative FIT
		+PM	+PM+FN	
AUC	0.93	0.93	0.94	0.984
Cut-off [a]	≥ 50 ng hHb/ml			≥ 50 ng hHb/ml
Sensitivity	94.1%	93.8%	93.8%	94.1%
Specificity	90.9%	95.2%	95.2%	95.5%
PPV	88.9%	93.8%	93.8%	94.1%
NPV	95.2%	95.2%	95.2%	95.5%
Accuracy	92.3%	94.6%	94.6%	94.9%

Abbreviations: FIT, fecal immunochemical test; PM, *P. micra*; FN, *F. nucleatum*; AUC, area under receiver operating characteristics curve; NPV, negative predictive value; PPV, positive predictive value. [a]the cut-off value of 50 ng hHb/ml was following by the manufacturer's instruction.

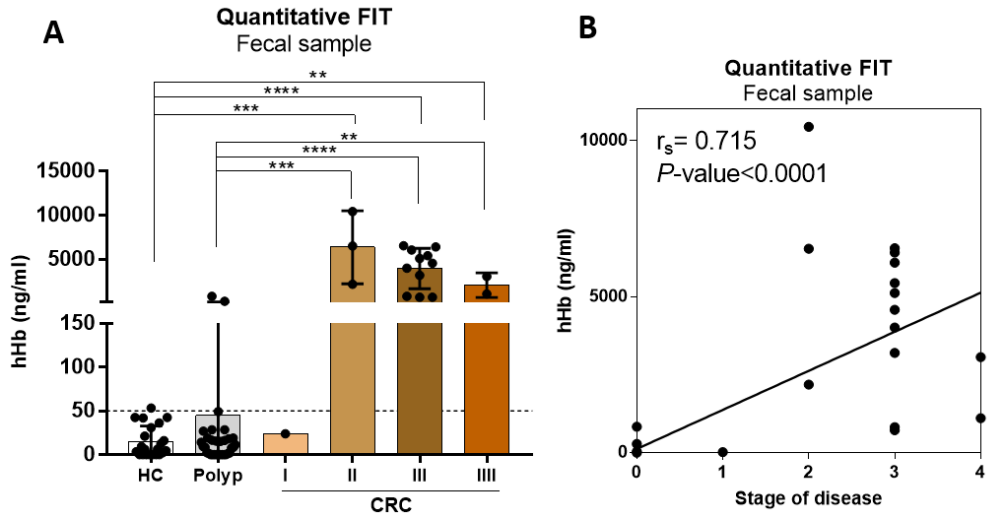


Figure 23 Spearman's rank correlation coefficient between the human hemoglobin (hHb) amount via quantitative FIT and stage of disease of HC versus CRC.

A) the hHb amounts according to stage of disease, B) the correlation between the hHb amount and stage of disease. Abbreviation: hHb, human hemoglobin; FIT, fecal immunochemical test; *, *P*<0.05; **, *P*<0.01; ***, *P*<0.001; ****, *P*<0.0001. Note: each test was considered positive at threshold of 50 hHb ng/ml buffer.

5.11. The combination of fecal microbial markers and FIT to improve the screening efficacy for adenoma detection

The ability of the fecal bacteria was analyzed to enhance differentiation between adenoma patients and control subjects. Because of poor differentiation performance of single bacterium assay, combined bacteria might improve the efficacy of detection. The combinations of three bacteria; SG, FS, and EC, that gave the highest AUC in the single bacterium assay were conducted. The combination of all three bacteria showed the best performance, i.e., AUC of 0.97, in discriminating patients with adenoma from control subjects (Figure 24A). At the best cut-off, the combination assay of three bacteria could detect adenomas with sensitivity of 100%, specificity of 83.33%, PPV of 94.1%, and NPV of 100% (Figure 24B, Table 12). On the other hand, although FIT is currently used for CRC screening in the clinic, the poor performance of FIT in adenoma detection gave a lower AUC (0.53) compared with

the bacterial combination assay (all AUC ≥ 0.65) (Figure 24A). At the selected cut-off value of FIT provided sensitivity of 96.77%, specificity of 9.09%, PPV of 60%, and NPV of 66.7% (Figure 24B, Table 12). These results suggested the potential of the combination assay of three bacteria without FIT to differentiate adenoma patients from healthy controls. However, there was a limitation because all of SG, FS and EC were simultaneously found in only 52% of fecal samples from adenoma patients (Figure 24F).

To improve the adenomatous polyp screening test, the complementary fecal bacterial models with FIT was performed. The combination of single bacterium detection and FIT slightly increased AUC of FS with FIT (0.65) and EC with FIT (0.62) (Figure 24C), but this type of combination did not enhance sensitivity and specificity as compared with the bacterial detection alone (Figure 21D and 24E). Besides, the addition of FIT to other combination assays of two or all three bacteria did not provide better discrimination between adenomas and healthy controls (Figure 24D, 24E) when compared with the combination assay of bacteria alone (Figure 24A). These results demonstrated that the absolute quantification of fecal microbial markers without FIT was more sensitive and specific than FIT.

Owing to the low prevalence of SG, FS, and EC in fecal samples of adenoma patients, the combinations of other bacteria with FIT were analyzed to find alternative fecal bacterial markers with higher prevalence in patients with polyp. The combinations of five bacteria with the FIT (five bacteria: FN+PM+EC+FS+Bla) gave the AUC of 0.735 (Figure 25A). At the best cut-off value, this model of five bacteria combined with FIT could discriminate adenomas from controls with sensitivity of 83.3%, specificity of 64.7%, PPV of 76.9%, and NPV of 73.3% (Figure 25B, Table 12). Although the combination of the five bacteria and FIT showed no better performance in adenomas detection than the previous combination of three bacteria, these five bacteria were found more often (81%) in stools of the polyp group (Figure 25C).

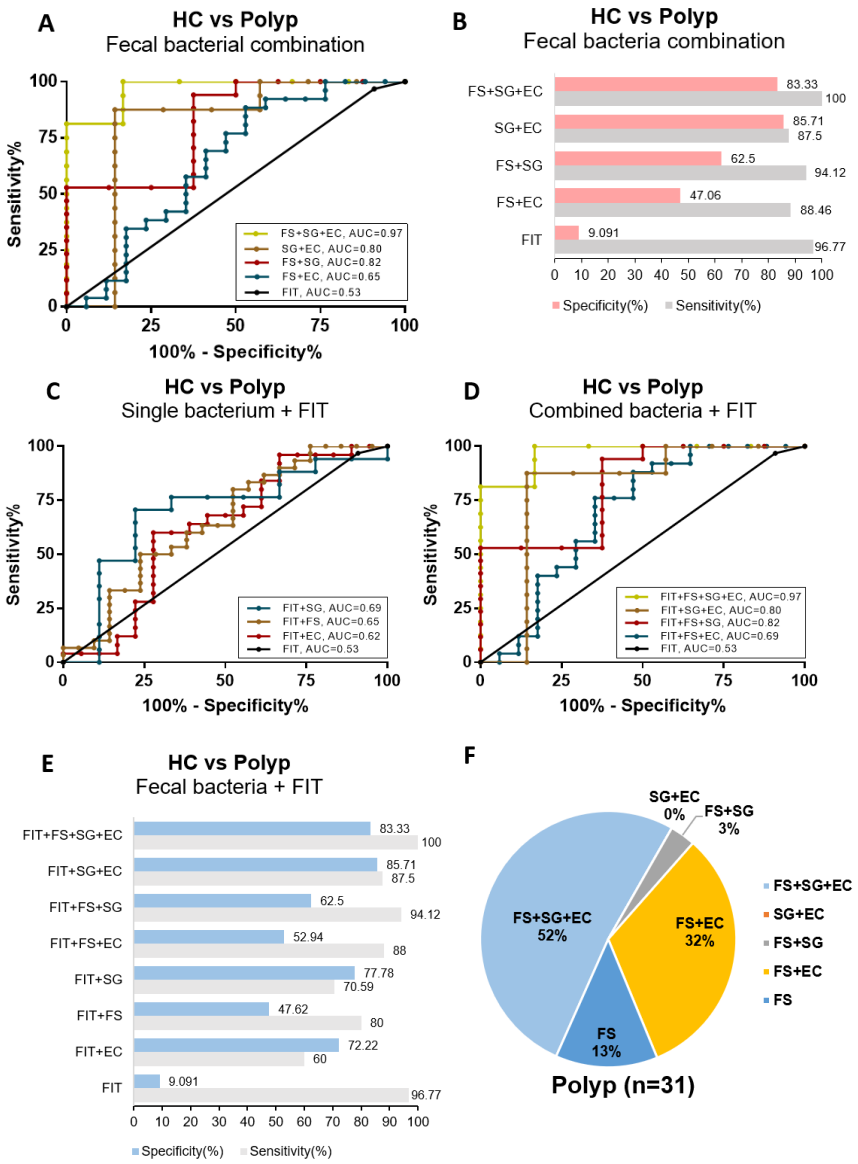


Figure 24 Receiving operating characteristic (ROC) curve displaying the sensitivity and the specificity for the combination of fecal bacterial markers and FIT in distinguishing patients with polyp versus control groups.

A) ROC curve of fecal bacteria combination B) sensitivity and specificity of fecal bacterial combination C) ROC curve of a single bacterium with FIT D) ROC curve of combined bacteria and FIT E) sensitivity and specificity of combined bacteria and FIT F) the distribution of specific bacterial markers in feces of adenoma patients.

Abbreviations: FN, *F. nucleatum*; SG, *S. gallolyticus*; PM, *P. micra*; EC, colibactin positive bacteria; Bla, *Blautia* spp.; FS, *F. saccharivorans.*

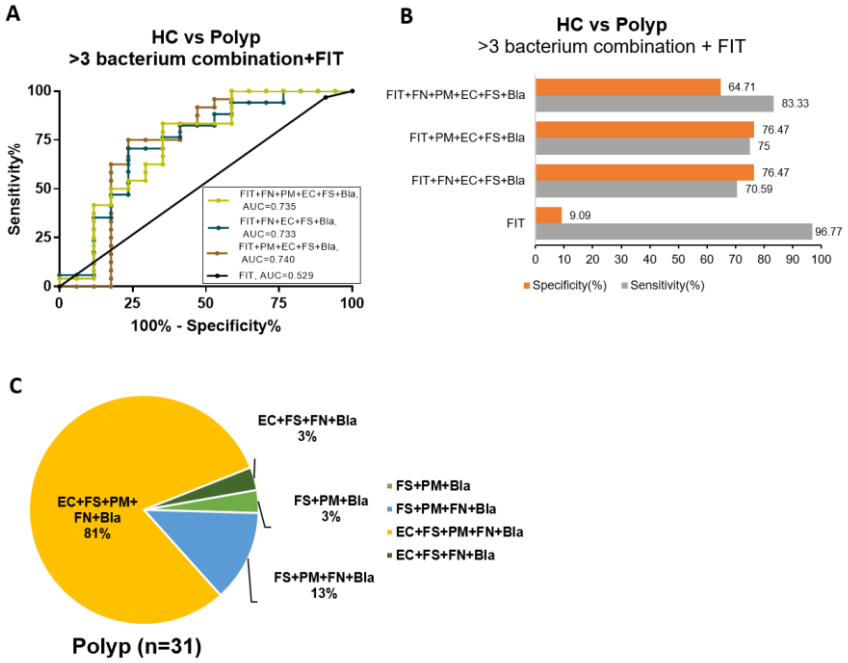


Figure 25 Receiving operating characteristic (ROC) curve displaying the sensitivity and the specificity for the alternative combination of fecal bacterial markers and FIT in distinguishing patients with polyp versus control groups.

A) ROC curve of fecal bacterial combination B) sensitivity and specificity of fecal bacterial combination C) the distribution of specific bacterial markers in feces of adenoma patients. Abbreviations: FN, *F. nucleatum*; SG, *S. gallolyticus*; PM, *P. micra*; EC, colibactin positive bacteria; Bla, *Blautia* spp.; FS, *F. saccharivorans.*

Table 12 Performance of FIT alone and in combination with selected fecal bacterial markers for adenoma screening test

Variables	Qualitative FIT			FS+SG +EC	Quantitative FIT
		+PM+EC+ FS+Bla	+FN+PM+EC+ FS+Bla		
AUC	0.53	0.74	0.735	0.97	0.5007
Cut-off[a]					< 29.75 ng hHb/ml
Sensitivity	96.8%	75.0%	83.3%	100.0 %	90.6%
Specificity	9.1%	76.5%	64.7%	83.3%	27.3%
PPV	60.0%	81.8%	76.9%	94.1%	64.4%
NPV	66.7%	68.4%	73.3%	100.0 %	15.8%
Accuracy	60.4%	75.6%	75.6%	95.5%	64.8%

Abbreviations: FIT, fecal immunochemical test; PM, *P. micra*; EC, colibactin positive bacteria; FS, *F. saccharivorans*; Bla, *Blautia* spp.; FN, *F. nucleatum*; SG, *S. gallolyticus*; AUC, area under receiver operating characteristics curve; NPV, negative predictive value; PPV, positive predictive value.

[a] The optimal cut-off value of 29.75 ng hHb/ml was calculated from Youden's index.

CHAPTER VI

DISCUSSION

Colorectal cancer (CRC) is one of the most common cancers (2) and becomes a major public health problem worldwide including in Thailand. The risk factors of CRC are the complex interplay among genetics, dietary, lifestyle, and environmental factors (90). In the last decade, accumulating evidence has supported a hypothesis that the alteration of intestinal microbiota composition possibly affects the initiation and progression of CRC (91-93). The high-throughput sequencing approach including the 16S rRNA gene sequencing has been extensively used as an efficient method to examine the total bacterial component in a particular environment (94). Although mucosa-associated microbiota were speculated to directly interact with the host, a limited number of studies performed gut microbiome analysis using colon tissue samples of CRC patients (52, 53, 95, 96). Matched non-tumor tissue from CRC patients was usually used for comparison (10, 85). In addition, studies are rarely conducted in both stool and mucosal samples to build a comprehensive picture (14, 49). In this study, the 16S rRNA gene sequencing tool was utilized to compare the bacterial composition among three groups of the Thai population consisting of healthy control (HC), patients with adenomas (polyp), and patients with CRC. Since the design of this study was a cross-sectional study, patients with adenomas were included to explore the gut microbiome across the intermediate state between HC group and CRC patients to better understand the association of gut microbiota and CRC according to the adenoma-adenocarcinoma sequence (3). Moreover, the sample types including feces and mucosal tissue representing lumen- and mucosa-associated microbiota, respectively, were investigated in this study. As expected, the structural segregation of gut microbiota between mucosal tissue and fecal samples showed significant differences in terms of bacterial richness, diversity, and overall microbial profile (Figure 6 and 7). This finding agrees with previous studies that lumen-associated microbiota only partially correlated with mucosa-associated microbiota (14, 97, 98). The relative abundance of the phyla Firmicutes was more dominant in

stool than mucosal tissue (Figure 7). Two previous studies reported that approximately 90% of fecal microbiota was predominated by Bacteroidetes and Firmicutes (41, 42). The phylum Firmicutes, which has been reported to increase energy harvest from host's diet (99), was greatly elevated in the intestinal lumen (97). By contrast, the higher proportions of phyla Proteobacteria and Fusobacteria were highly enriched on mucosal tissue (Figure 7) which are consistent with previous report (97). Therefore, exclusive use of fecal samples may not truly represent the microenvironment on the mucus layer (98).

The intestinal microbiota is spatially stratified throughout the colon in terms of longitudinal and cross-sectional axes. The different distributions of physical and chemical features (pH and oxygen concentration), as well as nutrient gradients, along the large intestines, can affect the microbial composition in distinct colon locations (100). For these reasons, this study collected the tissue samples from both proximal and distal colon of the same healthy donor to determine the bacterial pattern in the HC group. The alpha and beta diversity results showed no significant difference between the proximal and distal colon of the HC group (Figure A7). Thus, both tissues were used in further analyses as controls.

According to the results of biopsy tissue, the mucosal microbiota composition in samples originating from the CRC patients was considerably different from those of HC individuals (Figure 8D and A1), whereas the results of adenoma group and others were not statistically different. These finding were in accordance with a study of Irish CRC patients (14), which showed different trend in CRC patients from that of controls. Individuals with CRC had higher relative abundance of genera *Bacteroides*, *Parabacteroides*, *Collinsella*, *Erysipelatoclostridium*, *Flavonifractor* than those in HC groups (Figure 9, 10) whereas *Escherichia-Shigella*, and *Faecalibacterium* were lower abundant in CRC cases (Figure 9, 10). Moreover, members in the phylum Firmicutes, the major producers of beneficial short chain fatty acid (SCFA) (101), showed dissimilar distribution in which the genus *Flavonifractor* was elevated in the patients with CRC while the genus *Faecalibacterium* was more abundant in the HC group.

These results indicated that the microorganisms from the same taxonomic clade could play distinct functional roles in the microenvironment depending on their virulence factors and the interaction with their surrounding (46).

Tjalsma and his team proposed a bacterial driver-passenger model for CRC development in 2012 (102), in which certain driver bacteria in the colon can initiate in multistep development to colorectal carcinogenesis consisting of induced inflammation, increased cell proliferation, and/or produced genotoxin. After epithelial DNA damage, colorectal tumorigenesis is caused by alteration of the gut microenvironment that facilitates the outnumber of colonic commensals with either tumor-promoting or tumor-suppressive features considered as passenger bacteria. Furthermore, bacterial drivers may be replaced by passenger bacteria that could take the growth benefit in the cancerous microenvironment. In this study, *Flavobacterium*, an opportunistic pathogen in immunocompromised patients (103), was significantly over-represented in the mucosal tissue of the adenoma group but not in CRC (Figure 10), which is consistent with a previous study in patients with intestinal metaplasia (104). However, the bacteria were not in high abundance in fecal samples of the adenoma patients (Figure 14B), indicating that this genus might be associated with the early stage of CRC and act as driver bacteria but outcompeted by passenger bacteria in CRC patients.

As for passenger bacteria, the interesting and consistent observation was the significant enrichment of *Flavonifractor plautii* on the mucosal tissue of CRC patients (Figure 11D). It has been identified in Indian CRC population (105). *F. plautii* is able to degrade beneficial flavonoids found in plant-based diet, i.e. green tea, wine and cocoa (106) and then generates polyphenolic compounds, which play a role in the prevention of cancer (107). Apart from *F. plautii*, *Parabacteroides distasonis* in tissue samples (Figure 11F) was also associated with CRC. The protective role of this bacterium in tumor development and the maintenance of gut barrier has been proposed in tumor-bearing mice (108). Taken together, these findings suggested that

these bacteria play potential roles of passenger bacteria with tumor-suppressive features.

On the other hand, other remarkable observation of mucosal microbiota including *Erysipelatoclostridium ramosum*, *Bacteriodes thetaiotaomicron*, and *P. merdae* were positively associated with CRC (Figure 11). These bacteria could be found in the human gastrointestinal tract, but no study surveyed on their characteristics in CRC. However, they could be opportunistic pathogens in immunocompromised hosts (109) or found higher level in hypertension (110). In addition, *E. ranosum*, which has human immunoglobin A protease function (111), was less abundant in the HC and adenoma groups while dramatically elevated in the cancer tissues (Figure 11A). This finding might be explained by their IgA protease translocating across the colonic mucosa that could increase the host susceptibility (111). These results suggest they might serve as passenger bacteria with tumor-promoting roles that poorly colonize in non-disease colon but preferentially colonize in the tumor microenvironment. Moreover, the total microbial alterations found in this study and putative functions reported in previous studies are summarized in Table 13.

However, *Escherichia-Shigella*, which were more abundant in the HC tissues (Figure 10A, 11G), are generally referred as normal flora with potential pathogenic aspect in large intestine (112), in disagreement with previous findings (46). Certain *E. coli* strains can produce genotoxin (113) and usually are considered in causing CRC like driver passenger (102). The discrepancy between our study and other studies might be explained by 1) ethnic difference in the susceptibility to colonization by *Escherichia* and *Shigella,* 2) the different virulence mechanisms and functional roles among strains, and 3) the possibility as a high risk group for CRC of currently healthy subjects used as HC in this study.

Interestingly, 7 bacterial species from mucosal tissues and 3 bacterial species from stool samples were identified as putative microbial biomarkers to discriminate the microbial structure of CRC patients and healthy individuals (Figure 16A).

Table 13 Overall enrichments of lumen- and mucosa- associated microbiota in each group identified in the study.

Phylum	Family	Genus	Species	Sample	Purported functions/ previous reports
Actinobacteria	*Coriobacteriaceae*	*Collinsella*	-	Tissue	High abundance in T2DM (114), atherosclerosis (115, 116)
	Eggerthellaceae	*Eggerthella*	*E. lenta*	Feces	Bacteremia in immunocompromised host (117, 118), catechin converting property (119)
Firmicutes	*Erysipelotrichaceae*	*Erysipelato-clostridium*	*E. ramosum*	Tissue Feces	Invasive infections in immunocompromised hosts (109), producing IgA proteinase (111)
		-	-	Feces	High immunogenicity (120), role in the inflammation-related disorders of gut (121) and metabolic disorders (122)
	Ruminococcaceae	*Faecalibacterium*	-	Tissue Feces	Butyrate producing property and anti-inflammation features (123, 124)
		Oscillibacter	-	Feces	Negative correlation with gut barrier function (125), producing valeric acid feature (126), the association with depression (127)
		Flavonifractor	*F. plautii*	Tissue	Alleviating immune response in mice (128), the association in Indian CRC patients (105), catechin converting property (119)

Phylum	Family	Genus	Species	Sample	Purported functions/ previous reports
	Lachnospiraceae	*Agathobacter*	-	Feces	Positive correlation with fiber intake (129)
		Fusicatenibacter		Feces	SCFA producing features (130) , decreased in UC cases (131)
		-	-	Tissue	SCFA producing features (132)
Proteobacteria	*Enterobacteriaceae*	*Escherichia_ Shigella*	*Escherichia_ Shigella_coli*	Tissue	Genotoxin producing features (113)
Bacteroidetes	*Bacteroidaceae*	*Bacteroides*	*B. thetaio-taomicron*	Tissue	Promoting enteric infection of EHEC (133), breaking down indigestible dietary plant polysaccharides (134)
			B. vulgatus	Feces	Reducing gut LPS in atherosclerosis-prone mice (135)
	Tannerellaceae	*Parabacteroides*	*P. merdae*	Tissue	High level in hypertension patients (110)
			P. distasonis	Tissue	Anti-inflammation and anti-tumor activities (108)
	Odoribacteraceae	*Butyricimonas*	-	Tissue Feces	Butyrate producing property (136)
	Flavobacteriaceae	*Flavobacterium*	-	Tissue	Opportunistic pathogens in immunocompromised host (103)

Note: green box, enriched in CRC patients; yellow box, enriched in HC group; blue box, enriched in adenoma group.

Abbreviations: NADH, nonalcoholic steatohepatitis; T2DM, type 2 diabetes; EHEC, enterohemorrhagic *Escherichia coli.*; UC, ulcerative colitis; LPS, lipopolysaccharide; Ig, immunoglobulin

Nevertheless, the utilization of tissue samples in CRC prediction is the invasive approach. Thus, fecal samples are more practical to apply the biomarkers. *E. ramosum* was solely found significant abundance in both fecal and tissue samples (Figure 16B). This study provided a population-specific biomarker, which may potentially be utilized in noninvasive screening of Thai CRC.

Confounding factors such as dietary (137) or medical treatment (138, 139) may directly affect gut flora. In the present study, no subjects were taking antimicrobial drugs, consuming probiotics products, and receiving chemotherapy or radiation. Moreover, sampling fecal samples after colonoscopy within at least 1 month was also avoided as shown by the evidence of Drago et al. (140) that polyethylene glycol bowel cleansing preparation could have an impact on reverting to resemble the baseline intestinal microbiota profile. Therefore, these confounders could not considerably alter the gut microbiota composition. Nonetheless, the CRC group had significantly more patients with diabetes mellitus (DM) than the HC group, which had no DM cases (Table 6). The DM factor could bias the CRC-associated microbiota result. However, exclusion of DM cases roughly did not alter the main results (Figure A5). As a multifactorial disease, it was difficult to exclude all co-founding factors of CRC, therefore the study still included the results from cases with DM to perform the microbiome analysis.

This microbiome study has some limitations that might explain different results compared with other studies. As for disparate protocols, the selection of universal 16S rRNA gene primers is one of the factors that can cause different results of gut microbiota profiling between studies (12, 141). Most microbiome studies used the primers targeting on V3 (41, 46), V4 (35, 42, 49, 51), or V3/V4 (14, 37, 44) of hypervariable regions of 16S rRNA gene, but all of these primers resulted in low sensitivity of tissue samples in this study (Figure A6). Therefore, this study performed the 16S rRNA gene sequencing using primers targeted V1/V2, in consistent with some previous studies (52, 142, 143). Moreover, the distinct choices of specimen handling procedures including different manufacturers of sample preservation solution can

cause dissimilar patterns of gut microbiota (141, 144). Consequently, the study utilized the preservation solution in order to minimize the microbial community change over time and prevent DNA degradation by temperature fluctuation. Additionally, the samples were kept on ice during transportation by volunteers and collected at -80 °C for long term storage. In addition, the samples in each group were limited in number and restricted to one hospital in Bangkok that might not generally represent the Thai population.

The second objective of this study was to validate the microbiome analysis results of six CRC-associated gut microbiota that were well-known in recent studies by quantitative PCR (qPCR) assay. To investigate the alteration of bacterial composition along adenomas-adenocarcinomas sequence, qPCR was used to compare the bacterial quantity in both stool and tissue samples obtained from different sites of colon of patients with CRC, patients with adenomas, and HC. This study showed that CRC-related bacteria of patients with CRC were distinct from those of non-CRC volunteers (Figure 19). This is consistent with previous studies (50, 78, 145-148). Furthermore, a marked increase in the absolute abundance of *Fusobacterium nucleatum* in both feces and mucosal tissues from CRC patients was observed (Figure 19A). *F. nucleatum* was previously shown to be related to colorectal malignancy (85, 146, 149, 150), promote a pro-inflammatory environment (151), and possess virulence factors that promote their adhesiveness to host epithelial cells (55, 152, 153) and the ability to invade into epithelial cells (55, 154). Consequently, *F. nucleatum* may drive colorectal carcinogenesis. *Parvimonas micra* was also significantly higher in stool and mucosal samples of cancer cases (Figure 17C). which is in agreement with previous studies (14, 145, 146, 155). Therefore, *P. micra* is another potential CRC-promoting microorganism. Like *F. nucleatum*, it is an obligate anaerobic bacterium that could be an oral pathogen (156). Little is known about the participation of *P. micra* in colonic tumorigenesis. A study of Marchesan and collaborators (157) found *P. micra* was able to interrupt the regular function of the nucleotide-binding oligomerization domain 2 (NOD2) stimulatory activity in

periodontitis. This might perhaps give rise to a pro-tumorigenic and inflammatory environment. In addition, *Blautia* spp. significantly elevated in mucosal tissues of CRC patients in this study (Figure 19D). *Blautia* is an anaerobic bacteria that present widely in the mammal feces and colon (158). On the contrary, some studies reported the presence of bacteria belonging to the family *Lachnospiraceae* like *Blautia* spp. significantly diminished or vanished in fecal samples (159) and tissue samples (97) of patients with CRC. *Blautia* has been shown to have potential probiotic properties including producing short chain fatty acid (SCFA) (158). It was also involved in alleviating metabolic diseases (160) and inflammatory diseases (161), and inhibiting specific pathogens (162). Nevertheless, there is a contradiction in the correlation of *Blautia* with human diseases. Less *Blautia* was found in patients with type 2 diabetes (160), obesity (163), colon cancer (97, 159), and Crohn's disease (161), whereas more *Blautia* was present in irritable bowel syndrome (IBS) (164), inflammatory bowel diseases (165), breast cancer (166). Besides, there is no more-in-depth study of this genus and only a few strains of this bacterium have been isolated and characterized (158). The discrepancy of these studies might be explained by 1) ethnic difference in the susceptibility to harbor by *Blautia* or 2) geographical differences in *Blautia* strains including those discovered in Thailand. In contrast to the previously mentioned bacteria, *Fusicatenibacter saccharivorans* decreased significantly in stool samples of cancer cases compared with non-cancer cases (Figure 19), which was consistent with previous studies (159, 167). *F. saccharivorans* is the only strain in the genus *Fusicatenibacter* of the family *Lachnospiraceae* and was successfully isolated and cultured in 2013 (168). *F. saccharivorans* also has the SCFA-producing properties (168). A study of Takeshita et al. (169) reported that *F. saccharivorans* could suppress intestinal inflammation via inducing interleukin-10 in murine colitis and ulcerative colitis (UC) patients. Nonetheless, the role of *F. saccharivorans* in the pathogenesis of CRC remains unknown.

In this study, bacterial levels of *F. nucleatum,* colibactin^{+} *strains*, and *P. micra* in fecal samples showed positive correlation with clinicopathological features.

Moreover, *Blautia spp*. in tissue samples was significantly correlated with the stage of disease especially late-stage (stage III) of colorectal cancer (Figure 20). This information indicates that patients with advanced cancers may be more susceptible to the colonization of certain bacteria. The abundant nutrients, low oxygen, and immune suppression were found within the hypoxic tumor at late stages that supports bacterial growth (170). Another perspective following the concept of alpha-bug theory, there were primary bacteria could remodel the whole microenvironment to facilitate colonization via other bacteria and drive pro-inflammatory immune responses leading to oncogenic development of colonic epithelial cells (171). A previous study of de Carvalho et al. (172) suggested positive association between a high level of *F. nucleatum* in tumor tissue and high-depth invasion of cancer and poor prognosis. Likewise, Bonnet et al. (173) observed a relationship between cyclomodulin positive strains in the CRC mucosa and stage III/IV colon cancer. Consistent with Xu et al. study (174), fecal *P. micra* was predominantly enriched in the early and late stages of CRC. Although the association between *Blautia* spp. and TNM staging of CRC has never been reported, a study in breast cancer showed fecal *Blautia* spp. significantly escalated in stage II and III compared with stage 0 and I of breast cancer (166). The relationship between specific bacteria and clinical stage of disease might provide promising information to enhance the prediction of cancer progression and patient outcome. Currently, the qPCR technique for detection is a reasonably priced and widely used method in clinical laboratories. In case of stronger evidence, the quantity of bacteria in feces or colon tissue samples might be applied for diagnosis and progression of CRC in the future. However, the association of CRC-related bacteria mentioned above did not necessarily imply their oncogenicity or CRC etiology because many bacteria found in this study are present in high percentage of all groups including HC. In addition, CRC is a multifactorial disease that may be accompanied by other risk factors including genetics, dietary habits, and environment factor (90).

Based on the Asia Pacific Consensus Recommendations on colorectal cancer screening in 2015 (4), the fecal immunochemical tests (FITs) including qualitative FIT and quantitative FIT are widely used as stool-based screening test for colorectal cancer to select high-risk subjects for colonoscopy. The quantitative FIT is operated by an automated system which could reduce observer variability and is suitable for high-throughput screening in medical laboratories (175). On the other hand, the qualitative FIT is simpler in terms of no requirement of the additional medical device or automated analyzer, therefore it is more practical for office-based testing (176). The quantitative FIT could detect CRC with high sensitivity and high specificity (51). Nevertheless, Baxter et al. (51) found quantitative FIT missed detection of a subset of colonic lesions that could be detected by the quantitative FIT complemented with microbiome-based markers. Therefore, this study determined the performance of the microbiota abundance in distinguishing individuals with adenocarcinomas or adenomas from healthy colons and further evaluated by binary logistic regression models. For cancer detection (Figure 22, Table 11), the combination of qualitative FIT and detection of *P. micra* alone or together with *F. nucleatum* had higher specificity (95.2%) and higher positive predictive values (PPV) (93.8%) than qualitative FIT alone (90.9% specificity and 88.9% PPV) but had a slight decline in sensitivity (93.8% vs 94.1%). This information suggested fecal *P. micra* and *F. nucleatum* as the putative bacterial biomarkers for CRC, which are consistent with previous studies (78, 146-148). In addition, they could complement the accuracy of the existing screening test, qualitative FIT. On the contrary, the combination of fecal bacteria and quantitative FIT could not improve the CRC detection (Figure A4). The quantitative FIT alone had higher sensitivity (94.1%) and higher specificity (95.5%) in detecting adenocarcinoma than the combined test (Table 11). This is inconsistent with previous studies published in 2016 (51, 78). This discrepancy could be the result of improving performance of current FITs by manufacturers compared to those use in previous studies. Despite the superior performance of quantitative FIT in CRC detection, the

combined test with qualitative FIT can be an alternative option in resource-limited settings where automated machines are not available.

As for adenoma detection, the FIT was designed to detect occult fecal hemoglobin so its performance in detecting nonbleeding colonic lesions was a challenge. In this study, both qualitative (Figure 20, 21, Table 12) and quantitative FIT (Table 12, Figure A4) showed poor detection of precancerous lesions as demonstrated by the suboptimal area under the curve (AUC) of 0.53 and 0.50, respectively, which is in the agreement with previous studies (51, 78, 146). In addition, the combination of *S. gallolyticus*, *F. saccharivorans*, and colibactin positive strains without FIT could detect adenomas with sensitivity of 100% and specificity of 83.3% (Figure 20E, Table 12). Even though the detection of these three bacteria to differentiate adenomas patients from healthy controls was efficient, the occurrence rate that found all three bacteria simultaneously was limited since fecal *S. gallolyticus* was found 54% of study volunteers (Figure 20F). Similarly, Dumke et al. reported 63% of *S. gallolyticus* fecal carriage in healthy volunteers (177). With the exception of *S. gallolyticus*, an alternative test that combined other five bacteria composing of *F. saccharivorans*, colibactin positive strains, *P. micra*, *F. nucleatum*, and *Blautia* spp. with qualitative FIT could detect adenomatous polyp with sensitivity of 83.3% and specificity of 64.7% (Figure 21B, Table 12). Even if the performance of this bacterial combination test did not superior to the test with *S. gallolyticus*, the co-occurrence of these five bacteria in feces was higher (81%) (Figure 21C) than the combined bacteria with *S. Gallolyticus* (52%) (Figure 20F). These findings supported the potential role of microbiota in the fecal samples as novel candidate biomarkers for detection of precancerous lesions.

It is noteworthy that colonoscopy was implemented as the preferred strategy for CRC screening by the American College of Gastroenterology (ACG) (5). Nevertheless, the accessibility to colonoscopy is limited partly due to requiring high-cost equipment, well-trained endoscopic nurses, and board-certified endoscopists in Thailand (178). According to the Asia Pacific Consensus Recommendations for CRC

screening, a two-step method for CRC screening in average-risk subjects is recommended in the resource-limited countries (4, 179). The FIT-based screening is used as the first step followed by the second step of colonoscopy in case of positive FIT (4, 179). Consequently, the major advantage of the bacterial markers is the ability to detect even non-bleeding polyps in asymptomatic patients. Moreover, stool-based tests such as FIT combined with bacterial markers are non-invasive and may be more affordable than colonoscopy. Therefore, this method can be used generally for early detection of colonic lesions in primary healthcare.

This study showed that the combined detection of CRC-related bacteria with/without qualitative FIT offer acceptable performance in CRC or adenomatous polyp screening. The disadvantage of bacterial markers is limited prevalence of these bacteria in patient's stool, and it may be dependent on the population's characteristics. Thus, the development of these microbial models needs further investigation to improve their accuracy. The complementary approaches using different bacterial markers and FIT for adenocarcinoma or adenoma screening should be validated in larger population.

In conclusion, this study identified the dysbiosis signature of gut microbiota according to the adenoma-adenocarcinoma sequence for the Thai population, especially the significant difference of mucosal microbiota between CRC patients and the HC group. Apart from the host factors, the study suggested that the imbalance of colonic microorganisms accompanied by driver and passenger bacteria might be involved in CRC tumorigenesis. The microbiome analysis in this study also uncovered the putative biomarkers for CRC from mucosal and luminal microbiota. The subsequent validation study of candidate microbiota in the same population revealed the significant difference of most bacterial quantity among three groups. These findings remained to be confirmed by a larger population to improve the statistical power and identify the role of different microbiota. Furthermore, the functional profiles of gut microbiota that interact with human host or studies in animal models are needed to illustrate the causal role in CRC carcinogenesis. In

addition, the candidate biomarkers in CRC or adenoma detection also require the validation step by larger sample size.

APPENDIX A

MATERIALS

BUFFER AND REAGENT

1. Luria-Bertani broth (LB broth)

   To 950 ml of deionized H_2O, add:

Tryptone	10 g/L
Yeast extract	5 g/L
NaCl	10 g/L

   Sterile by autoclaving at 15 psi (1.05 kg/cm2) for 15 min.

2. SOC medium

   To 950 ml of deionized H_2O, add:

Tryptone	20 g/L
Yeast extract	5 g/L
NaCl	0.5 g/L
250 mM KCl	10 mL

   Adjust the medium to pH 7.0 with 5 N NaOH

   After autoclaving, add 20 ml of filtrated 1 M glucose and 5 ml of filtrated 2 M $MgCl_2$

3. TB solution

10 mM Pipes	(PIPES	=	3.021 g/L)
55 mM $MnCl_2$	($MnCl_2.4H_2O$	=	10.885 g/L)
15 mM $CaCl_2$	($CaCl_2.2H_2O$	=	2.205 g/L)
250 mM KCl	(KCl	=	18.637 g/L)

   All the components except for $MnCl_2$ were mixed and adjusted to pH 6.7 with KOH. Then, MnCl2 was dissolved, the solution was sterilized by filtration and stored at 4ºC, all salts were added as solids, always kept, and used in cold.

APPENDIX B

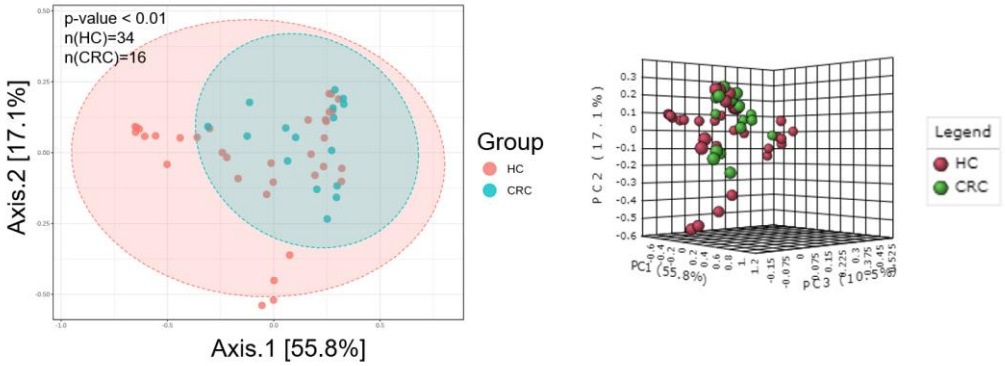


Figure A1. Beta diversity of mucosal-associated microbiota between HC and CRC groups.

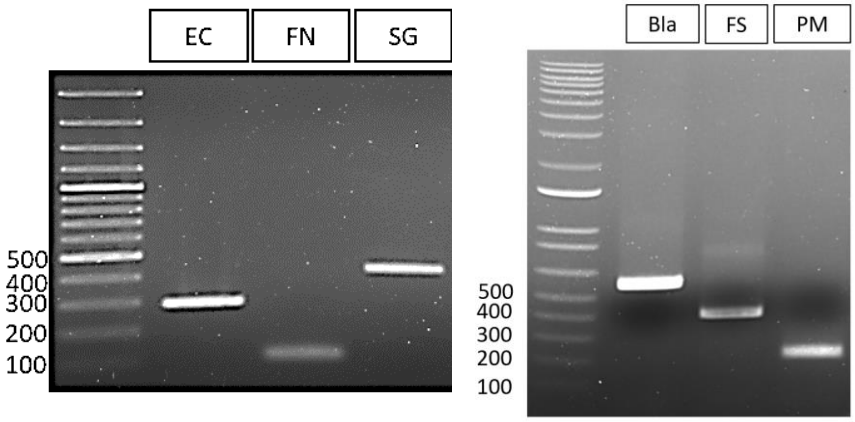


Figure A2. Agarose gel electrophoresis of polymerase chain reaction (PCR) products using primers specific for each bacterium. Abbreviations: FN, *F.nucleatum* (112 bp).; EC, colibactin positive strains (283 bp); PM, *P.micra* (200 bp); Bla, *Blautia* spp. (559 bp); SG, *S.gallolyticus* (408 bp); FS, *F.saccharivorans* (489 bp).

Table A1. Quality and quantity of plasmid DNA used as DNA standard curve.

Plasmid DNA	Concentration (ng/ul)	260/280	260/230
E. coli	154.6	2.07	2.36
F. nucleatum	97.6	2.06	1.98
S. gallolyticus	61.0	2.04	2.11
F. saccharivorans	78.2	1.93	2.06
P. micra	97.2	1.91	2.04
Blautia spp.	96.6	1.87	2.08

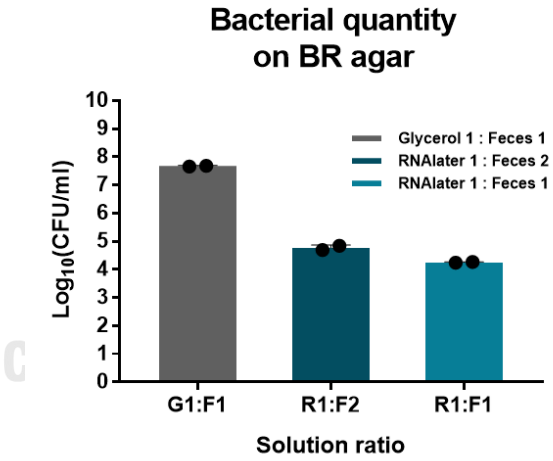


Figure A3. The quantity of fecal bacteria on brucella blood (BR) agar after fecal samples were preserved with glycerol or RNA later reagents for 12 hours**.** The graph represented in Log_{10} of colony forming unit per ml. Abbreviation: G, glycerol; F, feces; R, DNA/RNA protection reagent (RNA later).

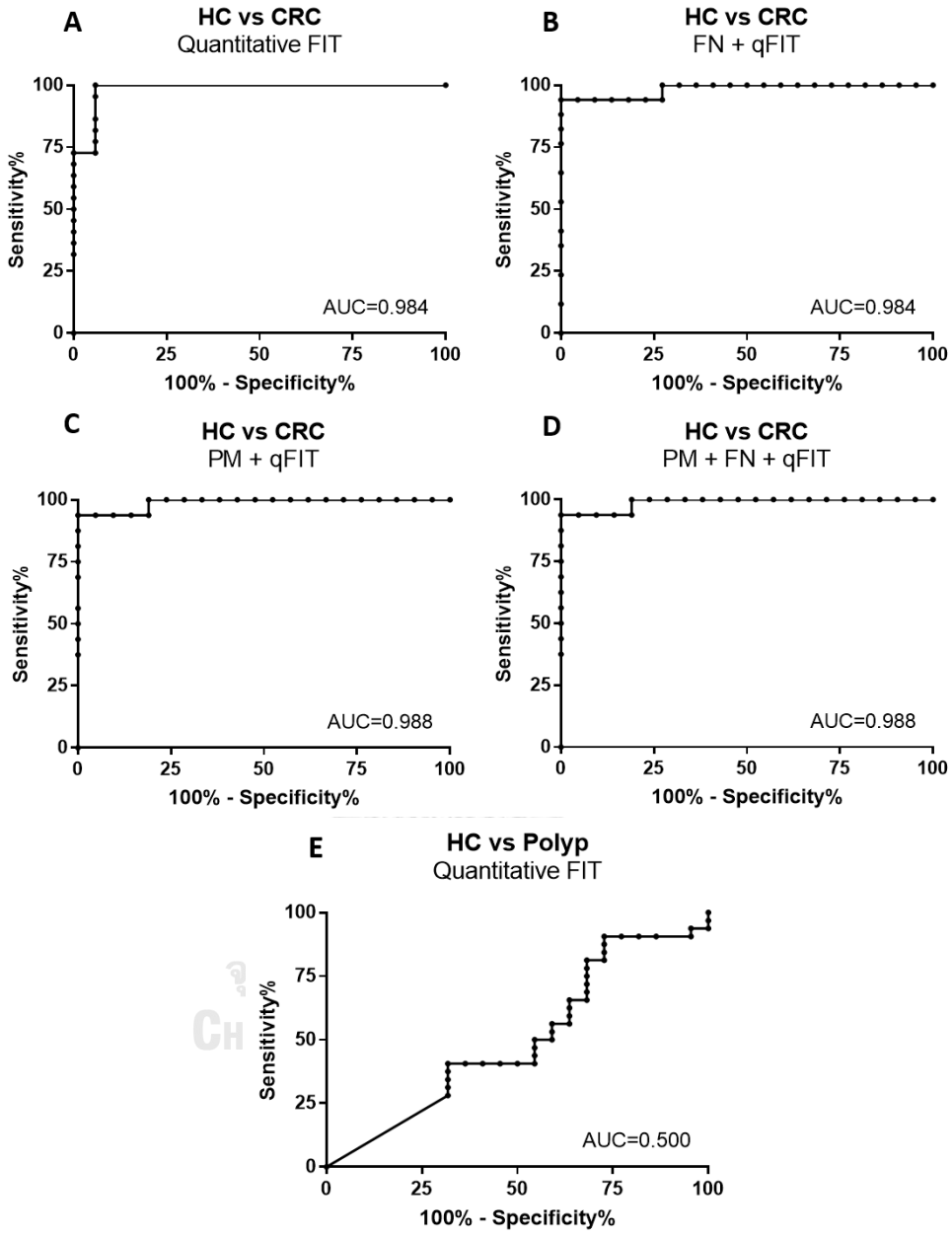


Figure A4. The receiver operating characteristic (ROC) curves displaying the specificity and the sensitivity for the combination of PM, FN, and quantitative FIT to detect CRC and adenomas. Abbreviations: PM, *P. micra*; FN, *F. nucleatum*; FIT, fecal immunochemical test, AUC, area under the curve; HC, Healthy controls; CRC, colorectal cancer.

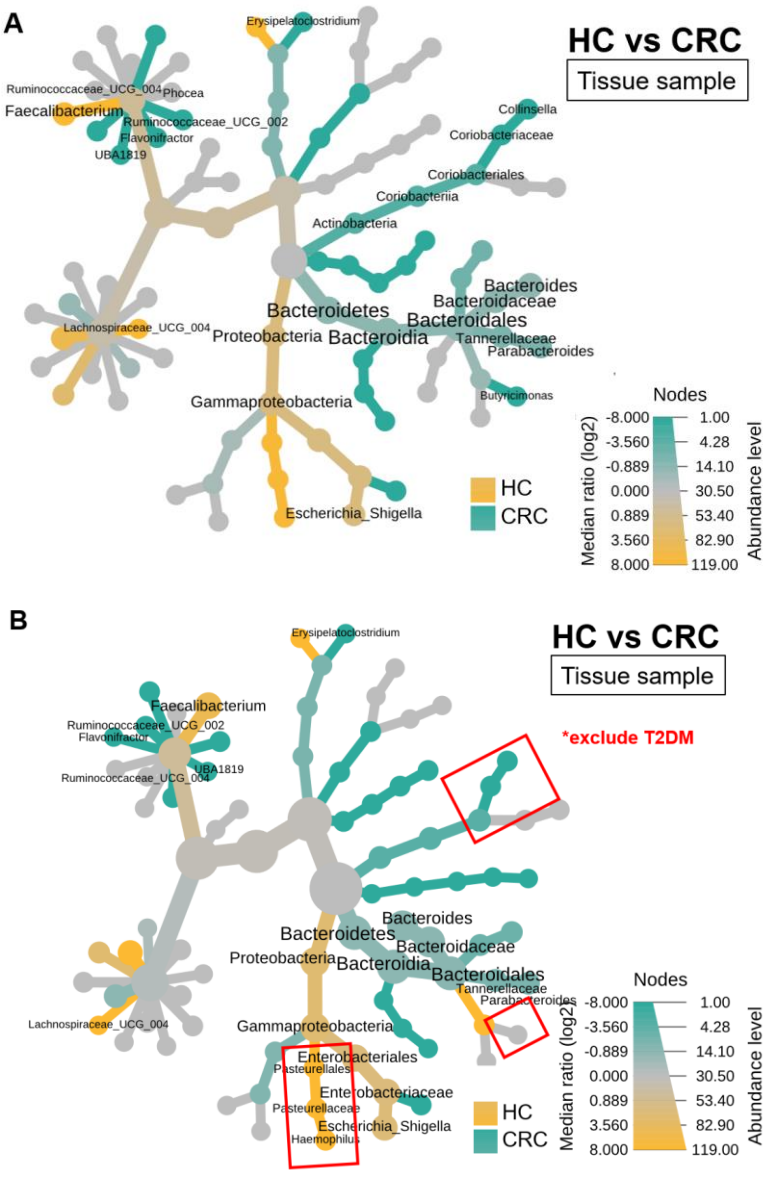


Figure A5. The difference of mucosa-associated microbiota between before and after excluding the CRC patients with T2DM. A) before, B) after.

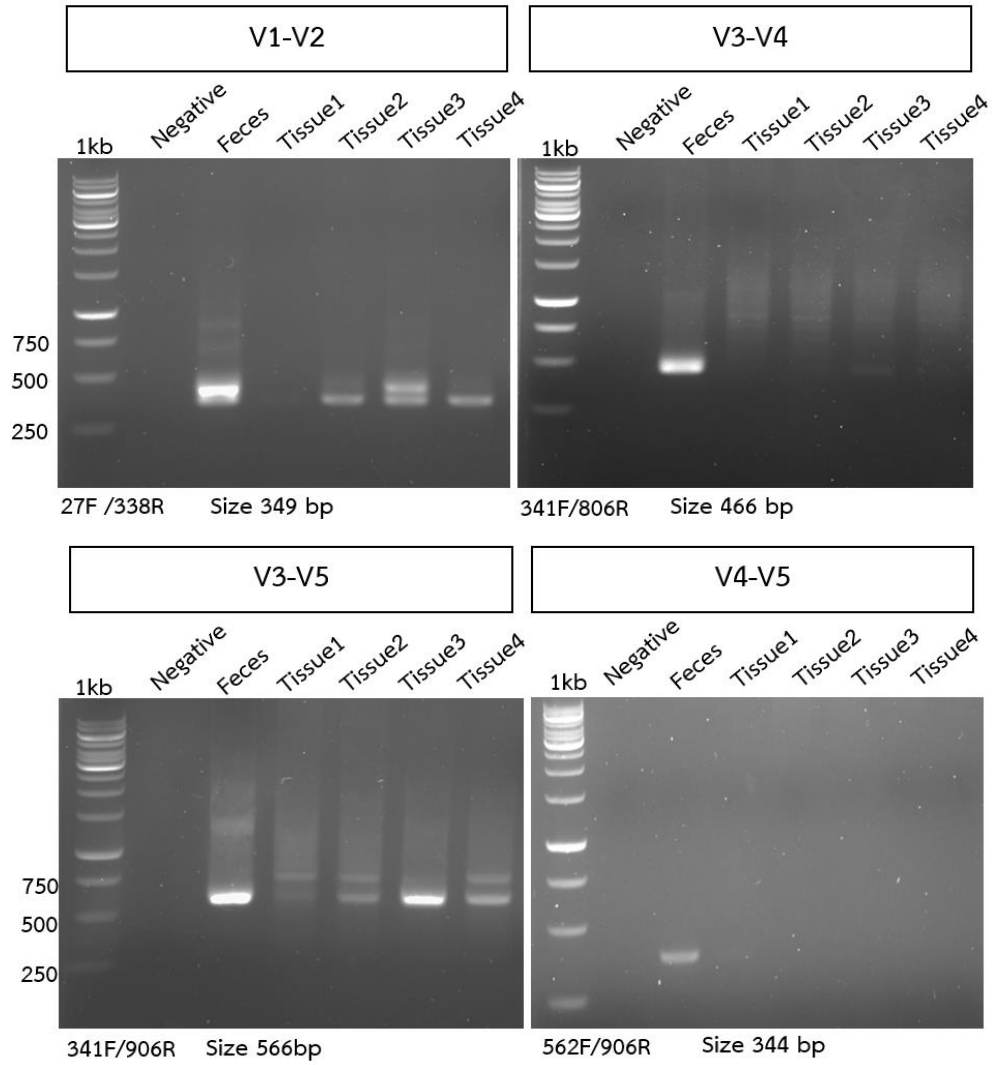


Figure A6. Agarose gel electrophoresis of polymerase chain reaction (PCR) products using different primers of each hypervariable region of 16S rRNA gene in tissue and fecal samples.

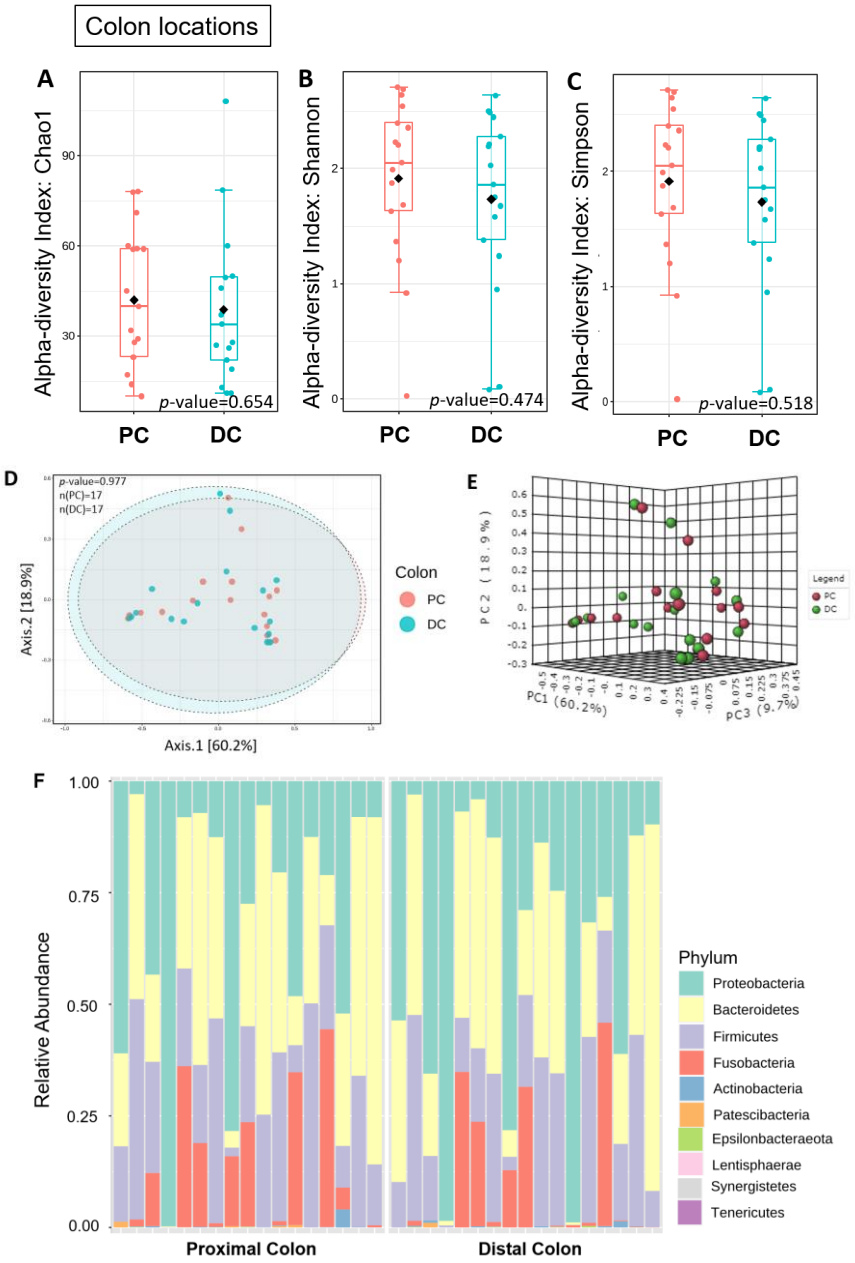
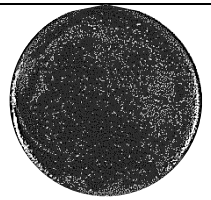
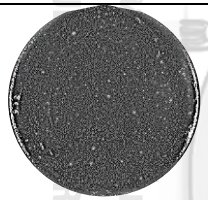
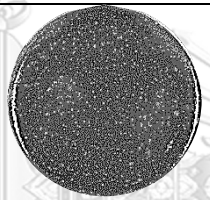
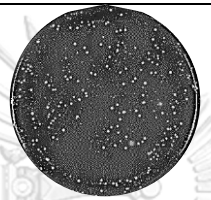
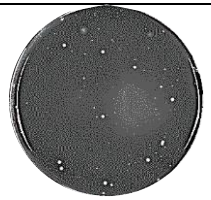
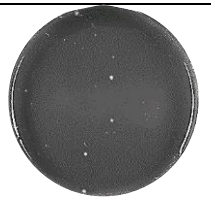
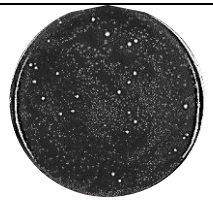
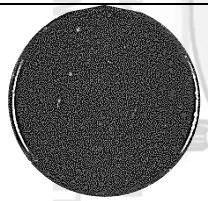
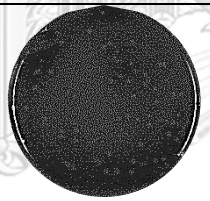
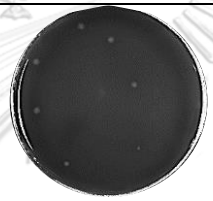
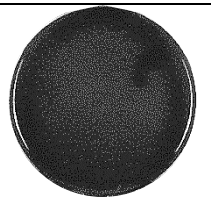
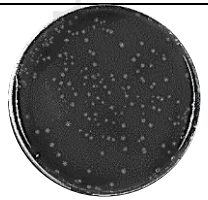
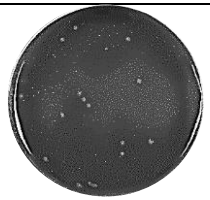


Figure A7. Alpha and beta diversity analysis of proximal and distal colon's HC group. A) Chao1's index, B) Shannon's index, C) Simpson's index, D) PCoA in 2-Dimension, E) PCoA in 3-Dimension, F) Taxonomic composition between proximal colon and distal colon. Data are shown as relative abundance.

Table A2. The effect of DNA/RNA protection reagent on fecal bacterial growth after fecal samples were preserved with the reagents for 12 hours. The table shows the enumeration of fecal bacterial colonies on brucella blood (BR) agar according to each dilution and different ratios of feces and reagent.

Ratio	10^{-1} dilution	10^{-2} dilution	10^{-3} dilution	10^{-4} dilution	10^{-5} dilution	10^{-6} dilution
G:F 1:1						
R:F 1:2						
R:F 1:1						

Abbreviation: G, glycerol; F, feces; R, DNA/RNA protection reagent.

Fusobacterium nucleatum subsp. animalis strain ChDC F332, complete genome

Sequence ID: CP022124.1 Length: 2322902 Number of Matches: 1

Range 1: 552780 to 552883 GenBank Graphics ▼ Next Match ▲ Previou

Score	Expect	Identities	Gaps	Strand
187 bits(101)	3e-44	103/104(99%)	0/104(0%)	Plus/Minus

```
Query  6       GTTGACTTTACAGAAGGAGATTATGTAAAAATCTTAAAAGGTTCATTTAAAGATCAAGAA  65
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  552883  GTTGACTTTACAGAAGGAGATTATGTAAAAATCTTAAAAGGTTCATTTAAAGATCAAGAA  552824

Query  66      GGACAAGTTGCTGAAATTGATCATGAACATGGTAGAGTTAAAGT  109
               |||||||||||||||| ||||||||||||||||||||||||||
Sbjct  552823  GGACAAGTTGCTGAAATAGATCATGAACATGGTAGAGTTAAAGT  552780
```

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Fusobacterium nucleatum subsp. animalis strain ChDC F332, complete genome	Fusobacteri...	187	187	95%	3e-44	99.04%	2322902	CP022124.1
☑	Fusobacterium nucleatum subsp. animalis strain KCOM 1325, complete genome	Fusobacteri...	187	187	95%	3e-44	99.04%	2310804	CP012715.1
☑	Fusobacterium nucleatum subsp. animalis 7_1, complete genome	Fusobacteri...	187	187	95%	3e-44	99.04%	2507720	CP007062.1
☑	Fusobacterium nucleatum subsp. animalis 4_8, complete genome	Fusobacteri...	187	187	95%	3e-44	99.04%	2261267	CP003723.1
☑	Fusobacterium nucleatum subsp. animalis strain KCOM 1279, complete genome	Fusobacteri...	182	182	95%	2e-42	98.08%	2549353	CP012713.1
☑	Fusobacterium nucleatum subsp. nucleatum ATCC 23726 chromosome, complete g...	Fusobacteri...	180	180	94%	5e-42	98.06%	2299539	CP028109.1
☑	Fusobacterium nucleatum subsp. nucleatum strain 25586 chromosome, complete g...	Fusobacteri...	180	180	94%	5e-42	98.06%	2180101	CP028101.1
☑	Fusobacterium nucleatum subsp. nucleatum strain ChDC F317, complete genome	Fusobacteri...	180	180	94%	5e-42	98.06%	2233010	CP022122.1
☑	Fusobacterium nucleatum subsp. nucleatum strain KCOM 1250, complete genome	Fusobacteri...	180	180	94%	5e-42	98.06%	2290405	CP012717.1

Figure A8. Sequencing result with nucleotide BLAST of *Fusobacterium nucleatum* (*nusG* gene) that used as a positive control for qPCR.

Escherichia coli strain ATCC 25922 chromosome, complete genome

Sequence ID: CP032085.1 Length: 5152857 Number of Matches: 1

Range 1: 1173441 to 1173716 GenBank Graphics ▼ Next Match ▲ Previous M

Score	Expect	Identities	Gaps	Strand
510 bits(276)	5e-141	276/276(100%)	0/276(0%)	Plus/Plus

```
Query  6        GCGCATCCTCAAGAGTAAATACGCCAGCAATACAGGCCGCAACATACTCGCCCAGACTGT  65
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1173441  GCGCATCCTCAAGAGTAAATACGCCAGCAATACAGGCCGCAACATACTCGCCCAGACTGT  1173500

Query  66       GCCCGATCATACTGTCCGGCGTAATCCCCAGATCAATCAAGGTGCGCGCTAGGCTGTATT  125
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1173501  GCCCGATCATACTGTCCGGCGTAATCCCCAGATCAATCAAGGTGCGCGCTAGGCTGTATT  1173560

Query  126      CGACACAGAACAACGCCGGTTGTGTGAACTGCGTTTGATCAATATCCCGCGTATCCTCAT  185
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1173561  CGACACAGAACAACGCCGGTTGTGTGAACTGCGTTTGATCAATATCCCGCGTATCCTCAT  1173620

Query  186      CGTCAAACAACAACGCCTTGAGATCGACCGTCGAATAGCGTTGGAATGCAGCAAAGCAGC  245
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1173621  CGTCAAACAACAACGCCTTGAGATCGACCGTCGAATAGCGTTGGAATGCAGCAAAGCAGC  1173680

Query  246      GATCCATGTGCTGGCGAAACATGGGTTGATGAGCAT  281
                ||||||||||||||||||||||||||||||||||||
Sbjct  1173681  GATCCATGTGCTGGCGAAACATGGGTTGATGAGCAT  1173716
```

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Escherichia coli strain EcPF5 chromosome, complete genome	Escherichia coli	510	510	98%	5e-141	100.00%	5147412	CP054236.1
☑	Escherichia coli strain SCU-488 chromosome, complete genome	Escherichia coli	510	510	98%	5e-141	100.00%	5065998	CP054449.1
☑	Escherichia coli strain FDAARGOS 1386 chromosome, complete genome	Escherichia coli	510	510	98%	5e-141	100.00%	5044147	CP077294.1
☑	Klebsiella pneumoniae strain KP59 chromosome	Klebsiella pneumoniae	510	510	98%	5e-141	100.00%	6064517	CP076322.1
☑	Escherichia coli strain F11 chromosome, complete genome	Escherichia coli	510	510	98%	5e-141	100.00%	5048308	CP076123.1
☑	Klebsiella pneumoniae strain 2016_49 chromosome, complete genome	Klebsiella pneumoniae	510	510	98%	5e-141	100.00%	5597599	CP068015.1
☑	Escherichia coli strain JA0072 chromosome, complete genome	Escherichia coli	510	510	98%	5e-141	100.00%	5030087	CP070227.1

Figure A9. Sequencing result with nucleotide BLAST of colibactin positive strains (*clbB* gene) that used as a positive control for qPCR.

Streptococcus gallolyticus subsp. gallolyticus strain ATCC 9809 manganese-dependent superoxide dismutase (sodA) gene, partial cds

Sequence ID: GU991733.1 Length: 409 Number of Matches: 1

Range 1: 7 to 409 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
745 bits(403)	0.0	403/403(100%)	0/403(0%)	Plus/Minus

```
Query  3    GCTTTTCCTTGTGAAATTGGTGTGTCTTGATTAGCTGTCGAGAGCACTTCAAGTTTGCCA  62
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  409  GCTTTTCCTTGTGAAATTGGTGTGTCTTGATTAGCTGTCGAGAGCACTTCAAGTTTGCCA  350

Query  63   TTTTCATTCACCACAAGCCAAGCCCAACCTGACCCAAAGCGAGTTGTCGCAGCTTGCGTG  122
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  349  TTTTCATTCACCACAAGCCAAGCCCAACCTGACCCAAAGCGAGTTGTCGCAGCTTGCGTG  290

Query  123  AAAGCAGCTTTGAATTCGTCAAATGAGCCAAAAGCTTCCTCAATCGCAGCCAACACTTGC  182
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  289  AAAGCAGCTTTGAATTCGTCAAATGAGCCAAAAGCTTCCTCAATCGCAGCCAACACTTGC  230

Query  183  GCTGTTGGTTCTTGTTTTTCAGGCGATAACAATTCCCAGAAAAGGGCGTGATTCAAATGC  242
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  229  GCTGTTGGTTCTTGTTTTTCAGGCGATAACAATTCCCAGAAAAGGGCGTGATTCAAATGC  170
```

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Streptococcus gallolyticus subsp. gallolyticus strain ATCC 9809 manganese-dependent superoxide dism...	Streptococcus ...	745	745	99%	0.0	100.00%	409	GU991733.1
☑	Streptococcus gallolyticus subsp. gallolyticus strain LMG 17956 manganese-dependent superoxide dism...	Streptococcus ...	743	743	99%	0.0	99.75%	415	GU991767.1
☑	Streptococcus gallolyticus subsp. gallolyticus strain LMG 16005 manganese-dependent superoxide dism...	Streptococcus ...	743	743	99%	0.0	99.75%	422	GU991763.1
☑	Streptococcus gallolyticus subsp. gallolyticus strain LMG 14876 manganese-dependent superoxide dism...	Streptococcus ...	743	743	99%	0.0	99.75%	417	GU991759.1
☑	Streptococcus gallolyticus subsp. gallolyticus isolate 12932 manganese-dependent superoxide dismutas...	Streptococcus ...	743	743	99%	0.0	99.75%	425	FJ042703.1
☑	Streptococcus gallolyticus strain 4-C11 manganese-dependent superoxide dismutase (sodA) gene, parti...	Streptococcus ...	743	743	99%	0.0	99.75%	436	AY035714.1

Figure A10. Sequencing result with nucleotide BLAST of *Streptococcus gallolyticus* (*sodA* gene) that used as a positive control for qPCR.

Parvimonas micra strain NCTC11808 genome assembly, chromosome: 1

Sequence ID: LR134472.1 Length: 1677398 Number of Matches: 1

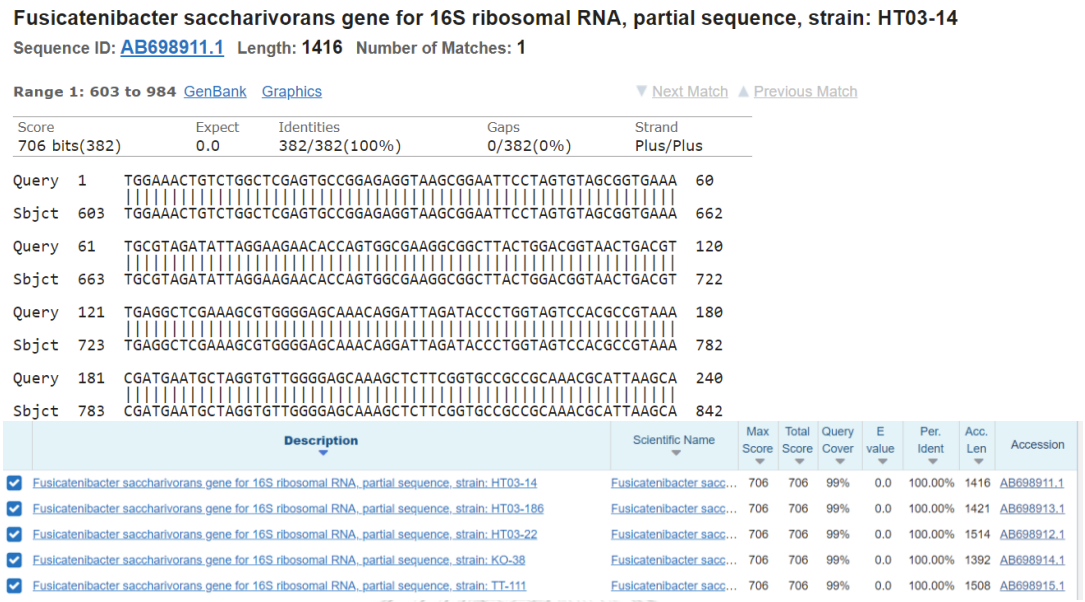
Range 1: 1116773 to 1116967 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
355 bits(192)	2e-94	194/195(99%)	0/195(0%)	Plus/Plus

```
Query  1        ACGGAAGAATTTGTCCGATAGAAACTCCAGAGGGTCCAAACATCGGTCTTATTACTTCTC  60
                ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Sbjct  1116773  ACGGAAGAATTTGTCCGATAGAAACTCCAGAAGGTCCAAACATCGGTCTTATTACTTCTC  1116832

Query  61       TTACAACTTATGCAAGAGTTGATCAATATGGATTTATTGAAACACCATATCGTGTTGTAA  120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1116833  TTACAACTTATGCAAGAGTTGATCAATATGGATTTATTGAAACACCATATCGTGTTGTAA  1116892

Query  121      ATAATGGAATTGCTACAAAGGACATTGTTTATTTAACTGCTGATGAAGAAGATGAAGTTA  180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1116893  ATAATGGAATTGCTACAAAGGACATTGTTTATTTAACTGCTGATGAAGAAGATGAAGTTA  1116952

Query  181      TTATCGCTCAAGCCA  195
                |||||||||||||||
Sbjct  1116953  TTATCGCTCAAGCCA  1116967
```

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Parvimonas micra strain NCTC11808 genome assembly, chromosome: 1	Parvimonas micra	355	355	100%	2e-94	99.49%	1677398	LR134472.1
☑	Parvimonas micra strain KCOM 1535, complete genome	Parvimonas micra	355	355	100%	2e-94	99.49%	1627009	CP009761.1
☑	Parvimonas micra strain KCOM 1037 chromosome, complete genome	Parvimonas micra	350	350	100%	8e-93	98.97%	1661863	CP031971.1
☑	Riemerella sp. IPDH 98/90 RpoB (rpoB) gene, partial cds	Riemerella sp. IPDH 98/90	87.9	87.9	57%	7e-14	80.70%	493	FJ999744.1
☑	Spiroplasma melliferum strain AS576 chromosome, complete genome	Spiroplasma melliferum	78.7	78.7	83%	4e-11	75.46%	1320490	CP029202.1

Figure A11. Sequencing result with nucleotide BLAST of *Parvimonas micra* (*rpoB* gene) that used as a positive control for qPCR.



Fusicatenibacter saccharivorans gene for 16S ribosomal RNA, partial sequence, strain: HT03-14

Sequence ID: AB698911.1 Length: 1416 Number of Matches: 1

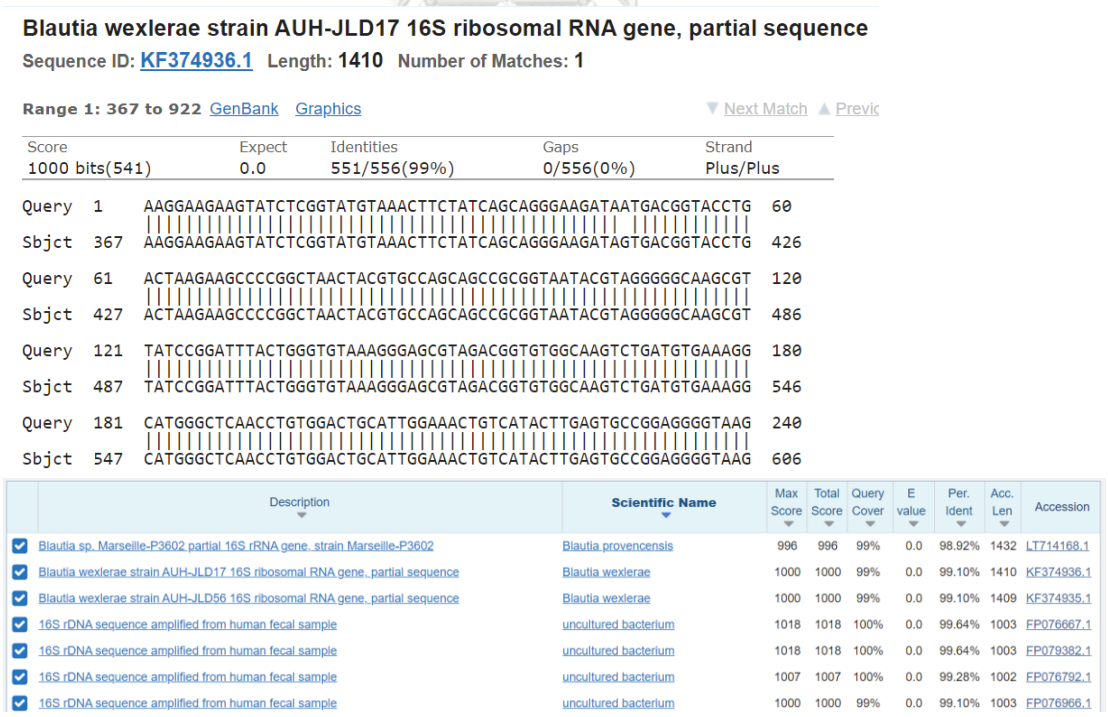
Range 1: 603 to 984 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
706 bits(382)	0.0	382/382(100%)	0/382(0%)	Plus/Plus

```
Query  1    TGGAAACTGTCTGGCTCGAGTGCCGGAGAGGTAAGCGGAATTCCTAGTGTAGCGGTGAAA  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  603  TGGAAACTGTCTGGCTCGAGTGCCGGAGAGGTAAGCGGAATTCCTAGTGTAGCGGTGAAA  662

Query  61   TGCGTAGATATTAGGAAGAACACCAGTGGCGAAGGCGGCTTACTGGACGGTAACTGACGT  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  663  TGCGTAGATATTAGGAAGAACACCAGTGGCGAAGGCGGCTTACTGGACGGTAACTGACGT  722

Query  121  TGAGGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAA  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  723  TGAGGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAA  782

Query  181  CGATGAATGCTAGGTGTTGGGGAGCAAAGCTCTTCGGTGCCGCCGCAAACGCATTAAGCA  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  783  CGATGAATGCTAGGTGTTGGGGAGCAAAGCTCTTCGGTGCCGCCGCAAACGCATTAAGCA  842
```

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Fusicatenibacter saccharivorans gene for 16S ribosomal RNA, partial sequence, strain: HT03-14	Fusicatenibacter sacc...	706	706	99%	0.0	100.00%	1416	AB698911.1
Fusicatenibacter saccharivorans gene for 16S ribosomal RNA, partial sequence, strain: HT03-186	Fusicatenibacter sacc...	706	706	99%	0.0	100.00%	1421	AB698913.1
Fusicatenibacter saccharivorans gene for 16S ribosomal RNA, partial sequence, strain: HT03-22	Fusicatenibacter sacc...	706	706	99%	0.0	100.00%	1514	AB698912.1
Fusicatenibacter saccharivorans gene for 16S ribosomal RNA, partial sequence, strain: KO-38	Fusicatenibacter sacc...	706	706	99%	0.0	100.00%	1392	AB698914.1
Fusicatenibacter saccharivorans gene for 16S ribosomal RNA, partial sequence, strain: TT-111	Fusicatenibacter sacc...	706	706	99%	0.0	100.00%	1508	AB698915.1

Figure A12. Sequencing result with nucleotide BLAST of *Fusicatenibacter saccharivorans* (*16S rRNA* gene) that used as a positive control for qPCR.



Blautia wexlerae strain AUH-JLD17 16S ribosomal RNA gene, partial sequence

Sequence ID: KF374936.1 Length: 1410 Number of Matches: 1

Range 1: 367 to 922 GenBank Graphics ▼ Next Match ▲ Previc

Score	Expect	Identities	Gaps	Strand
1000 bits(541)	0.0	551/556(99%)	0/556(0%)	Plus/Plus

```
Query  1    AAGGAAGAAGTATCTCGGTATGTAAACTTCTATCAGCAGGGAAGATAATGACGGTACCTG  60
            ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Sbjct  367  AAGGAAGAAGTATCTCGGTATGTAAACTTCTATCAGCAGGGAAGATAGTGACGGTACCTG  426

Query  61   ACTAAGAAGCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGT  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  427  ACTAAGAAGCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGT  486

Query  121  TATCCGGATTTACTGGGTGTAAAGGGAGCGTAGACGGTGTGGCAAGTCTGATGTGAAAGG  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  487  TATCCGGATTTACTGGGTGTAAAGGGAGCGTAGACGGTGTGGCAAGTCTGATGTGAAAGG  546

Query  181  CATGGGCTCAACCTGTGGACTGCATTGGAAACTGTCATACTTGAGTGCCGGAGGGGTAAG  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  547  CATGGGCTCAACCTGTGGACTGCATTGGAAACTGTCATACTTGAGTGCCGGAGGGGTAAG  606
```

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Blautia sp. Marseille-P3602 partial 16S rRNA gene, strain Marseille-P3602	Blautia provencensis	996	996	99%	0.0	98.92%	1432	LT714168.1
Blautia wexlerae strain AUH-JLD17 16S ribosomal RNA gene, partial sequence	Blautia wexlerae	1000	1000	99%	0.0	99.10%	1410	KF374936.1
Blautia wexlerae strain AUH-JLD56 16S ribosomal RNA gene, partial sequence	Blautia wexlerae	1000	1000	99%	0.0	99.10%	1409	KF374935.1
16S rDNA sequence amplified from human fecal sample	uncultured bacterium	1018	1018	100%	0.0	99.64%	1003	FP076667.1
16S rDNA sequence amplified from human fecal sample	uncultured bacterium	1018	1018	100%	0.0	99.64%	1003	FP079382.1
16S rDNA sequence amplified from human fecal sample	uncultured bacterium	1007	1007	100%	0.0	99.28%	1002	FP076792.1
16S rDNA sequence amplified from human fecal sample	uncultured bacterium	1000	1000	99%	0.0	99.10%	1003	FP076966.1

Figure A13. Sequencing result with nucleotide BLAST of *Fusicatenibacter saccharivorans* (16S rRNA gene) that used as a positive control for qPCR.

Homo sapiens isolate CHM13 chromosome 17

Sequence ID: CP068261.2 **Length:** 84276897 **Number of Matches:** 1

Range 1: 13743979 to 13744154 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
320 bits(173)	4e-85	176/177(99%)	1/177(0%)	Plus/Plus

```
Query  1         AGTCCAATGATGGTAGTAATAGTGATTGTCATTGTGTTGTTATTATTGCAGTTCCTAGCC  60
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  13743979  AGTCCAATGATGGTAGTAATAGTGATTGTCATTGTGTTGTTATTATTGCAGTTCCTAGCC  13744038

Query  61        CAGTGCTATTATTCATCATAACTTACAACCTACCCCCATTCCTACCCTATTTTAAATATA  120
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  13744039  CAGTGCTATTATTCATCATAACTTACAACCTACCCCCATTCCTACCCTATTTTAAATATA  13744098

Query  121       TTTTAGTTCACCATTCATAGGCCACTTAAATATATATTCACTCACAAGGAAAAGCAC  177
                 ||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Sbjct  13744099  TTTTAGTTCACCATTCATAGGCCACTTAAATATATATTCACTCACAAGGAAA-GCAC  13744154
```

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Homo sapiens BAC clone VMRC59-374G12 from chromosome unknown, complete sequence	Homo sapiens	320	320	97%	4e-85	99.44%	133892	AC270119.1
☑	Homo sapiens isolate CHM13 chromosome 17	Homo sapiens	320	320	97%	4e-85	99.44%	84276897	CP068261.2
☑	Homo sapiens DNA, chromosome 17, nearly complete genome	Homo sapiens	320	320	97%	4e-85	99.44%	80688777	AP023477.1
☑	Homo sapiens chromosome 17, clone RP11-952N18, complete sequence	Homo sapiens	320	320	97%	4e-85	99.44%	97676	AC005497.9

Figure A14. Sequencing result with nucleotide BLAST of non-specific product of *S. gallolyticus* primers in tissue samples

Human DNA sequence from clone RP11-376F21 on chromosome 9, complete sequence

Sequence ID: AL583849.10 **Length:** 48173 **Number of Matches:** 1

Range 1: 3364 to 3449 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
132 bits(71)	1e-28	82/87(94%)	1/87(1%)	Plus/Minus

```
Query  22    TCTTTTAGGCAGAGGGTGGGGAAGTACTGCGCCCTTGGACTAAGGTCCTGCCGGGCTGGC  81
             |||| |||||||||| ||||||||||||| ||||||| |||||||||||||||||||||
Sbjct  3449  TCTTCTAGGCAGAGG-TGGGGAAGTACTGAGCCCTTGAACTAAGGTCCTGCCGGGCTGGC  3391

Query  82    ACTGGGGCCCTGTGGTGACTCCGGTGC  108
             |||||||||||||||||||||| ||||
Sbjct  3390  ACTGGGGCCCTGTGGTGACTCCAGTGC  3364
```

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Human DNA sequence from clone RP11-376F21 on chromosome 9, complete sequence	Homo sapiens	132	132	79%	1e-28	94.25%	48173	AL583849.10

Figure A15. Sequencing result with nucleotide BLAST of non-specific product of *F. saccharivorans* primers in tissue samples

REFERENCES

1. Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, et al. Colorectal cancer statistics, 2020. CA Cancer J Clin. 2020;70(3):145-64.

2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021.

3. Thrumurthy SG, Thrumurthy SS, Gilbert CE, Ross P, Haji A. Colorectal adenocarcinoma: risks, prevention and diagnosis. Bmj. 2016;354:i3590.

4. Sung JJ, Ng SC, Chan FK, Chiu HM, Kim HS, Matsuda T, et al. An updated Asia Pacific Consensus Recommendations on colorectal cancer screening. Gut. 2015;64(1):121-32.

5. Rex DK, Johnson DA, Anderson JC, Schoenfeld PS, Burke CA, Inadomi JM. American College of Gastroenterology Guidelines for Colorectal Cancer Screening 2008. Official journal of the American College of Gastroenterology | ACG. 2009;104(3).

6. Nistal E, Fernández-Fernández N, Vivas S, Olcoz JL. Factors Determining Colorectal Cancer: The Role of the Intestinal Microbiota. Front Oncol. 2015;5:220.

7. Dulal S, Keku TO. Gut microbiome and colorectal adenomas. Cancer J. 2014;20(3):225-31.

8. Zackular JP, Baxter NT, Iverson KD, Sadler WD, Petrosino JF, Chen GY, et al. The gut microbiome modulates colon tumorigenesis. mBio. 2013;4(6):e00692-13.

9. Cuevas-Ramos G, Petit CR, Marcq I, Boury M, Oswald E, Nougayrède JP. Escherichia coli induces DNA damage in vivo and triggers genomic instability in mammalian cells. Proc Natl Acad Sci U S A. 2010;107(25):11537-42.

10. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, et al. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. Cell Host Microbe. 2013;14(2):207-15.

11. Wu S, Lim KC, Huang J, Saidi RF, Sears CL. Bacteroides fragilis enterotoxin cleaves the zonula adherens protein, E-cadherin. Proc Natl Acad Sci U S A. 1998;95(25):14979-84.

12. Osman M-A, Neoh H-m, Ab Mutalib N-S, Chin S-F, Jamal R. 16S rRNA Gene Sequencing for Deciphering the Colorectal Cancer Gut Microbiome: Current Protocols and Workflows. Frontiers in Microbiology. 2018;9(767).

13. Villéger R, Lopès A, Veziant J, Gagnière J, Barnich N, Billard E, et al. Microbial markers in colorectal cancer detection and/or prognosis. World J Gastroenterol. 2018;24(22):2327-47.

14. Flemer B, Lynch DB, Brown JM, Jeffery IB, Ryan FJ, Claesson MJ, et al. Tumour-associated and non-tumour-associated microbiota in colorectal cancer. Gut. 2017;66(4):633-43.

15. Levine JS, Ahnen DJ. Clinical practice. Adenomatous polyps of the colon. N Engl J Med.

2006;355(24):2551-7.

16. Stryker SJ, Wolff BG, Culp CE, Libbe SD, Ilstrup DM, MacCarty RL. Natural history of untreated colonic polyps. Gastroenterology. 1987;93(5):1009-13.

17. Migliore L, Migheli F, Spisni R, Coppedè F. Genetics, Cytogenetics, and Epigenetics of Colorectal Cancer. Journal of Biomedicine and Biotechnology. 2011;2011:792362.

18. Keum N, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. Nat Rev Gastroenterol Hepatol. 2019;16(12):713-32.

19. Issa IA, Noureddine M. Colorectal cancer screening: An updated review of the available options. World J Gastroenterol. 2017;23(28):5086-96.

20. Maida M, Macaluso FS, Ianiro G, Mangiola F, Sinagra E, Hold G, et al. Screening of colorectal cancer: present and future. Expert Rev Anticancer Ther. 2017;17(12):1131-46.

21. Bevan R, Rutter MD. Colorectal Cancer Screening-Who, How, and When? Clinical endoscopy. 2018;51(1):37-49.

22. Rutter MD, Nickerson C, Rees CJ, Patnick J, Blanks RG. Risk factors for adverse events related to polypectomy in the English Bowel Cancer Screening Programme. Endoscopy. 2014;46(2):90-7.

23. Vital signs: Colorectal cancer screening, incidence, and mortality--United States, 2002-2010. MMWR Morb Mortal Wkly Rep. 2011;60(26):884-9.

24. Carroll MR, Seaman HE, Halloran SP. Tests and investigations for colorectal cancer screening. Clin Biochem. 2014;47(10-11):921-39.

25. Bray C, Bell LN, Liang H, Collins D, Yale SH. Colorectal Cancer Screening. Wmj. 2017;116(1):27-33.

26. Wu GD, Lewis JD. Analysis of the human gut microbiome and association with disease. Clin Gastroenterol Hepatol. 2013;11(7):774-7.

27. Gagnière J, Raisch J, Veziant J, Barnich N, Bonnet R, Buc E, et al. Gut microbiota imbalance and colorectal cancer. World J Gastroenterol. 2016;22(2):501-18.

28. Tlaskalová-Hogenová H, Stepánková R, Hudcovic T, Tucková L, Cukrowska B, Lodinová-Zádníková R, et al. Commensal bacteria (normal microflora), mucosal immunity and chronic inflammatory and autoimmune diseases. Immunol Lett. 2004;93(2-3):97-108.

29. Hullar MA, Burnett-Hartman AN, Lampe JW. Gut microbes, diet, and cancer. Cancer Treat Res. 2014;159:377-99.

30. Thompson-Chagoyán OC, Maldonado J, Gil A. Colonization and impact of disease and other factors on intestinal microbiota. Dig Dis Sci. 2007;52(9):2069-77.

31. Reddy BS, Weisburger JH, Narisawa T, Wynder EL. Colon carcinogenesis in germ-free rats with 1,2-dimethylhydrazine and N-methyl-n'-nitro-N-nitrosoguanidine. Cancer Res.

1974;34(9):2368-72.

32. Uronis JM, Mühlbauer M, Herfarth HH, Rubinas TC, Jones GS, Jobin C. Modulation of the intestinal microbiota alters colitis-associated colorectal cancer susceptibility. PLoS One. 2009;4(6):e6026.

33. Li Y, Kundu P, Seow SW, de Matos CT, Aronsson L, Chin KC, et al. Gut microbiota accelerate tumor growth via c-jun and STAT3 phosphorylation in APCMin/+ mice. Carcinogenesis. 2012;33(6):1231-8.

34. Vannucci L, Stepankova R, Kozakova H, Fiserova A, Rossmann P, Tlaskalova-Hogenova H. Colorectal carcinogenesis in germ-free and conventionally reared rats: different intestinal environments affect the systemic immunity. Int J Oncol. 2008;32(3):609-17.

35. Zackular JP, Baxter NT, Chen GY, Schloss PD. Manipulation of the Gut Microbiota Reveals Role in Colon Tumorigenesis. mSphere. 2016;1(1).

36. Boleij A, Tjalsma H. Gut bacteria in health and disease: a survey on the interface between intestinal microbiology and colorectal cancer. Biol Rev Camb Philos Soc. 2012;87(3):701-30.

37. Ahn J, Sinha R, Pei Z, Dominianni C, Wu J, Shi J, et al. Human gut microbiome and risk for colorectal cancer. J Natl Cancer Inst. 2013;105(24):1907-11.

38. Huipeng W, Lifeng G, Chuang G, Jiaying Z, Yuankun C. The differences in colonic mucosal microbiota between normal individual and colon cancer patients by polymerase chain reaction-denaturing gradient gel electrophoresis. J Clin Gastroenterol. 2014;48(2):138-44.

39. Shen XJ, Rawls JF, Randall T, Burcal L, Mpande CN, Jenkins N, et al. Molecular characterization of mucosal adherent bacteria and associations with colorectal adenomas. Gut Microbes. 2010;1(3):138-47.

40. Sanapareddy N, Legge RM, Jovov B, McCoy A, Burcal L, Araujo-Perez F, et al. Increased rectal microbial richness is associated with the presence of colorectal adenomas in humans. Isme j. 2012;6(10):1858-68.

41. Wu N, Yang X, Zhang R, Li J, Xiao X, Hu Y, et al. Dysbiosis signature of fecal microbiota in colorectal cancer patients. Microb Ecol. 2013;66(2):462-70.

42. Weir TL, Manter DK, Sheflin AM, Barnett BA, Heuberger AL, Ryan EP. Stool microbiome and metabolome differences between colorectal cancer patients and healthy adults. PLoS One. 2013;8(8):e70803.

43. Wang T, Cai G, Qiu Y, Fei N, Zhang M, Pang X, et al. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. Isme j. 2012;6(2):320-9.

44. Sobhani I, Tap J, Roudot-Thoraval F, Roperch JP, Letulle S, Langella P, et al. Microbial dysbiosis in colorectal cancer (CRC) patients. PLoS One. 2011;6(1):e16393.

45. Kasai C, Sugimoto K, Moritani I, Tanaka J, Oya Y, Inoue H, et al. Comparison of human

gut microbiota in control subjects and patients with colorectal carcinoma in adenoma: Terminal restriction fragment length polymorphism and next-generation sequencing analyses. Oncol Rep. 2016;35(1):325-33.

46. Gao Z, Guo B, Gao R, Zhu Q, Qin H. Microbiota disbiosis is associated with colorectal cancer. Front Microbiol. 2015;6:20.

47. Burns MB, Lynch J, Starr TK, Knights D, Blekhman R. Virulence genes are a signature of the microbiome in the colorectal tumor microenvironment. Genome Medicine. 2015;7(1):55.

48. Gao R, Kong C, Huang L, Li H, Qu X, Liu Z, et al. Mucosa-associated microbiota signature in colorectal cancer. Eur J Clin Microbiol Infect Dis. 2017;36(11):2073-83.

49. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. Mol Syst Biol. 2014;10(11):766.

50. Zackular JP, Rogers MA, Ruffin MTt, Schloss PD. The human gut microbiome as a screening tool for colorectal cancer. Cancer Prev Res (Phila). 2014;7(11):1112-21.

51. Baxter NT, Ruffin MT, Rogers MAM, Schloss PD. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. Genome Medicine. 2016;8(1):37.

52. Mira-Pascual L, Cabrera-Rubio R, Ocon S, Costales P, Parra A, Suarez A, et al. Microbial mucosal colonic shifts associated with the development of colorectal cancer reveal the presence of different bacterial and archaeal biomarkers. Journal of Gastroenterology. 2015;50(2):167-79.

53. Nakatsu G, Li X, Zhou H, Sheng J, Wong SH, Wu WKK, et al. Gut mucosal microbiome across stages of colorectal carcinogenesis. Nature Communications. 2015;6(1):8727.

54. Ternes D, Karta J, Tsenkova M, Wilmes P, Haan S, Letellier E. Microbiome in Colorectal Cancer: How to Get from Meta-omics to Mechanism? Trends Microbiol. 2020;28(5):401-23.

55. Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW. Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/β-catenin signaling via its FadA adhesin. Cell Host Microbe. 2013;14(2):195-206.

56. Nosho K, Sukawa Y, Adachi Y, Ito M, Mitsuhashi K, Kurihara H, et al. Association of Fusobacterium nucleatum with immunity and molecular alterations in colorectal cancer. World J Gastroenterol. 2016;22(2):557-66.

57. Saito T, Nishikawa H, Wada H, Nagano Y, Sugiyama D, Atarashi K, et al. Two FOXP3(+)CD4(+) T cell subpopulations distinctly control the prognosis of colorectal cancers. Nat Med. 2016;22(6):679-84.

58. Gur C, Ibrahim Y, Isaacson B, Yamin R, Abed J, Gamliel M, et al. Binding of the Fap2 protein of Fusobacterium nucleatum to human inhibitory receptor TIGIT protects tumors from immune cell attack. Immunity. 2015;42(2):344-55.

59. Tomkovich S, Yang Y, Winglee K, Gauthier J, Mühlbauer M, Sun X, et al. Locoregional Effects of Microbiota in a Preclinical Model of Colon Carcinogenesis. Cancer Res. 2017;77(10):2620-32.

60. Lee YK, Mehrabian P, Boyajian S, Wu WL, Selicha J, Vonderfecht S, et al. The Protective Role of Bacteroides fragilis in a Murine Model of Colitis-Associated Colorectal Cancer. mSphere. 2018;3(6).

61. Keenan JI, Aitchison A, Purcell RV, Greenlees R, Pearson JF, Frizelle FA. Screening for enterotoxigenic Bacteroides fragilis in stool samples. Anaerobe. 2016;40:50-3.

62. Wu S, Morin PJ, Maouyo D, Sears CL. Bacteroides fragilis enterotoxin induces c-Myc expression and cellular proliferation. Gastroenterology. 2003;124(2):392-400.

63. Chung L, Thiele Orberg E, Geis AL, Chan JL, Fu K, DeStefano Shields CE, et al. Bacteroides fragilis Toxin Coordinates a Pro-carcinogenic Inflammatory Cascade via Targeting of Colonic Epithelial Cells. Cell Host Microbe. 2018;23(2):203-14.e5.

64. Casterline BW, Hecht AL, Choi VM, Bubeck Wardenburg J. The Bacteroides fragilis pathogenicity island links virulence and strain competition. Gut Microbes. 2017;8(4):374-83.

65. Wernke KM, Xue M, Tirla A, Kim CS, Crawford JM, Herzon SB. Structure and bioactivity of colibactin. Bioorganic & Medicinal Chemistry Letters. 2020;30(15):127280.

66. Buc E, Dubois D, Sauvanet P, Raisch J, Delmas J, Darfeuille-Michaud A, et al. High prevalence of mucosa-associated E. coli producing cyclomodulin and genotoxin in colon cancer. PLoS One. 2013;8(2):e56964.

67. Ellmerich S, Djouder N, Schöller M, Klein J-P. PRODUCTION OF CYTOKINES BY MONOCYTES, EPITHELIAL AND ENDOTHELIAL CELLS ACTIVATED BY STREPTOCOCCUS BOVIS. Cytokine. 2000;12(1):26-31.

68. Boleij A, Muytjens CM, Bukhari SI, Cayet N, Glaser P, Hermans PW, et al. Novel clues on the specific association of Streptococcus gallolyticus subsp gallolyticus with colorectal cancer. J Infect Dis. 2011;203(8):1101-9.

69. Kumar R, Herold JL, Taylor J, Xu J, Xu Y. Variations among Streptococcus gallolyticus subsp. gallolyticus strains in connection with colorectal cancer. Scientific Reports. 2018;8(1):1514.

70. Abdulamir AS, Hafidh RR, Mahdi LK, Al-jeboori T, Abubaker F. Investigation into the controversial association of Streptococcus gallolyticus with colorectal cancer and adenoma. BMC Cancer. 2009;9(1):403.

71. Boleij A, Tjalsma H. The itinerary of Streptococcus gallolyticus infection in patients with colonic malignant disease. Lancet Infect Dis. 2013;13(8):719-24.

72. Sears CL, Garrett WS. Microbes, microbiota, and colon cancer. Cell Host Microbe. 2014;15(3):317-28.

73. Valdes AM, Walter J, Segal E, Spector TD. Role of the gut microbiota in nutrition and

health. BMJ. 2018;361:k2179.

74. Ursell LK, Metcalf JL, Parfrey LW, Knight R. Defining the human microbiome. Nutr Rev. 2012;70 Suppl 1(Suppl 1):S38-44.

75. Barko PC, McMichael MA, Swanson KS, Williams DA. The Gastrointestinal Microbiome: A Review. J Vet Intern Med. 2018;32(1):9-25.

76. Ostfeld RS, Keesing F. Biodiversity and Human Health. In: Levin SA, editor. Encyclopedia of Biodiversity (Second Edition). Waltham: Academic Press; 2013. p. 357-72.

77. Yu LC, Wei SC, Ni YH. Impact of microbiota in colorectal carcinogenesis: lessons from experimental models. Intest Res. 2018;16(3):346-57.

78. Wong SH, Kwong TNY, Chow TC, Luk AKC, Dai RZW, Nakatsu G, et al. Quantitation of faecal Fusobacterium improves faecal immunochemical test in detecting advanced colorectal neoplasia. Gut. 2017;66(8):1441-8.

79. Amitay EL, Krilaviciute A, Brenner H. Systematic review: Gut microbiota in fecal samples and detection of colorectal neoplasms. Gut Microbes. 2018;9(4):293-307.

80. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods. 2016;13(7):581-3.

81. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(Database issue):D590-6.

82. Chong J, Liu P, Zhou G, Xia J. Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. Nat Protoc. 2020;15(3):799-821.

83. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. Genome Biology. 2011;12(6):R60.

84. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics. 2012;13(1):134.

85. Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, et al. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res. 2012;22(2):299-306.

86. Arthur JC, Gharaibeh RZ, Mühlbauer M, Perez-Chanona E, Uronis JM, McCafferty J, et al. Microbial genomic analysis reveals the essential role of inflammation in bacteria-induced colorectal cancer. Nat Commun. 2014;5:4724.

87. Abdulamir AS, Hafidh RR, Bakar FA. Molecular detection, quantification, and isolation of Streptococcus gallolyticus bacteria colonizing colorectal tumors: inflammation-driven potential of carcinogenesis via IL-1, COX-2, and IL-8. Mol Cancer. 2010;9:249.

88. Kurakawa T, Ogata K, Matsuda K, Tsuji H, Kubota H, Takada T, et al. Diversity of

Intestinal Clostridium coccoides Group in the Japanese Population, as Demonstrated by Reverse Transcription-Quantitative PCR. PLoS One. 2015;10(5):e0126226.

89. Inoue H, Nojima H, Okayama H. High efficiency transformation of Escherichia coli with plasmids. Gene. 1990;96(1):23-8.

90. Rodriguez-Castaño GP, Caro-Quintero A, Reyes A, Lizcano F. Advances in Gut Microbiome Research, Opening New Strategies to Cope with a Western Lifestyle. Front Genet. 2016;7:224.

91. Rajagopala SV, Vashee S, Oldfield LM, Suzuki Y, Venter JC, Telenti A, et al. The Human Microbiome and Cancer. Cancer Prevention Research. 2017;10(4):226-34.

92. García-Castillo V, Sanhueza E, McNerney E, Onate SA, García A. Microbiota dysbiosis: a new piece in the understanding of the carcinogenesis puzzle. J Med Microbiol. 2016;65(12):1347-62.

93. Wang X, Yang Y, Huycke MM. Microbiome-driven carcinogenesis in colorectal cancer: Models and mechanisms. Free Radic Biol Med. 2017;105:3-15.

94. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences. 2011;108(Supplement 1):4516-22.

95. Dejea CM, Wick EC, Hechenbleikner EM, White JR, Mark Welch JL, Rossetti BJ, et al. Microbiota organization is a distinct feature of proximal colorectal cancers. Proc Natl Acad Sci U S A. 2014;111(51):18321-6.

96. Geng J, Song Q, Tang X, Liang X, Fan H, Peng H, et al. Co-occurrence of driver and passenger bacteria in human colorectal cancer. Gut Pathog. 2014;6:26.

97. Chen W, Liu F, Ling Z, Tong X, Xiang C. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. PLoS One. 2012;7(6):e39743.

98. Zmora N, Zilberman-Schapira G, Suez J, Mor U, Dori-Bachash M, Bashiardes S, et al. Personalized Gut Mucosal Colonization Resistance to Empiric Probiotics Is Associated with Unique Host and Microbiome Features. Cell. 2018;174(6):1388-405.e21.

99. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006;444(7122):1027-31.

100. Donaldson GP, Lee SM, Mazmanian SK. Gut biogeography of the bacterial microbiota. Nature Reviews Microbiology. 2016;14(1):20-32.

101. 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. J Lipid Res. 2013;54(9):2325-40.

102. Tjalsma H, Boleij A, Marchesi JR, Dutilh BE. A bacterial driver-passenger model for

colorectal cancer: beyond the usual suspects. Nat Rev Microbiol. 2012;10(8):575-82.

103. Manfredi R, Nanetti A, Ferri M, Mastroianni A, Coronado OV, Chiodo F. Flavobacterium spp. organisms as opportunistic bacterial pathogens during advanced HIV disease. J Infect. 1999;39(2):146-52.

104. Gong J, Li L, Zuo X, Li Y. Change of the duodenal mucosa-associated microbiota is related to intestinal metaplasia. BMC Microbiology. 2019;19(1):275.

105. Gupta A, Dhakan Darshan B, Maji A, Saxena R, Prasoodanan PKV, Mahajan S, et al. Association of Flavonifractor plautii, a Flavonoid-Degrading Bacterium, with the Gut Microbiome of Colorectal Cancer Patients in India. mSystems.4(6):e00438-19.

106. Braune A, Blaut M. Bacterial species involved in the conversion of dietary flavonoids in the human gut. Gut Microbes. 2016;7(3):216-34.

107. Batra P, Sharma AK. Anti-cancer potential of flavonoids: recent trends and future perspectives. 3 Biotech. 2013;3(6):439-59.

108. Koh GY, Kane AV, Wu X, Crott JW. Parabacteroides distasonis attenuates tumorigenesis, modulates inflammatory markers and promotes intestinal barrier integrity in azoxymethane-treated A/J mice. Carcinogenesis. 2020;41(7):909-17.

109. Milosavljevic MN, Kostic M, Milovanovic J, Zaric RZ, Stojadinovic M, Jankovic SM, et al. Antimicrobial treatment of Erysipelatoclostridium ramosum invasive infections: a systematic review. Rev Inst Med Trop Sao Paulo. 2021;63:e30.

110. Yan Q, Gu Y, Li X, Yang W, Jia L, Chen C, et al. Alterations of the Gut Microbiome in Hypertension. Frontiers in Cellular and Infection Microbiology. 2017;7(381).

111. Kosowska K, Reinholdt J, Rasmussen LK, Sabat A, Potempa J, Kilian M, et al. The Clostridium ramosum IgA Proteinase Represents a Novel Type of Metalloendopeptidase*. Journal of Biological Chemistry. 2002;277(14):11987-94.

112. Kaper JB, Nataro JP, Mobley HLT. Pathogenic Escherichia coli. Nature Reviews Microbiology. 2004;2(2):123-40.

113. Nougayrède JP, Homburg S, Taieb F, Boury M, Brzuszkiewicz E, Gottschalk G, et al. Escherichia coli induces DNA double-strand breaks in eukaryotic cells. Science. 2006;313(5788):848-51.

114. Lambeth SM, Carson T, Lowe J, Ramaraj T, Leff JW, Luo L, et al. Composition, Diversity and Abundance of Gut Microbiome in Prediabetes and Type 2 Diabetes. J Diabetes Obes. 2015;2(3):1-7.

115. Astbury S, Atallah E, Vijay A, Aithal GP, Grove JI, Valdes AM. Lower gut microbiome diversity and higher abundance of proinflammatory genus Collinsella are associated with biopsy-proven nonalcoholic steatohepatitis. Gut Microbes. 2020;11(3):569-80.

116. Karlsson FH, Fåk F, Nookaew I, Tremaroli V, Fagerberg B, Petranovic D, et al.

Symptomatic atherosclerosis is associated with an altered gut metagenome. Nature Communications. 2012;3(1):1245.

117. Ugarte-Torres A, Gillrie MR, Griener TP, Church DL. Eggerthella lenta Bloodstream Infections Are Associated With Increased Mortality Following Empiric Piperacillin-Tazobactam (TZP) Monotherapy: A Population-based Cohort Study. Clinical Infectious Diseases. 2018;67(2):221-8.

118. Wong D, Aoki F, Rubinstein E. Bacteremia caused by Eggerthella lenta in an elderly man with a gastrointestinal malignancy: A case report. Can J Infect Dis Med Microbiol. 2014;25(5):e85-6.

119. Kutschera M, Engst W, Blaut M, Braune A. Isolation of catechin-converting human intestinal bacteria. J Appl Microbiol. 2011;111(1):165-75.

120. Palm NW, de Zoete MR, Cullen TW, Barry NA, Stefanowski J, Hao L, et al. Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. Cell. 2014;158(5):1000-10.

121. Dey N, Soergel DA, Repo S, Brenner SE. Association of gut microbiota with post-operative clinical course in Crohn's disease. BMC Gastroenterol. 2013;13:131.

122. Zhang H, DiBaise JK, Zuccolo A, Kudrna D, Braidotti M, Yu Y, et al. Human gut microbiota in obesity and after gastric bypass. Proc Natl Acad Sci U S A. 2009;106(7):2365-70.

123. Louis P, Flint HJ. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. FEMS Microbiology Letters. 2009;294(1):1-8.

124. Ferreira-Halder CV, Faria AVdS, Andrade SS. Action and function of Faecalibacterium prausnitzii in health and disease. Best Practice & Research Clinical Gastroenterology. 2017;31(6):643-8.

125. Lam YY, Ha CW, Campbell CR, Mitchell AJ, Dinudom A, Oscarsson J, et al. Increased gut permeability and microbiota change associate with mesenteric fat inflammation and metabolic dysfunction in diet-induced obese mice. PLoS One. 2012;7(3):e34233.

126. Iino T, Mori K, Tanaka K, Suzuki KI, Harayama S. Oscillibacter valericigenes gen. nov., sp. nov., a valerate-producing anaerobic bacterium isolated from the alimentary canal of a Japanese corbicula clam. Int J Syst Evol Microbiol. 2007;57(Pt 8):1840-5.

127. Naseribafrouei A, Hestad K, Avershina E, Sekelja M, Linløkken A, Wilson R, et al. Correlation between the human fecal microbiota and depression. Neurogastroenterol Motil. 2014;26(8):1155-62.

128. Ogita T, Yamamoto Y, Mikami A, Shigemori S, Sato T, Shimosato T. Oral Administration of Flavonifractor plautii Strongly Suppresses Th2 Immune Responses in Mice. Front Immunol. 2020;11:379.

129. Kulecka M, Fraczek B, Mikula M, Zeber-Lubecka N, Karczmarski J, Paziewska A, et al.

The composition and richness of the gut microbiota differentiate the top Polish endurance athletes from sedentary controls. Gut Microbes. 2020;11(5):1374-84.

130. Jin M, Kalainy S, Baskota N, Chiang D, Deehan EC, McDougall C, et al. Faecal microbiota from patients with cirrhosis has a low capacity to ferment non-digestible carbohydrates into short-chain fatty acids. Liver Int. 2019;39(8):1437-47.

131. Takeshita K, Mizuno S, Mikami Y, Sujino T, Saigusa K, Matsuoka K, et al. A Single Species of Clostridium Subcluster XIVa Decreased in Ulcerative Colitis Patients. Inflamm Bowel Dis. 2016;22(12):2802-10.

132. Vacca M, Celano G, Calabrese F, Portincasa P, Gobbetti M, De Angelis M. The Controversial Role of Human Gut Lachnospiraceae. Microorganisms. 2020;8:573.

133. Curtis MM, Hu Z, Klimko C, Narayanan S, Deberardinis R, Sperandio V. The gut commensal Bacteroides thetaiotaomicron exacerbates enteric infection through modification of the metabolic landscape. Cell Host Microbe. 2014;16(6):759-69.

134. Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC, et al. A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. Science. 2003;299(5615):2074-6.

135. Yoshida N, Emoto T, Yamashita T, Watanabe H, Hayashi T, Tabata T, et al. Bacteroides vulgatus and Bacteroides dorei Reduce Gut Microbial Lipopolysaccharide Production and Inhibit Atherosclerosis. Circulation. 2018;138(22):2486-98.

136. Jhangi S, Gandhi R, Glanz B, Cook S, Nejad P, Ward D, et al. Increased Archaea Species and Changes with Therapy in Gut Microbiome of Multiple Sclerosis Subjects (S24.001). Neurology. 2014;82(10 Supplement):S24.001.

137. Maldonado-Gómez MX, Martínez I, Bottacini F, O'Callaghan A, Ventura M, van Sinderen D, et al. Stable Engraftment of Bifidobacterium longum AH1206 in the Human Gut Depends on Individualized Features of the Resident Microbiome. Cell Host Microbe. 2016;20(4):515-26.

138. Palleja A, Mikkelsen KH, Forslund SK, Kashani A, Allin KH, Nielsen T, et al. Recovery of gut microbiota of healthy adults following antibiotic exposure. Nature Microbiology. 2018;3(11):1255-65.

139. Montassier E, Gastinne T, Vangay P, Al-Ghalith GA, Bruley des Varannes S, Massart S, et al. Chemotherapy-driven dysbiosis in the intestinal microbiome. Aliment Pharmacol Ther. 2015;42(5):515-28.

140. Drago L, Toscano M, De Grandi R, Casini V, Pace F. Persisting changes of intestinal microbiota after bowel lavage and colonoscopy. Eur J Gastroenterol Hepatol. 2016;28(5):532-7.

141. Chen Z, Hui PC, Hui M, Yeoh YK, Wong PY, Chan MCW, et al. Impact of Preservation Method and 16S rRNA Hypervariable Region on Gut Microbiota Profiling. mSystems. 2019;4(1):e00271-18.

142. Allali I, Delgado S, Marron PI, Astudillo A, Yeh JJ, Ghazal H, et al. Gut microbiome

compositional and functional differences between tumor and non-tumor adjacent tissues from cohorts from the US and Spain. Gut Microbes. 2015;6(3):161-72.

143. Geng J, Fan H, Tang X, Zhai H, Zhang Z. Diversified pattern of the human colorectal cancer microbiome. Gut Pathog. 2013;5(1):2.

144. Song Se J, Amir A, Metcalf Jessica L, Amato Katherine R, Xu Zhenjiang Z, Humphrey G, et al. Preservation Methods Differ in Fecal Microbiome Stability, Affecting Suitability for Field Studies. mSystems.1(3):e00021-16.

145. Yu J, Feng Q, Wong SH, Zhang D, Liang QY, Qin Y, et al. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. Gut. 2017;66(1):70-8.

146. Löwenmark T, Löfgren-Burström A, Zingmark C, Eklöf V, Dahlberg M, Wai SN, et al. Parvimonas micra as a putative non-invasive faecal biomarker for colorectal cancer. Scientific Reports. 2020;10(1):15250.

147. Eklöf V, Löfgren-Burström A, Zingmark C, Edin S, Larsson P, Karling P, et al. Cancer-associated fecal microbial markers in colorectal cancer detection. Int J Cancer. 2017;141(12):2528-36.

148. Liang Q, Chiu J, Chen Y, Huang Y, Higashimori A, Fang J, et al. Fecal Bacteria Act as Novel Biomarkers for Noninvasive Diagnosis of Colorectal Cancer. Clin Cancer Res. 2017;23(8):2061-70.

149. Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, Earl AM, et al. Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. Genome Res. 2012;22(2):292-8.

150. Tahara T, Yamamoto E, Suzuki H, Maruyama R, Chung W, Garriga J, et al. Fusobacterium in colonic flora and molecular features of colorectal carcinoma. Cancer Res. 2014;74(5):1311-8.

151. Krisanaprakornkit S, Kimball JR, Weinberg A, Darveau RP, Bainbridge BW, Dale BA. Inducible expression of human beta-defensin 2 by Fusobacterium nucleatum in oral epithelial cells: multiple signaling pathways and role of commensal bacteria in innate immunity and the epithelial barrier. Infect Immun. 2000;68(5):2907-15.

152. Uitto VJ, Baillie D, Wu Q, Gendron R, Grenier D, Putnins EE, et al. Fusobacterium nucleatum increases collagenase 3 production and migration of epithelial cells. Infect Immun. 2005;73(2):1171-9.

153. Bachrach G, Ianculovici C, Naor R, Weiss EI. Fluorescence based measurements of Fusobacterium nucleatum coaggregation and of fusobacterial attachment to mammalian cells. FEMS Microbiol Lett. 2005;248(2):235-40.

154. Swidsinski A, Dörffel Y, Loening-Baucke V, Theissig F, Rückert JC, Ismail M, et al. Acute appendicitis is characterised by local invasion with Fusobacterium

nucleatum/<em>necrophorum</em&gt. Gut. 2011;60(1):34.

155. Shah MS, DeSantis TZ, Weinmaier T, McMurdie PJ, Cope JL, Altrichter A, et al. Leveraging sequence-based faecal microbial community survey data to identify a composite biomarker for colorectal cancer. Gut. 2018;67(5):882.

156. Sundqvist G. Taxonomy, ecology, and pathogenicity of the root canal flora. Oral Surgery, Oral Medicine, Oral Pathology. 1994;78(4):522-30.

157. Marchesan J, Jiao Y, Schaff RA, Hao J, Morelli T, Kinney JS, et al. TLR4, NOD1 and NOD2 mediate immune recognition of putative newly identified periodontal pathogens. Mol Oral Microbiol. 2016;31(3):243-58.

158. Liu X, Mao B, Gu J, Wu J, Cui S, Wang G, et al. Blautia-a new functional genus with potential probiotic properties? Gut Microbes. 2021;13(1):1-21.

159. Mori G, Rampelli S, Orena BS, Rengucci C, De Maio G, Barbieri G, et al. Shifts of Faecal Microbiota During Sporadic Colorectal Carcinogenesis. Scientific Reports. 2018;8(1):10329.

160. Inoue R, Ohue-Kitano R, Tsukahara T, Tanaka M, Masuda S, Inoue T, et al. Prediction of functional profiles of gut microbiota from 16S rRNA metagenomic data provides a more robust evaluation of gut dysbiosis occurring in Japanese type 2 diabetic patients. J Clin Biochem Nutr. 2017;61(3):217-21.

161. Chen L, Wang W, Zhou R, Ng SC, Li J, Huang M, et al. Characteristics of fecal and mucosa-associated microbiota in Chinese patients with inflammatory bowel disease. Medicine (Baltimore). 2014;93(8):e51.

162. Kim SG, Becattini S, Moody TU, Shliaha PV, Littmann ER, Seok R, et al. Microbiota-derived lantibiotic restores resistance against vancomycin-resistant Enterococcus. Nature. 2019;572(7771):665-9.

163. Benítez-Páez A, Gómez Del Pugar EM, López-Almela I, Moya-Pérez Á, Codoñer-Franch P, Sanz Y. Depletion of Blautia Species in the Microbiota of Obese Children Relates to Intestinal Inflammation and Metabolic Phenotype Worsening. mSystems. 2020;5(2).

164. Rajilić-Stojanović M, Biagi E, Heilig HG, Kajander K, Kekkonen RA, Tims S, et al. Global and deep molecular analysis of microbiota signatures in fecal samples from patients with irritable bowel syndrome. Gastroenterology. 2011;141(5):1792-801.

165. Nishino K, Nishida A, Inoue R, Kawada Y, Ohno M, Sakai S, et al. Analysis of endoscopic brush samples identified mucosa-associated dysbiosis in inflammatory bowel disease. J Gastroenterol. 2018;53(1):95-106.

166. Luu TH, Michel C, Bard J-M, Dravet F, Nazih H, Bobin-Dubigeon C. Intestinal proportion of Blautia sp. is associated with clinical stage and histoprognostic grade in patients with early-stage breast cancer. Nutrition and cancer. 2017;69(2):267-75.

167. Zhang Y, Yu X, Yu E, Wang N, Cai Q, Shuai Q, et al. Changes in gut microbiota and

plasma inflammatory factors across the stages of colorectal tumorigenesis: a case-control study. BMC Microbiology. 2018;18(1):92.

168. Takada T, Kurakawa T, Tsuji H, Nomoto K. Fusicatenibacter saccharivorans gen. nov., sp. nov., isolated from human faeces. Int J Syst Evol Microbiol. 2013;63(Pt 10):3691-6.

169. Takeshita K, Mizuno S, Mikami Y, Sujino T, Saigusa K, Matsuoka K, et al. A Single Species of Clostridium Subcluster XIVa Decreased in Ulcerative Colitis Patients. Inflammatory Bowel Diseases. 2016;22(12):2802-10.

170. Cummins J, Tangney M. Bacteria and tumours: causative agents or opportunistic inhabitants? Infect Agent Cancer. 2013;8(1):11.

171. Sears CL, Pardoll DM. Perspective: alpha-bugs, their microbial partners, and the link to colon cancer. J Infect Dis. 2011;203(3):306-11.

172. de Carvalho AC, de Mattos Pereira L, Datorre JG, dos Santos W, Berardinelli GN, Matsushita MdM, et al. Microbiota Profile and Impact of Fusobacterium nucleatum in Colorectal Cancer Patients of Barretos Cancer Hospital. Frontiers in Oncology. 2019;9(813).

173. Bonnet M, Buc E, Sauvanet P, Darcha C, Dubois D, Pereira B, et al. Colonization of the human gut by E. coli and colorectal cancer risk. Clin Cancer Res. 2014;20(4):859-67.

174. Xu J, Yang M, Wang D, Zhang S, Yan S, Zhu Y, et al. Alteration of the abundance of Parvimonas micra in the gut along the adenoma-carcinoma sequence. Oncol Lett. 2020;20(4):106.

175. van Dam L, Kuipers EJ, van Leerdam ME. Performance improvements of stool-based screening tests. Best Pract Res Clin Gastroenterol. 2010;24(4):479-92.

176. Tannous B, Lee-Lewandrowski E, Sharples C, Brugge W, Bigatello L, Thompson T, et al. Comparison of conventional guaiac to four immunochemical methods for fecal occult blood testing: implications for clinical practice in hospital and outpatient settings. Clin Chim Acta. 2009;400(1-2):120-2.

177. Dumke J, Vollmer T, Akkermann O, Knabbe C, Dreier J. Case-control study: Determination of potential risk factors for the colonization of healthy volunteers with Streptococcus gallolyticus subsp. gallolyticus. PLoS One. 2017;12(5):e0176515.

178. Tiankanon K, Aniwan S, Rerknimitr R. Current Status of Colorectal Cancer and Its Public Health Burden in Thailand. Clin Endosc. 2021.

179. Aniwan S, Ratanachu Ek T, Pongprasobchai S, Limsrivilai J, Praisontarangkul OA, Pisespongsa P, et al. The Optimal Cut-Off Level of The Fecal Immunochemical Test For Colorectal Cancer Screening in a Country with Limited Colonoscopy Resources: A Multi-Center Study from Thailand. Asian Pac J Cancer Prev. 2017;18(2):405-12.

VITA

NAME	Nutta Iadsee
DATE OF BIRTH	12 November 1995
PLACE OF BIRTH	Bangkok
INSTITUTIONS ATTENDED	Faculty of Science, Khon Kaen University
HOME ADDRESS	11/67 Pak Phraek, Thung Song, Nakhon Si Thammarat