Identification and Characterization of Microbial Signatures and Potential Biomarkers for Irritable Bowel Syndrome in Bangladesh

PhD. Thesis

DEPERTMENT OF MICROBIOLOGY
UNIVERSITY OF DHAKA
DHAKA-1000

March, 2026

SUBMITTED BY
REGISTRATION NO. 10/2021-22
SESSION: 2021-2022

Identification and Characterization of Microbial Signatures and Potential Biomarkers for Irritable Bowel Syndrome in Bangladesh

A DISSERTATION SUBMITTED TO THE UNIVERSITY OF DHAKA IN THE FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPERTMENT OF MICROBIOLOGY
UNIVERSITY OF DHAKA
DHAKA-1000
March, 2026

SUBMTTED BY
EXAMINATION ROLL NO. 10/2021-22
REGISTRATION NO. 10/2021-22
SESSION: 2021-22

Dedicated to...

My Beloved Parents, My lovely wife and kids Who Always Support and Guide Me and Cherish My Life with Their Blessings

Quotation…

'I believe that we have been doing this not primarily to achieve riches or even honour, but rather because we were interested in the work, enjoyed doing it and felt very strongly that it was worthwhile.'

Frederick Sanger

Declaration

I hereby declare that the thesis entitled **"Identification and Characterization of Microbial Signatures and Potential Biomarkers for Irritable Bowel Syndrome in Bangladesh"** has been prepared for submission to the Department of Microbiology, Faculty of Biological Sciences, University of Dhaka, Dhaka-1000, Bangladesh, for partial fulfillment of the requirements for the degree of Doctor of Philosophy. I confirm that I am the sole author of this thesis. To the best of my knowledge, this thesis contains no material previously published by any other person, except where due acknowledgements have been made. Furthermore, this thesis contains no material that has been accepted as part of the requirements for any other academic degree or non-degree program, in English or any other language.

Signature of the Candidate

Certification

It is hereby certified that the student bearing Reg. No. 10/2021-22, Session-2021-22 has carried out the research work entitled "**Identification and Characterization of Microbial Signatures and Potential Biomarkers for Irritable Bowel Syndrome in Bangladesh**" for the partial fulfillment of his PhD Degree from University of Dhaka, Bangladesh, under our academic supervision and co-supervision at the Department of Microbiology, University of Dhaka.

Signature of Supervisor

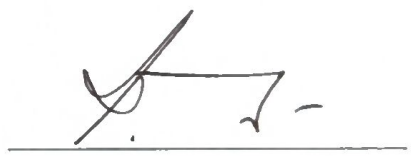
Dr. Donald James Gomes

Professor,

Department of Microbiology

University of Dhaka

Dhaka-1000, Bangladesh



Signature of Co-supervisor

Dr. Md. Mizanur Rahaman

Associate Professor,

Department of Microbiology

University of Dhaka

Dhaka-1000, Bangladesh

Acknowledgement

Foremost, I want to express my gratitude to Almighty Allah for giving me the strength and ability to conduct the research and complete it satisfactorily. Without His blessings, the completion of this task would not have been possible.

I want to sincerely thank my supervisor Professor Dr. Donald James Gomes, Dept. of Microbiology, University of Dhaka, and my co-supervisor Dr. Md. Mizanur Rahaman, Associate Professor, Dept. of Microbiology, University of Dhaka, to whom I am forever indebted for allowing me to work on a wonderful topic that introduced me to the fascinating world of metagenomics. I am grateful for receiving continuous support, motivation, and scholastic guidance throughout the project.

Besides my advisor, I owe my sincere Gratitude and thanks to Professor Dr. Shakila Nargis Khan, Chairman, Dept. of Microbiology, University of Dhaka for her kind cooperation and support during this research work.

I would like to express my heartfelt gratitude to Mr. Spencer Mark Mondol Lecturer, Department of Microbiology, University of Dhaka and Md. Rafiul Islam Ranga, Assistant Professor, Department of Microbiology, University of Dhaka, for their words of encouragement, guidance, unconditional support and unreserved help served as a beacon light throughout the course of study, research work and completion of this manuscript.

I am extremely grateful to, Md. Hasanul Banna Siam, Afroza Khan Huma, Md. Azizul Hoque, Md. Atiq Abrar Rahman, Fardin Farhad, Israt Islam, Nadira Naznin Rakhi, and for their immense help and support to get the work done in time. My cordial gratitude is also to Dr. Md. Kazi Alamgir Hossain, Md. Mahmuduzaman Rajib, Dipan Barai, Md. Raihan Reza, Md. Rakibul Islam, Md. Payel hasan, Md. Robiul Islam for helping me out when things got difficult for me. Finally, I would like to thank Foysal Ahmed, Md. Riyadul Islam, Sozib vai, our sincere laboratory assistant who landed a helping hand during my research work. Thanks to all the staffs of Department of Microbiology for their generous help through this work.

Last but not the least, I would like to thank my family for their love & inspiration and for supporting me spiritually throughout my life.

Author

March, 2026

Abstract

Introduction: Irritable Bowel Syndrome (IBS) is a functional gastrointestinal disorder increasingly linked to gut microbiome dysbiosis. However, metagenomic studies integrating taxonomic, functional, resistome, virulome, and genome-resolved analyses in Bangladeshi IBS patients are limited. This study applied whole-metagenome sequencing (WMS), comparative genomics, and machine-learning (ML) approaches to characterize gut microbiome alterations in IBS patients relative to healthy controls (HC).

Materials and Methods: A total of thirty (n=30) stool samples were collected from IBS-diagnosed patients (n=20) and HC (n=10) at the National Gastroliver Institute & Hospital, Mohakhali, Dhaka, Bangladesh. Of these, based on the quality and concentration of extracted total genomic DNA, ten IBS samples (n=10) and six HC samples (n=6) were selected for WMS. Sequencing was performed using the Illumina Next Generation Sequencing (MiSeq 4000) platform. The raw FASTQ sequences were subjected to quality control using the Trimmomatic tool. Downstream analyses were performed using two complementary pipelines: the Chan Zuckerberg ID (CZID, formerly IDseq) for host filtering, taxonomic profiling, diversity metrics, and assembly, as well as the Kraken2-based workflow for antimicrobial resistance (AMR) and virulence factor genes detection. Metagenome Assembled Genomes (MAGs) were reconstructed using MEGAHIT, MetaBAT2, and MaxBin2. Comparative global analysis involved a total of 1,109 samples (HC=492; IBS=617), comprising 1,093 publicly available gut metagenomes from across all continents, along with the Bangladeshi samples (n=16). ML models were trained on combined global and local taxonomic datasets after normalization, feature selection, and hyperparameter optimization.

Results: IBS patients exhibited clear dysbiosis in their gut microbiome characterized by a marked reduction of Actinobacteria, especially *Bifidobacterium* spp., and increased abundance of Firmicutes and pathobionts such as *Klebsiella pneumoniae*, *Enterobacter cloacae*, and *Sutterella wadsworthensis*. Notably, multiple *Bifidobacterium* species, including *B. breve, B. bifidum, B. catenulatum, B. pseudocatenulatum* and *B. longum* subspecies, were significantly enriched in the HC samples compared to IBS patients. AMR profiling revealed a relatively higher prevalence of tetracycline, macrolide, and β-lactam resistance genes in IBS samples. MAG analysis recovered 392 genomes, with IBS-derived MAGs harboring more AMR and virulence genes. Comparative

analysis of global datasets, including Bangladeshi samples, revealed significantly lower alpha diversity in IBS patients supported with Chao1 (p-value: 2.6226e-05), ACE (p-value: 1.8209e-05), and Observed (p-value: 3.0546e-05) indices, indicating higher species richness in HC. Beta diversity analysis showed clear and distinct clustering of IBS and HC samples, reflecting significant differences in overall community composition. LEfSe analysis (LDA score > 2.5, $p < 0.05$, FDR-adjusted) identified distinct microbial signatures between IBS patients and HC groups, with healthy individuals enriched in *Bacteroides* spp. (*B. stercoris*, *B. uniformis*) and *Bifidobacterium catenulatum*. In contrast, IBS patients showed increased abundances of *Klebsiella*, *Roseburia*, *Faecalibacterium*, and uniquely associated taxa including *Collinsella aerofaciens*, *Klebsiella pneumoniae*, *Streptococcus salivarius*, and *Clostridium* spp. Among the developed ML models, Logistic Regression demonstrated superior IBS prediction performance with an accuracy of 0.80 (ROC-AUC 0.88).

Conclusion: This multilayered metagenomic investigation revealed substantial microbial, functional, and genomic disruptions in Bangladeshi IBS patients, consistent with patterns observed in global datasets. The findings underscore the potential of microbiome-based ML models in early diagnosis and targeted therapeutics for IBS. Nevertheless, further studies with larger, multi-center, and longitudinal cohorts are warranted to validate these findings and resolve population- and subtype-specific microbiome signatures in IBS.

Contents

List of Tables

List of Figures

List of abbreviations

Chan Zuckerberg Infectious Disease (CZID)

Clostridium difficile infection (CDI)

Fecal microbial transplant (FMT)

Fermentable oligosaccharides, disaccharides, monosaccharides, and polyols (FODMAPs)

Healthy Control (HC)

High fat intake diet (HFD)

Institutional Review Board (IRB)

Interquartile range (IQR)

Intestinal epithelial cells (IEC)

Inflammatory Bowel Disease (IBD)

Irritable bowel syndrome (IBS)

Lipopolysaccharide (LPS)

Metagenomically assembled genomes (MAGs)

Muramyl dipeptide (MDP)

Multidrug resistant (MDR)

Myeloid-Derived Suppressor Cells (MDSC)

Natural killer T cells (NKT)

Nod-like receptor (NLR)

Non-metric multidimensional scaling (NMDS)

Nucleotide-binding oligomerization domain 2 (NOD2)

Pathogen associated molecular pattern (PAMP)

Polyunsaturated fatty acids (PUFA)

Principle coordinate analysis (PCoA)

Short-chain fatty acids (SCFA)

Toll-like receptor (TLR)

Chapter 01

Introduction and Literature Review

1.1 General Introduction

The gut microbiome, comprising a vast ecosystem of trillions of microorganisms residing within the human gastrointestinal tract, plays a pivotal role in maintaining host health by actively regulating digestion, modulating immune responses, and influencing neurological processes through the production of bioactive metabolites, including short-chain fatty acids, bile acids, and neurotransmitter precursors (Qin et al., 2010; Nicholson et al., 2012). This microbial genomic repertoire is often referred to as a "second genome", which exceeds the human genome in both size and complexity. This enables functions that human genome cannot provide alone. When balanced, the gut microbiome supports resilience, nutrient absorption, and immune tolerance. In contrast, disruption of this equilibrium, known as 'dysbiosis', can trigger cascades of inflammation, metabolic dysfunction, and gastrointestinal disorders (Mostafavi Abdolmaleky & Zhou, 2024).

Irritable Bowel Syndrome (IBS) is one of the clearest examples of how significant this homeostasis can be. IBS is a functional gastrointestinal disorder characterized by recurrent abdominal pain and altered bowel habits in the absence of structural abnormalities (Chey et al., 2015; Lovell & Ford, 2012). It affects 10–15% of population worldwide. Its burden is not limited to physical symptoms. Patients with IBS often experience impaired quality of life, reduced productivity, and psychological distress. The gut-brain axis is a bidirectional communication system linking the central nervous system and the enteric nervous system and plays a pivotal role in symptom generation. Stress and anxiety can amplify gastrointestinal disturbances and microbial imbalance, which in turn influence mood and cognition (Mayer et al., 2015a). This interplay highlights the multidimensional nature of IBS, which is not merely a gastrointestinal condition but a disorder that bridges biology, psychology, and social experience.

Global research has revealed that IBS patients consistently exhibit distinct microbiome signatures compared to healthy controls, although the specific alterations vary across regions. European IBS cohorts often show reduced abundance of beneficial taxa such as *Bifidobacterium* and *Faecalibacterium prausnitzii* but increases in pro-inflammatory *Enterobacteriaceae* (Ahmadi et al., 2025). However, different patterns are observed in Asian populations. Several studies from India reported a higher prevalence of *Prevotella* species consistent with carbohydrate-rich diets.

On the other hand, Japanese cohorts exhibited reductions in *Lactobacillus* and *Bifidobacterium* and increases in *Streptococcus* (Kassinen et al., 2007). Some African studies suggest that rural populations consuming high-fiber diets exhibit greater microbial diversity, which is potentially protective against IBS. Whereas the urbanized groups show reduced diversity and higher prevalence of IBS (Brewster et al., 2019). Collectively, these findings denote the critical role of diet, environment, and culture in shaping IBS-associated gut microbiome dysbiosis.

In Bangladesh, the traditional diet is dominated by rice, lentils, fish etc. Here, the spices such as turmeric and chili are widely consumed. This may further modulate microbial activity through antimicrobial and anti-inflammatory properties. But environmental challenges including poor sanitation and contaminated water supplies usually expose the individuals of Bangladesh to recurrent enteric infections that disrupt microbial balance (Shireen et al., 2017). Frequent antibiotic use in both clinical and agricultural settings also plays a role in this disruption, which reduces diversity and promotes dysbiosis (Scutari et al., 2020). The socioeconomic factors may add another layer of complexity. Limited awareness of functional gastrointestinal disorders, dependence on traditional remedies, and complex healthcare infrastructure restrict the access to advanced diagnostics such as microbiome sequencing (Bakhshi et al., 2014). Despite these types of influences, systematic investigations into IBS-associated microbiome alterations in Bangladesh remained unaddressed. This ignorance has left a critical gap in both local healthcare and global understanding.

To address this gap, advances in sequencing technologies have opened new possibilities. Traditional culture-based methods are not suitable, as cultivating anaerobic gut microbes is very challenging. On the other hand, 16S rRNA sequencing can provide taxonomic insights but it lacks proper identification at the species and strain level (Dethlefsen & Relman, 2011); (Michael & L., 2007). Whole-genome shotgun metagenomics has emerged as a transformative approach. It enables comprehensive characterization of microbial communities by sequencing all genetic material present in a sample (Qin et al., 2010). This method allows identification of microbial taxa at fine resolution. This method also offers functional profiling of metabolic pathways and detection of antibiotic resistance genes and virulence factors (Segata et al., 2012). Integration with other "omics" approaches metatranscriptomics, metabolomics, and proteomics, provides multidimensional insights into host-microbe interactions. (Integrative HMP, 2019; Sanches et al.,

2024). In IBS research, metagenomics has revealed consistent functional shifts such as reduced butyrate-producing pathways and increased methane-related genes, which correlate with clinical symptoms (Pozuelo et al., 2015).

Therapeutic strategies targeting the microbiome have shown varying degrees of promise. Probiotics such as *Bifidobacterium infantis* and *Lactobacillus plantarum* have shown promise in reducing IBS symptoms by restoring microbial balance (Didari et al., 2015). Prebiotics, which are usually known as nondigestible fibers that stimulate the growth of beneficial bacteria, have demonstrated efficacy in improving stool regularity and microbial diversity (Malfertheiner & Delchier, 2014). Fecal microbiota transplantation (FMT) is still in experimental stage. But this has shown symptom improvement in some IBS patients. It has shown improved symptoms particularly those with diarrhea-predominant IBS. Though in case of FMT, long-term safety and standardization remains as a concern (Johnsen et al., 2018). Dietary interventions such as the low FODMAP diet have proven effective in reducing symptoms by limiting substrates for gas-producing bacteria. (O'Halloran et al., 2017). Personalized medicine approaches hold promise for developing therapies for individual microbial imbalances (Kamboj et al., 2018). This approach integrates microbiome profiling with clinical and dietary data.

But there are still limitations with the methods. There are differences in how samples are collected and how DNA is extracted. There are also differences in sequencing platforms and bioinformatics pipelines. This makes it complicated to compare studies and reproduce results. The heterogeneity of IBS, marked by differences in subtype, symptom severity, and comorbidities, complicates the identification of consistent microbial signatures (Holowatyj et al., 2020). Diet, medication use, and psychosocial stress are several important factors that must be carefully controlled because they have a significant effect on the composition of the microbiome. To solve these problems, we need to develop standardized protocols, bigger and more diverse groups of people, and the use of multi-omics methods to better understand how host-microbe interactions work.

Against these backdrops, the present study aims to provide the first whole-genome metagenomic insight into IBS-associated dysbiosis in Bangladeshi patients. By sequencing stool samples from IBS patients and healthy controls, and comparing them with global datasets, we have targeted to identify microbial and functional signatures unique to this population. Our objectives are: first, to characterize the microbial composition of IBS patients in Bangladesh; second, to identify

functional pathways and metabolites that distinguish IBS-associated dysbiosis; third, to compare Bangladeshi microbiome signatures with international cohorts to highlight regional differences; and finally, to apply a machine learning approach to develop a predictive model for gut microbiome-based IBS diagnosis. Through this work, we aim not only to advance scientific understanding of IBS but also to provide insights that resonate with the lived realities of Bangladeshi patients, ensuring that microbiome research is both globally relevant and locally meaningful.

1.2 Review of Literature

1.2.1 Human Microbiota and Its Components

There are many distinct types of microorganisms that exist in different regions of the body, such as bacteria, archaea, fungi, viruses, and eukaryotes. The gut is home to the biggest and most complicated group (Bull & Plummer, 2014). When you look at its microbial genomes and metabolites, this ecosystem is commonly dubbed the microbiome. It acts like a "forgotten organ" or "meta-organism" that changes with the host to affect metabolism, immunity, and overall balance. There are trillions of microorganisms in the gut microbiome, and it has more than 3 million genes, which is a lot more than the human genome (Zhu et al., 2010). It makes bioactive substances such as short-chain fatty acids (SCFAs), bile acids (BAs), and tryptophan derivatives. More than 90% of this colony is made up of bacteria (Fusco et al., 2023). Archaea, methanogens, fungi like *Candida* and *Saccharomyces*, and viruses like bacteriophages all help keep everything in balance and strong. Microbial metabolites including acetate, propionate, and butyrate are important parts that help host-microbe interactions (Jyoti & Dey, 2025). Other things, such as polyamines, hydrogen sulfide, and indoles, help control how the immune system works and how well the barrier works. The microbiota helps in digestion, getting nutrients out of food, modulating the immune system, fighting off pathogens, and breaking down drugs (Colella et al., 2023). It can be affected by things like the host and the environment. Dysbiosis, or imbalance, throws this balance off and makes inflammation and disease more likely. In general, the gut microbiome is thought of as a "second genome" that changes features in ways that host genetics can't.

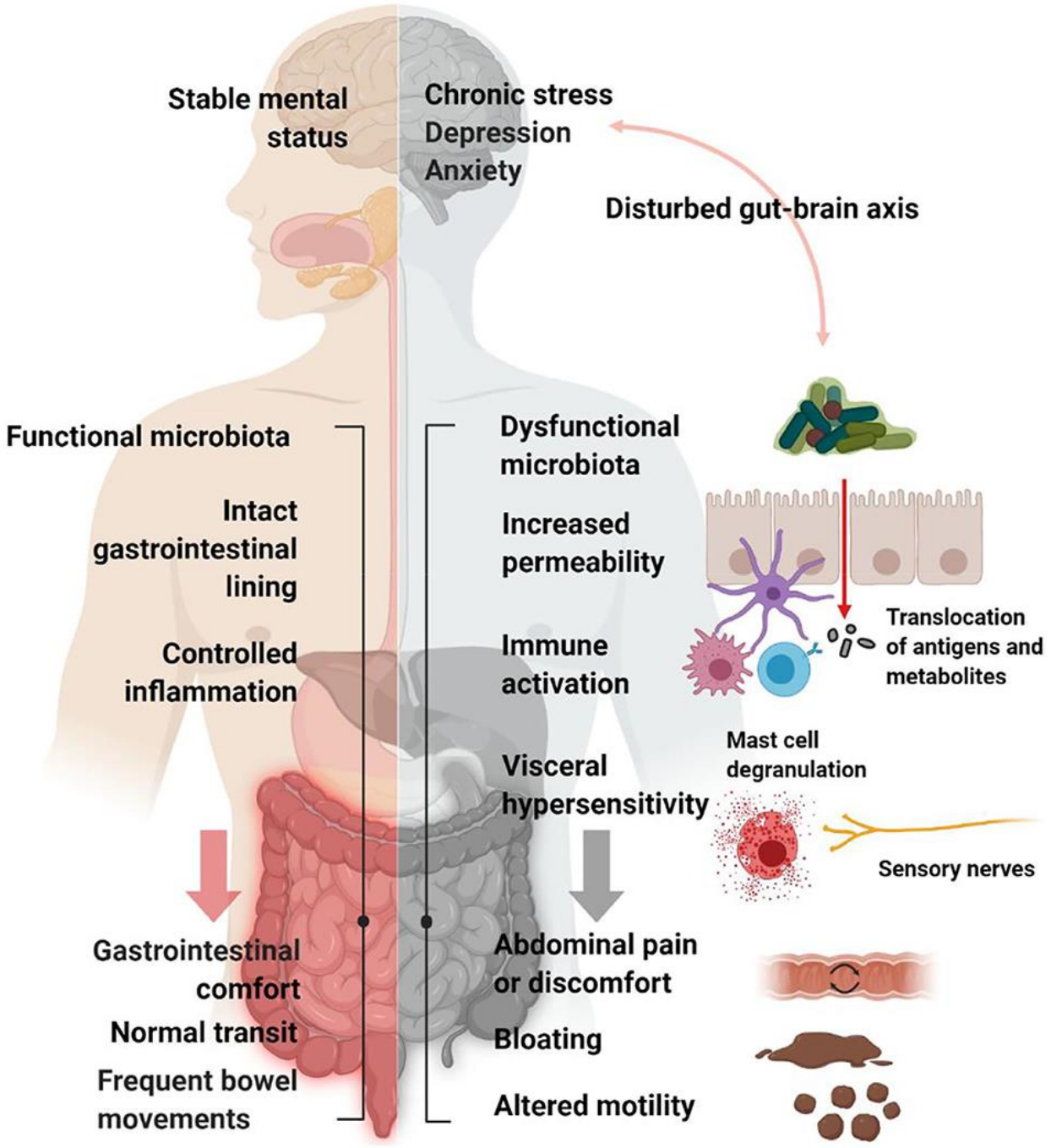


Figure. 1.1 Impact of Gut Microbiota Functionality on Gastrointestinal and Mental Health via the Gut-Brain Axis. Comparison of functional versus dysfunctional gut microbiota illustrating effects on intestinal integrity, immune activation, and gut-brain axis signaling, with implications for gastrointestinal symptoms and mental health outcomes (Carco et al., 2020).

1.2.2 Gut Microbiome: Composition, Function, and Health Implications

The human gut microbiome is a complex ecosystem composed of trillions of microorganisms, including bacteria, archaea, fungi, and viruses, which collectively outnumber human cells and contribute significantly to host physiology (Qin et al., 2010). The diversity of microbes in the gastrointestinal (GI) tract changes across different regions and over time. Among all sections, the colon hosts the greatest number and variety of microorganisms around 10^{11} to 10^{12} cells per gram of fecal content. This rich diversity is supported by the colon's oxygen-free (anaerobic) environment, neutral pH, and slow movement of material through it.

Bacteria dominate this community, with Firmicutes and Bacteroidetes being the most abundant phyla, followed by Actinobacteria and Proteobacteria (Sekirov et al., 2010). These microbial populations vary across individuals, influenced by genetics, diet, environment, and geography (Turnbaugh et al., 2009). Diversity gradients exist along the GI tract: the esophagus and stomach have lower diversity dominated by *Helicobacter* and *Lactobacillus*, the duodenum features *Prevotella* for carbohydrate degradation, the ileum enriches *Enterococcus* and *Bacteroides* for bile acid circulation, and the colon favors anaerobic fermenters like *Parabacteroides* and *Bifidobacterium*. Reduced diversity is linked to diseases such as inflammatory bowel disease (IBD), obesity, and diabetes, while higher diversity correlates with better immune modulation and metabolic health. Population-specific variations are notable, with higher diversity in non-Western (e.g., African, Asian) populations due to fiber-rich diets, contrasting with lower diversity in urbanized Western groups.

Functionally, the gut microbiome plays a central role in digestion and nutrient metabolism. It aids in the breakdown of complex polysaccharides, producing short-chain fatty acids (SCFAs) such as acetate, propionate, and butyrate, which serve as energy sources for colonocytes and regulate immune responses (Flint et al., 2012). The microbiome also synthesizes essential vitamins, including vitamin K and B-group vitamins, and contributes to bile acid metabolism (Pham et al., 2021). Beyond metabolism, the microbiome is crucial for immune system development, with microbial antigens shaping mucosal immunity and tolerance (Belkaid & Hand, 2014).

The gut microbiome also interacts with the central nervous system through the gut-brain axis, influencing mood, cognition, and behavior **(Figure 1.1)** (Cryan & Dinan, 2012). Neuroactive metabolites such as serotonin precursors and gamma-aminobutyric acid (GABA) are modulated

by microbial activity, linking gut health to psychological well-being **(Figure 1.2)** (Cryan & Dinan, 2012). Dysregulation of this axis has been implicated in functional gastrointestinal disorders, including IBS, as well as psychiatric conditions such as anxiety and depression (Mayer et al., 2015a).

Several factors influence microbiome composition. Diet is a major determinant, with fiber-rich diets promoting SCFA-producing bacteria, while high-fat, high-sugar diets favor pro-inflammatory taxa (David et al., 2014). Antibiotic use can cause profound disruptions, reducing microbial diversity and enabling opportunistic pathogens to thrive (Dethlefsen & Relman, 2011). Geographic and cultural differences also shape microbiome profiles; for example, rural African populations consuming plant-based diets exhibit higher microbial diversity compared to Western populations (De Filippo et al., 2010).

A healthy microbiome is characterized by diversity, stability, and resilience, whereas dysbiosis, an imbalance in microbial composition, is associated with a range of diseases, including IBS, obesity, diabetes, and inflammatory bowel disease (IBD) (Shreiner et al., 2015). Thus, understanding the composition and functions of the gut microbiome is essential for elucidating its role in health and disease, and for developing diagnostic tools and microbiome-targeted therapies.

1.2.3 Irritable Bowel Syndrome (IBS): Clinical and Epidemiological Perspectives

Irritable Bowel Syndrome (IBS) is a functional gastrointestinal disorder characterized by recurrent abdominal pain associated with altered bowel habits, in the absence of identifiable structural abnormalities (Chey et al., 2015). The Rome IV criteria, established in 2016, remain the gold standard for diagnosis, emphasizing symptom-based classification rather than exclusion of organic disease (Drossman, 2016) (Lacy & Patel, 2017). IBS is typically divided into four subtypes based on predominant stool pattern: constipation-predominant (IBS-C), diarrhea-predominant (IBS-D), mixed (IBS-M), and unclassified (IBS-U) (Lacy & Patel, 2017). Globally, IBS affects approximately 10–15% of the population, though prevalence varies widely depending on diagnostic criteria and geographic region (Lovell & Ford, 2012). Studies in North America and Europe report prevalence rates between 12–15% (Chan & Williams, 2020), while Asian countries often show lower rates, ranging from 5–10% (Veitch et al., 2015). According to community-based surveys, the prevalence of IBS in Bangladesh is estimated to be between 7 to 8%, with greater rates seen in women and younger adults (Bakhshi et al., 2014). Cultural factors, dietary patterns,

and healthcare-seeking behavior may contribute to these regional differences and it challenges in diagnosis and management of IBS in Bangladesh (Mannan et al., 2024).

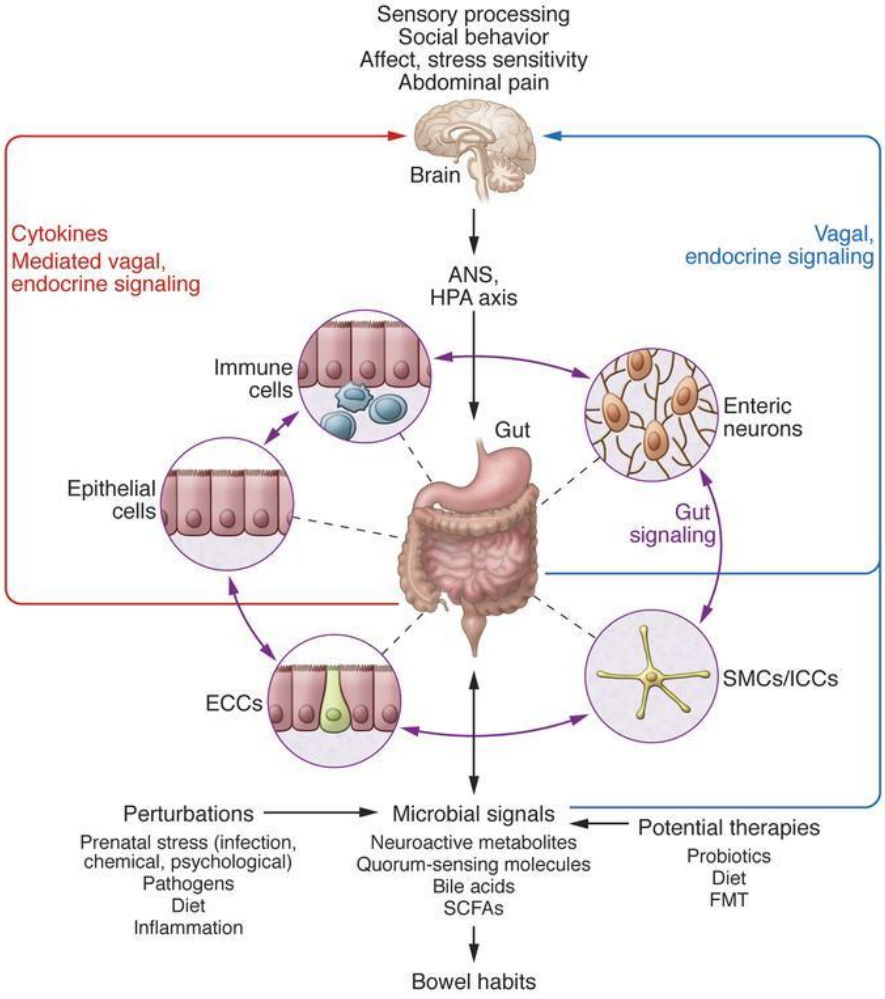


Figure 1.2 Cellular and Microbial Mediators of Gut-Brain Axis Signaling and Therapeutic Modulation. Diagram illustrating bidirectional communication between the gut and brain via immune, epithelial, neuronal, and endocrine pathways, influenced by microbial metabolites and environmental perturbations, with potential therapeutic interventions targeting bowel habits and neuroimmune regulation (Mayer et al., 2015b).

The pathophysiology of IBS is multifactorial, involving disturbances in gut motility, visceral hypersensitivity, immune activation, and psychosocial stressors (Brage et al., 2013). Recent research highlights the role of the gut-brain axis, where bidirectional communication between the central nervous system and enteric nervous system influences symptom generation (Mayer et al., 2015a). Low-grade mucosal inflammation and altered intestinal permeability have also been

implicated, suggesting that IBS may represent a spectrum of immune-mediated dysfunction rather than a purely functional disorder (Spiller & Major, 2016).

IBS imposes a significant burden on healthcare systems, with patients frequently undergoing repeated diagnostic tests and treatments (Canavan, 2014). Quality of life is often impaired, with symptoms affecting work productivity, social functioning, and psychological well-being (Gralnek et al., 2000). In Bangladesh, limited awareness and stigma surrounding gastrointestinal disorders further complicate diagnosis and management. This underscores the need for region-specific research to better understand IBS epidemiology and its unique manifestations in South Asian populations.

1.2.4 Dysbiosis and IBS: Linking Microbiome Alterations to Disease

Dysbiosis, defined as an imbalance in the composition and function of the gut microbiota, has emerged as a central mechanism in the pathophysiology of IBS (Shreiner et al., 2015). Several studies have demonstrated that IBS patients exhibit reduced microbial diversity compared to healthy controls, with alterations in the relative abundance of key bacterial taxa (Jeffery et al., 2012). Specifically, reductions in beneficial genera such as *Bifidobacterium* and *Lactobacillus* have been consistently reported, alongside increases in potentially pathogenic taxa such as *Enterobacteriaceae* (Rajilić–Stojanović et al., 2011).

Functional consequences of dysbiosis in IBS include altered fermentation processes and short-chain fatty acid (SCFA) production. Butyrate, a key SCFA that supports colonic epithelial health, is often reduced in IBS patients, contributing to impaired barrier function and low-grade inflammation (Moran, 2017). Dysbiosis also affects bile acid metabolism, with increased deconjugation and altered signaling contributing to diarrhea-predominant IBS (IBS-D) (Oakland et al., 2019).

Immune dysregulation is another hallmark of IBS-associated dysbiosis. Studies have shown that microbial imbalance can trigger mucosal immune activation, leading to increased levels of pro-inflammatory cytokines such as IL-6 and TNF-α (Barbara et al., 2011). This low-grade inflammation may sensitize visceral afferent pathways, contributing to abdominal pain and discomfort (Barbara et al., 2011). Moreover, dysbiosis has been linked to increased intestinal

permeability, often referred to as “leaky gut,” which facilitates translocation of microbial products and further immune activation (Macura et al., 2024).

The gut-brain axis provides another pathway through which dysbiosis influences IBS symptoms. Microbial metabolites, including tryptophan derivatives and GABA, modulate central nervous system activity, and dysbiosis may disrupt these signaling pathways, contributing to anxiety and depression frequently observed in IBS patients (Mayer et al., 2015a). This highlights the bidirectional relationship between psychological stress and microbial imbalance.

Comparative studies have revealed that dysbiosis patterns differ between IBS and other gastrointestinal disorders such as inflammatory bowel disease (IBD). While IBD is characterized by profound microbial depletion and dominance of pro-inflammatory taxa, IBS shows subtler but consistent shifts in microbial composition (Khan et al., 2019). Importantly, regional differences exist: Asian IBS patients often show distinct microbial signatures compared to Western cohorts, likely reflecting dietary and cultural influences (J. Li et al., 2024).

Taken together, these findings underscore that IBS is not merely a functional disorder but involves measurable alterations in gut microbial ecology. Hence, understanding these dysbiosis patterns is essential for developing microbiome-targeted therapies, including probiotics, prebiotics, and fecal microbiota transplantation.

1.2.5 Metagenomic Approaches to Studying the Gut Microbiome

Traditional microbiological methods relied heavily on culture-based techniques, which were limited because most gut microbes are anaerobic and difficult to cultivate (Zoetendal et al., 2008). The advent of molecular approaches, particularly 16S rRNA gene sequencing, revolutionized microbiome research by enabling taxonomic profiling of microbial communities without the need for cultivation (Turnbaugh et al., 2007). However, while 16S sequencing provides insights into microbial composition, it lacks resolution at the species and strain level and offers limited functional information (Michael & L., 2007).

Whole-genome shotgun metagenomics has emerged as a powerful alternative, allowing comprehensive characterization of microbial communities by sequencing all genetic material present in a sample (Qin et al., 2010). This approach enables identification of microbial taxa at the species and strain level, as well as functional profiling of metabolic pathways, antibiotic resistance

genes, and virulence factors (Sharpton, 2014). Metagenomics thus provides a more holistic view of the microbiome, linking composition to function.

Bioinformatics pipelines are central to metagenomic analysis. Instruments like MetaPhlAn, HUMAnN, and MEGAHIT enable taxonomic classification, functional annotation, and genome assembly from intricate datasets (Segata et al., 2012; Franzosa et al., 2018). Improvements in computational techniques have facilitated strain-level resolution, uncovering intra-species variability that could be pivotal in the pathophysiology of IBS (Nayfach et al., 2016). Moreover, integration of metagenomics with other "omics" approaches, such as metatranscriptomics, metabolomics, and proteomics, allows multi-dimensional insights into microbial activity and host-microbe interactions (Knight et al., 2018).

In IBS research, metagenomics has revealed consistent alterations in microbial functional capacity. For example, studies have shown reduced genes involved in butyrate synthesis and increased pathways related to methane production in IBS patients (Pozuelo et al., 2015). These functional shifts correlate with clinical symptoms such as diarrhea, constipation, and bloating, underscoring the importance of functional rather than purely compositional analyses.

Despite its advantages, metagenomics faces challenges, including high costs, computational complexity, and difficulties in distinguishing live from dead microbes (Quince et al., 2017). Additionally, variability in sample collection, DNA extraction methods, and sequencing platforms can introduce biases, complicating cross-study comparisons (Forry et al., 2024). Nevertheless, whole-genome metagenomics remains the most comprehensive approach for studying the gut microbiome and is particularly relevant for IBS research in diverse populations such as Bangladesh, where dietary and environmental factors may shape unique microbial signatures.

1.2.6 Global Evidence on Gut Microbiome in IBS

Research across diverse populations has consistently demonstrated that IBS patients exhibit distinct gut microbiome signatures compared to healthy controls, though the specific alterations vary geographically (Chan & Williams, 2020). In European cohorts, IBS has been associated with reduced abundance of *Bifidobacterium* and *Faecalibacterium prausnitzii*, both considered beneficial taxa due to their anti-inflammatory properties (Jeffery et al., 2012). Conversely,

increases in Enterobacteriaceae and *Clostridium* species have been reported, suggesting a shift toward pro-inflammatory microbial communities (Rajilić–Stojanović et al., 2011).

North American studies have highlighted functional changes in the microbiome of IBS patients, including reduced genes involved in butyrate synthesis and increased pathways related to methane production, which correlate with constipation-predominant IBS (IBS-C) (Pozuelo et al., 2015). These findings emphasize that microbial functional capacity, rather than composition alone, plays a critical role in symptom generation (Tap et al., 2017). In Asian populations, IBS microbiome profiles differ from Western cohorts, likely reflecting dietary and cultural influences. Indian studies indicate a greater incidence of Prevotella species in IBS patients, aligning with the carbohydrate-dense diets prevalent in South Asia (Jung & Myung, 2023). Similarly, Japanese studies have identified reductions in Lactobacillus and *Bifidobacterium*, alongside increases in *Streptococcus* spp., highlighting regional variability in dysbiosis patterns (Kassinen et al., 2007). African studies, though limited, suggest that rural populations consuming high-fiber diets exhibit greater microbial diversity, which may protect against IBS development (De Filippo et al., 2010). This contrasts with urban African cohorts, where Westernized diets are associated with reduced diversity and increased IBS prevalence (O'Keefe et al., 2015). Such findings underscore the importance of diet and environment in shaping IBS microbiome signatures. Meta-analyses have attempted to synthesize global evidence, revealing consistent reductions in microbial diversity and beneficial taxa across IBS patients, but also highlighting heterogeneity between studies due to methodological differences and geographic variability (Lamb et al., 2019). These global insights emphasize the need for region-specific research, particularly in underrepresented populations such as Bangladesh, where unique dietary and environmental factors may shape distinct microbiome patterns.

1.2.7 In the Context of Bangladesh: Gut Microbiome and IBS

Although global research on the gut microbiome in IBS has expanded rapidly, studies from Bangladesh remain limited. Bangladesh presents a unique context due to its dietary patterns, environmental exposures, and socioeconomic conditions, all of which strongly influence gut microbial composition (Bakhshi et al., 2014). Dietary patterns and nutritional status in Bangladesh: implications for gut health (Fahim et al., 2021). High consumption of spices such as turmeric and

chili may also modulate gut microbial activity, given their antimicrobial and anti-inflammatory properties (Compaoré et al., 2016).

Environmental factors play a significant role in shaping the gut microbiome in Bangladesh. Widespread exposure to enteric pathogens due to poor sanitation and contaminated water supplies contributes to recurrent gastrointestinal infections, which can disrupt microbial balance and predispose individuals to IBS (Shireen et al., 2017). Moreover, frequent use of antibiotics in both clinical and agricultural settings has been shown to reduce microbial diversity and promote dysbiosis (Batuman et al., 2024). These factors suggest that IBS patients in Bangladesh may exhibit distinct microbiome signatures compared to Western populations. A study compared the gut microbiome of healthy children aged 9–14 living in Bangladesh and the United States. It found that Bangladeshi children had greater bacterial diversity, distinct microbial composition, and less temporal stability compared to U.S. children (Lin et al., 2013). Another article explores how ethnicity influences gut microbiome composition in Bangladesh. The study compared the gut microbiota of Bengali people with several Indigenous ethnic groups (Chakma, Marma, Khyang, and Tripura) living in the Chittagong Hill Tracts. Researchers found distinct microbial signatures between these populations, showing that ethnicity and lifestyle strongly shape gut microbiome diversity and structure (Ahammad et al., 2024). Another significant article investigated how SARS-CoV-2 infection disrupts the oral and gut microbiomes in Bangladeshi patients. The study found that COVID-19 patients had reduced microbial diversity and richness compared to healthy controls, with a notable increase in opportunistic pathogens such as *Escherichia*, *Shigella*, *Bacteroides*, and *Streptococcus*. Concurrently, advantageous genera like as Prevotella were markedly diminished (Rafiqul Islam et al., 2022).

Socioeconomic conditions further complicate IBS management in Bangladesh. Limited awareness of functional gastrointestinal disorders, coupled with stigma surrounding digestive symptoms, often delays diagnosis and treatment. Patients frequently rely on traditional remedies or over-the-counter medications, which may alter gut microbial composition but remain poorly studied (Lee et al., 2020). Additionally, healthcare infrastructure constraints limit access to advanced diagnostic tools, including microbiome sequencing, underscoring the need for region-specific research using modern metagenomic approaches. Notwithstanding these obstacles, recent studies from South Asia offer significant insights. Indian research indicates that patients with IBS frequently exhibit

a higher prevalence of Prevotella species and a decreased presence of Bifidobacterium, highlighting dietary impacts (Ghoshal et al., 2023). Similar patterns are expected in Bangladesh, though systematic investigations are lacking. The application of whole-genome metagenomics in Bangladeshi IBS patients could therefore reveal novel microbial and functional signatures, contributing to global understanding while addressing local healthcare needs.

In summary, Bangladesh represents a critical but underexplored setting for IBS microbiome research. Dietary behaviors, environmental factors, and healthcare constraints generate a distinct microbial ecosystem that need exploration via thorough metagenomic methodologies. This thesis aims to fill that gap by providing the first whole-genome metagenomic analysis of IBS-associated dysbiosis in Bangladeshi patients.

1.2.8 Therapeutic Implications of Microbiome Research in IBS

The growing recognition of dysbiosis in IBS has led to exploration of microbiome-targeted therapies **(Figure 1.3)**. Probiotics are among the most widely studied interventions, with several randomized controlled trials showing improvements in abdominal pain, bloating, and stool consistency following supplementation with strains such as *Bifidobacterium infantis* and *Lactobacillus plantarum* (Didari et al., 2015); (Ianiro et al., 2020). These probiotics are thought to restore microbial balance, enhance barrier function, and modulate immune responses, though efficacy varies depending on strain and patient subtype (Heeney et al., 2018).

Prebiotics, nondigestible fibers that stimulate growth of beneficial bacteria, have also been investigated. Supplementation with fructo-oligosaccharides and galacto-oligosaccharides has been shown to increase *Bifidobacterium* abundance and improve stool regularity in IBS patients. Synbiotics, which combine probiotics and prebiotics, may offer synergistic benefits, though evidence remains limited (Heeney et al., 2018).

Fecal microbiota transplantation (FMT) represents a more radical approach, aiming to restore microbial diversity by transferring stool from healthy donors. Early trials have reported symptom improvement in subsets of IBS patients, particularly those with diarrhea-predominant IBS (IBS-D), though results are inconsistent and long-term safety remains uncertain (Johnsen et al., 2018). Standardization of donor selection, preparation, and delivery methods is needed before FMT can be widely adopted in IBS management (El-Salhy et al., 2020).

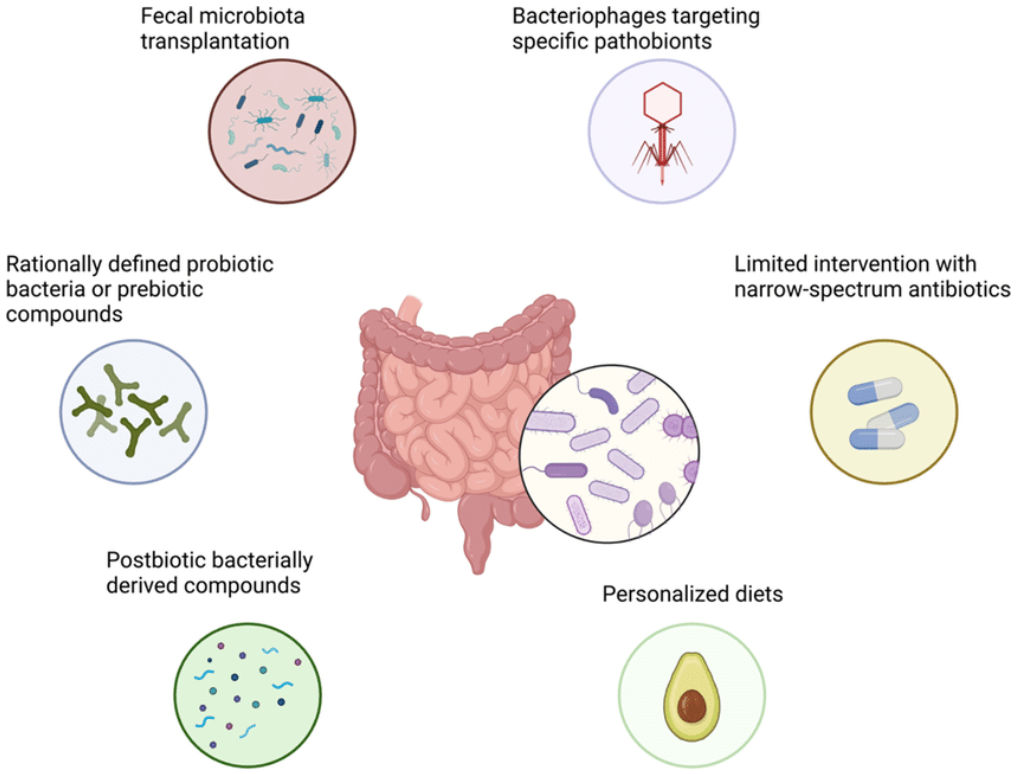


Figure 1.3 Therapeutic Strategies for Gut Microbiome Modulation. Overview of current and emerging interventions to modulate the gut microbiota, including fecal microbiota transplantation, bacteriophage therapy, narrow-spectrum antibiotics, personalized diets, postbiotics, and rationally selected probiotics or prebiotics (Pethakamsetty et al., 2023).

Dietary interventions also play a central role in modulating the microbiome. The low FODMAP diet, which restricts fermentable oligosaccharides, disaccharides, monosaccharides, and polyols, has been shown to reduce IBS symptoms by limiting substrates for gas-producing bacteria (O'Halloran et al., 2017). However, long-term adherence may reduce microbial diversity, raising concerns about sustainability (Halmos et al., 2015).

Personalized medicine approaches are increasingly emphasized, with microbiome profiling used to predict treatment response. For example, patients with reduced butyrate-producing bacteria may benefit more from butyrate-enhancing probiotics or prebiotics (Moran, 2017). Integration of metagenomic data with clinical and dietary information could enable tailored therapies that address individual microbial imbalances (Leventi et al., 2018).

Despite promising advances, challenges remain in translating microbiome research into clinical practice. Variability in study designs, small sample sizes, and heterogeneity in patient populations complicate interpretation of results (Ianiro et al., 2020). Moreover, regulatory and ethical considerations surrounding interventions such as FMT highlight the need for rigorous clinical trials and standardized protocols. Nonetheless, microbiome-targeted therapies represent a frontier in IBS management, offering potential for more effective and personalized treatments.

1.2.9 Methodological Challenges and Considerations in IBS Microbiome Research

Despite significant progress in microbiome science, methodological challenges continue to complicate IBS research. One major issue is variability in sample collection and storage. Stool samples are the most common source for microbiome analysis, yet differences in collection methods, transport conditions, and preservation techniques can introduce biases in microbial composition (Filardo et al., 2024). For instance, postponed freezing or the application of non-standard preservatives may modify the relative abundances of anaerobic organisms (Jin et al., 2016).

DNA extraction methods also influence results. Different kits vary in their ability to lyse Gram-positive bacteria, leading to underrepresentation of taxa such as *Bifidobacterium* and *Clostridium* (Costea et al., 2017). This variability complicates cross-study comparisons and highlights the need for standardized protocols in IBS microbiome research (Knight et al., 2018).

Sequencing platform choice further impacts outcomes. While 16S rRNA sequencing remains widely used, it suffers from limited taxonomic resolution and primer bias (Michael & L., 2007). Whole-genome metagenomics provides greater resolution and functional insights but requires higher sequencing depth and computational resources (Quince et al., 2017). Differences in bioinformatics pipelines, including taxonomic classifiers and reference databases, can also yield divergent results from the same dataset (Nayfach et al., 2016).

Another challenge is the heterogeneity of IBS itself. Patients differ in subtype, symptom severity, and comorbidities, making it difficult to identify consistent microbial signatures (Leventi et al., 2018). Limited statistical power and generalizability are often due to small sample sizes and the absence of stratification by IBS subtype (Tap et al., 2017). Longitudinal designs are rarely

employed, leaving uncertainty about whether dysbiosis precedes IBS onset or results from chronic symptoms (Tap et al., 2017).

Ultimately, confounding variables like as nutrition, pharmacological interventions, and psychological stressors must be meticulously regulated. Dietary consumption significantly affects microbiome composition; yet, several research depend on self-reported food frequency surveys, which may be imprecise (Tap et al., 2017). Exposure to antibiotics and the utilization of proton pump inhibitors might substantially modify microbial populations, hence confounding the interpretation of data specific to IBS (Tap et al., 2017).

Addressing these methodological challenges is essential for advancing IBS microbiome research. Standardization of protocols, adoption of multi-omics approaches, and careful cohort design will improve reproducibility and enable more precise identification of microbial signatures relevant to IBS pathophysiology.

1.2.10 Research Gaps and Future Directions

Despite significant advances in understanding the gut microbiome in IBS, several gaps remain. First, most studies have been conducted in Western populations, with limited representation from South Asia and virtually none from Bangladesh (Lamb et al., 2019). Given that diet, environment, and genetics strongly influence microbiome composition, findings from Western cohorts may not be directly applicable to Bangladeshi patients (Dong & Gupta, 2019). This highlights the need for region-specific investigations that account for local dietary patterns, sanitation conditions, and healthcare practices (Bakhshi et al., 2014).

Second, methodological limitations persist. Many studies rely on 16S rRNA sequencing, which provides taxonomic but limited functional insights (Michael & L., 2007) Whole-genome metagenomics offers greater resolution, enabling identification of strain-level differences and functional pathways, but remains underutilized in IBS research due to cost and computational challenges (Quince et al., 2017). Future studies should integrate metagenomics with meta transcriptomics, metabolomics, and proteomics to capture dynamic host-microbe interactions (Knight et al., 2018). Third, heterogeneity in IBS subtypes complicates interpretation of microbiome data. Diarrhea-predominant (IBS-D) and constipation-predominant (IBS-C) patients often exhibit distinct microbial and functional signatures, yet many studies fail to stratify cohorts

accordingly (Tap et al., 2017). Future research should adopt subtype-specific designs to better link microbial alterations with clinical phenotypes (Leventi et al., 2018).Fourth, longitudinal data are scarce. Most studies are cross-sectional, limiting understanding of temporal dynamics in microbiome changes and their relationship to symptom fluctuations (Caruana et al., 2015). Longitudinal metagenomic studies could clarify whether dysbiosis precedes IBS onset or results from chronic symptoms, thereby informing causality (Jeffery et al., 2012). Fifth, therapeutic translation remains limited. While probiotics, prebiotics, and fecal microbiota transplantation show promise, variability in patient response underscores the need for personalized approaches (Lamb et al., 2019). Microbiome-based diagnostics and predictive models could guide individualized therapies, but require validation in diverse populations (Leventi et al., 2018). Finally, ethical and practical considerations must be addressed. Issues such as donor selection in FMT, data privacy in microbiome sequencing, and equitable access to advanced diagnostics are critical for responsible implementation of microbiome-based interventions (El-Salhy et al., 2020). Addressing these challenges will ensure that microbiome research translates into safe, effective, and accessible therapies.

In summary, future IBS microbiome research should prioritize region-specific studies, employ advanced multi-omics approaches, stratify by IBS subtype, adopt longitudinal designs, and focus on personalized therapeutic translation. This thesis contributes to filling these gaps by applying whole-genome metagenomics to IBS patients in Bangladesh, thereby generating novel insights into dysbiosis in a unique population.

1.3 Aims and objectives of this study

IBS is increasingly recognized as one of the most common functional gastrointestinal disorders affecting millions of individuals across diverse populations in both developed and developing regions worldwide. In Bangladesh, IBS cases are on the rise, likely driven by rapid urbanizations and transitions in diet, environment, and lifestyle. Despite this growing public health concern, high-resolution microbial studies in the Bangladeshi population remain limited. To date, no whole-genome shotgun metagenomic study has comprehensively characterized the gut microbiome of IBS patients in Bangladesh, leaving critical gaps in our understanding of region-specific microbial signatures, functional pathways, and potential disease mechanisms.

Hence, the primary aim of this study was to generate the first whole-genome shotgun metagenomic dataset of IBS patients in Bangladesh, providing a comprehensive characterization of their gut microbiome. This study focused on uncovering the taxonomic composition, functional capacities, and potential virulence-associated features of the gut microbiota in this population. This project aims to better diagnostic methods, deepen mechanistic understanding, and inspire future therapy efforts for IBS in Bangladesh and other resource-limited environments by developing a basic microbial framework.

The specific objectives of this study were:

- To comprehensively characterize the gut microbial composition of Bangladeshi IBS patients.
- To find investigate potential virulence factors and pathogenic elements in the gut microbiome of IBS patients.
- To identify distinct molecular and/or microbial signatures linked with IBS.
- To enable prompt identification and accurate categorization of IBS utilizing microbiological indicators.
- To investigate the therapeutic ramifications of the characteristic microorganisms linked to health.
- To assess the therapeutic efficacy of probiotics and prebiotics, informed by metagenomic data.
- To develop advanced machine learning based models for detecting IBS at its early progression using metagenomic features.

Chapter 02

Materials and Methods

2. Materials and Methods:

The Metagenomic based approach toward profiling the gut microbiome of IBS patients and healthy individuals followed two established pipelines. A schematic representation of the overall workflow is shown in **Figure 2.1**

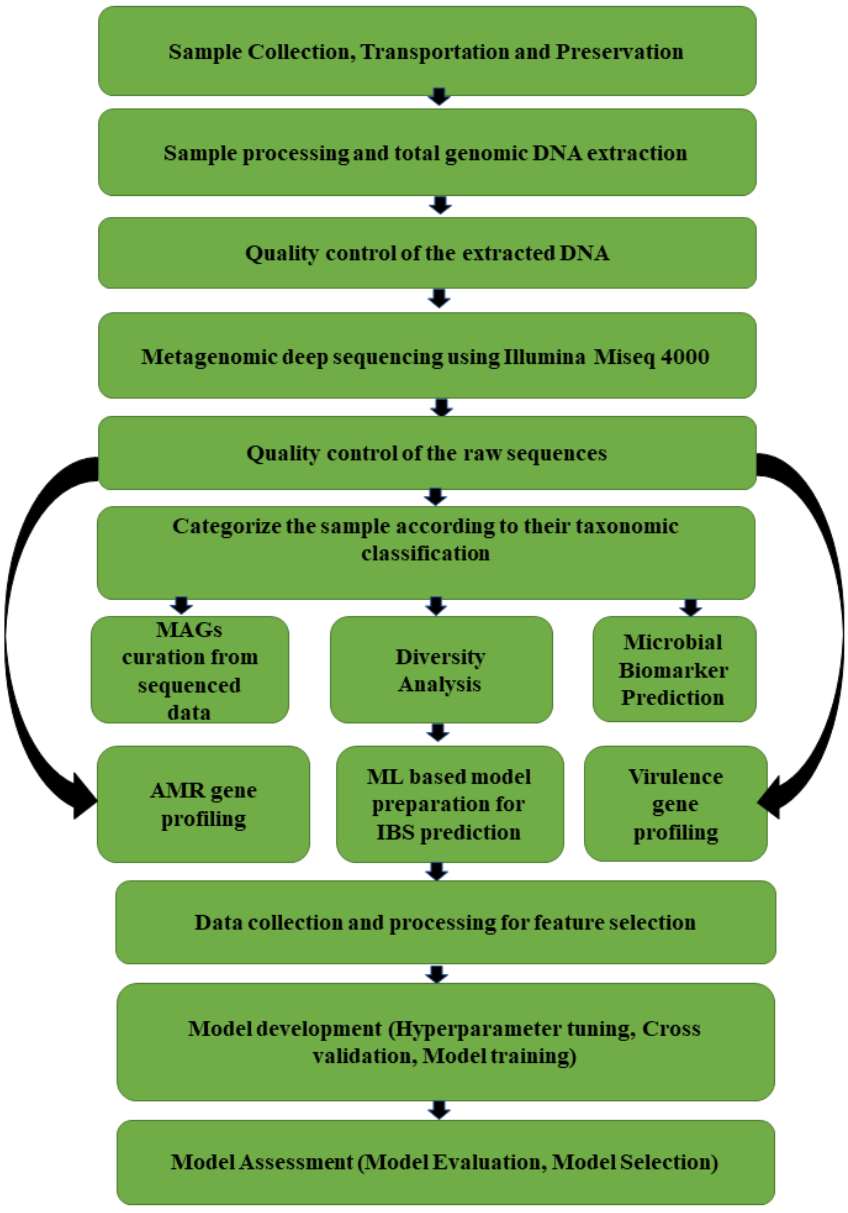


Figure 2.1 Workflow for Microbiome-Based IBS Prediction and Functional Profiling. Sequential overview of metagenomic analysis pipeline from hypothesis development to faecal sample processing, sequencing, and bioinformatic profiling, culminating in machine learning based prediction of IBS and characterization of antimicrobial resistance, diversity, and virulence biomarkers.

2.1 Site selection and patient identification

The National Gastroliver Institute & Hospital, Mohakhali, Dhaka was selected as the site of sample collection since this is the first specialized hospital in Bangladesh for patients with gastrointestinal complications (**Figures 2.2 and 2.3**). Patients having IBS were mainly identified and diagnosed based on Rome IV criteria and the identification was carried out in cooperation with resident physicians at the hospital. A total of thirty (n = 30) stool samples were collected to conduct this study. Among the thirty samples, twenty (n = 20) were collected from IBS-diagnosed patients and ten (n = 10) stool samples were collected from healthy individuals for comparative analysis.

Figure 2.2: National Gastroliver Institute & Hospital

2.2: Sample collection and storage

Appropriate ethical clearance was obtained for a human sample collection from the Ethical Review Committee, Faculty of Biological Sciences, University of Dhaka (Ref. No. 190/Biol. Scs.). The study did not involve the risk of psychological or informational harm. All experiments were performed in accordance with relevant guidelines and regulations. Signed informed consent was

obtained from the participants. They were adequately informed of the nature and purpose of the study and were assured of maintaining the confidentiality of the data being used in the research. A structured questionnaire was prepared to record basic demographic information and disease-specific questions. Patients were adequately informed about the appropriate procedure for stool sample collection. The sample was collected in a 10 mL plastic container that was sterilized in advance. Sample collecting containers were promptly placed in a temporary ice box and thereafter transported to the Microbial Genetics and Bioinformatics Laboratory (MGBL) at the Department of Microbiology, University of Dhaka, where they were instantly frozen at -80 ℃ for further processing.

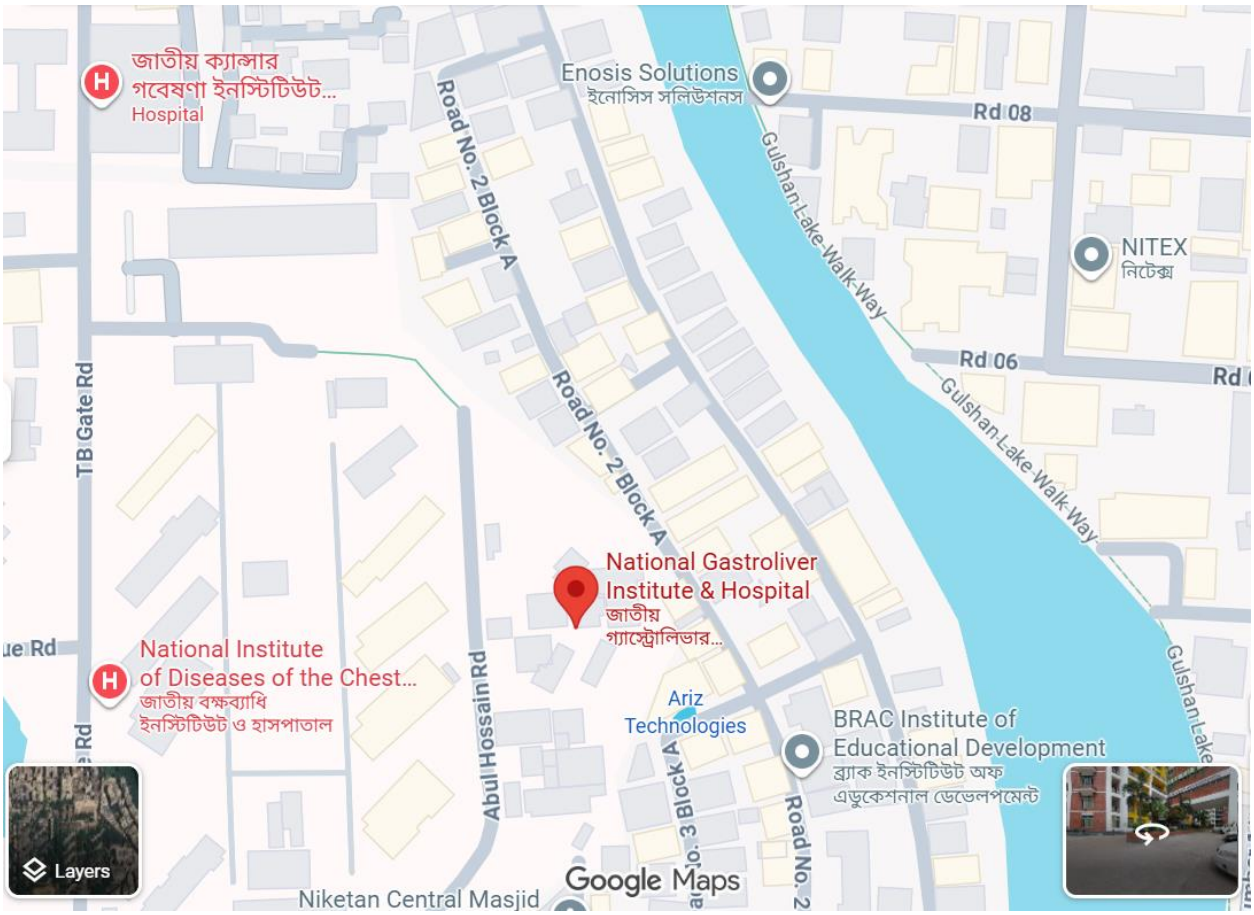


Figure 2.3: Location map of National Gastroliver Institute & Hospital.

2.3 DNA extraction

Total DNA was extracted from each sample using QIAamp PowerFecal Pro DNA Kit (mention the manufacturer's name and specification origin country etc.) as per the manufacturer's protocol.

According to the protocol, 250 mg of the samples was transferred to a fresh sterilized container (Mirsepasi et al., 2014). The kit allows the extraction of highly pure DNA to be used for subsequent use in NGS-based sequencing. The purity of the DNA was checked using a Nanodrop spectrophotometer (mention the machine's specification, country etc.). Samples DNA concentration higher than 15 ng/ml with 200 uL were stored for metagenomic sequencing.

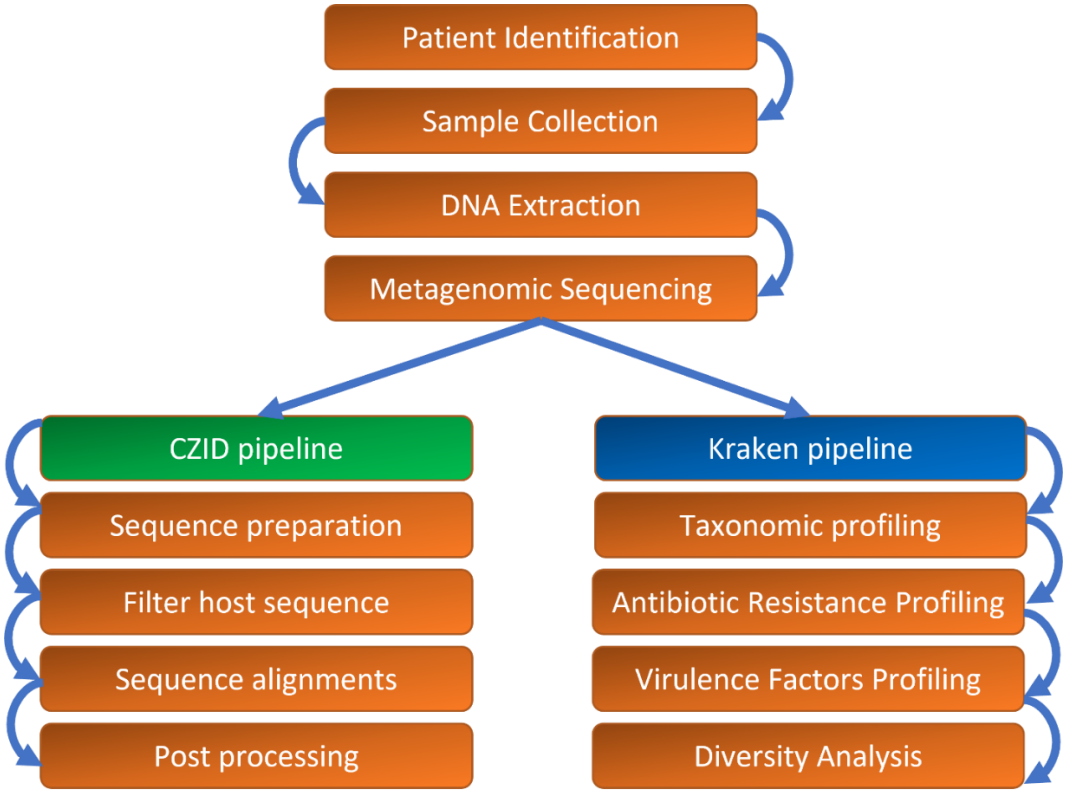


Figure 2.4: A schematic diagram of the workflow to profile the gut microbiome of IBS and healthy groups. Workflow diagram illustrating the metagenomic sequencing and analysis process, highlighting parallel pipelines using CZID and Kraken for microbial profiling, resistance gene detection, virulence factor identification, and diversity analysis.

2.4 Metagenomic sequencing

Illumina-based shotgun sequencing was done through commercial service from EzBiome where the sequencing depth was 100 M to 120 M reads/sample. A customized, automated library construction approach was used to create unpaired, shotgun fragment sequencing libraries. Following library preparation, each sample was subjected to PCR amplification and Illumina

sequencing, as directed by the manufacturer. The Illumina processing program was used to filter short, mixed, and low-quality reads from raw sequence data. Whole metagenomic shotgun sequencing generated about 3-5 Gb of high-quality data per sample.

Figure 2.5: Metagenomic sequencing from sample to analysis. Schematic representation of the shotgun metagenomic sequencing workflow, illustrating the progression from sample collection to taxonomic profiling and functional annotation of microbial communities.

2.5 Metagenomic analysis

2.5.1 Chan Zuckerberg Infectious Disease-based pipeline

CZID (Chan Zuckerberg Infectious Disease) is a powerful web-based automated metagenomics analysis tool (Langelier et al., 2024) that can process Illumina-based shotgun sequencing data to yield meaningful taxonomic and disease-specific information. CZID leverages a comprehensive robust pipeline of different established algorithms to provide statistically significant results. The process is simplified by uploading FASTA/FASTQ sequence files to the CZID server with appropriate metadata. The pipeline begins its tasks in four steps: (1) sequence preparation, (2) host genome filtering, (3) alignments, and (4) post-processing.

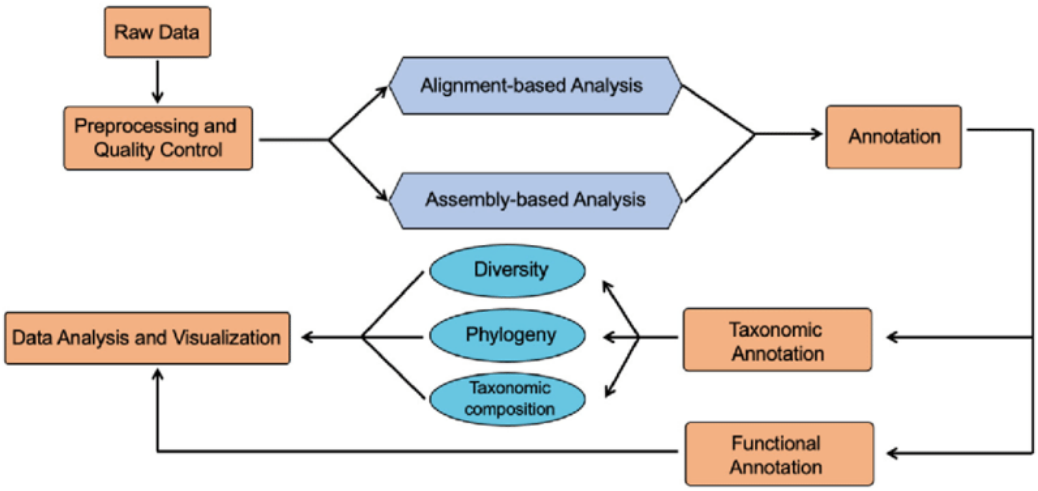


Figure 2.6: Whole metagenomic sequence analysis pipeline. Flowchart depicting a bioinformatics pipeline for processing raw sequencing data, including preprocessing, alignment or assembly-based analysis, annotation, and downstream steps such as diversity assessment, phylogenetic analysis, and taxonomic and functional annotation for data interpretation and visualization.

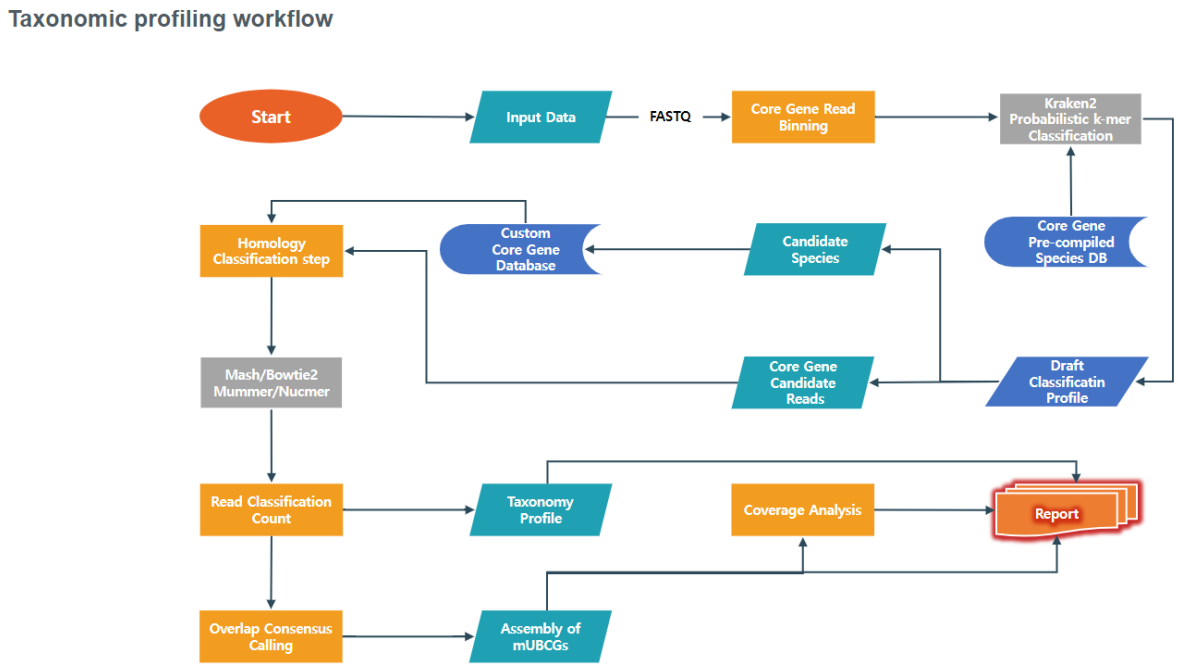


Figure 2.7: Taxonomic profiling of processed data. Detailed workflow of a taxonomic profiling pipeline, illustrating sequential steps from input data processing and core gene binning to probabilistic classification, homology analysis, consensus calling, and final report generation.

2.5.1.1 Sequence preparation

Sequence preparation is initiated by truncating the fastq format files into 75 million fragments. A Trimmomatic (Bolger et al., 2014) software package was used to trim the Illumina sequence adapters. Low-quality reads were subsequently identified and eliminated using the PriceSeq tool. Duplicated reads and low-complexity read from both forward and reverse sequences were cleaned up by a custom-built CZID-debug program and LZW, respectively.

2.5.1.2 Filter host sequence

Filtering of human DNA sequences was performed using the Bowtie2 program using a consensus reference human genome (Langmead & Salzberg, 2012). Bowtie2 is a tool for aligning sequencing reads to large reference sequences that is both quick and memory efficient. It excels at matching reads with 50 to 100 characters to genomes that are quite long (e.g., mammalian). In the next phase, randomly 1 million fragments were subsampled and filtered through the STAR program (T. Wang et al., 2013) followed by another round of the Bowtie2-based host removal process. Finally, GSNAP (T. D. Wu et al., 2016) was used to remove all potentially human sequences.

2.5.1.3 Sequence alignments

The clean sequences were aligned against the NCBI reference databases using Rapsearch2 (Zhao et al., 2012) and GSNAP. Rapsearch2 gave us taxonomy-based counts, and GSNAP gave us FASTA files with annotations. The sequences were ready to be put together with the taxon counts and notes (**Figure 2.6**).

2.5.1.4 Post-processing

For assembly, metaSPAdes (Nurk et al., 2017) was used to generate contigs and scaffolds. The contigs were run using Blast and refined under GSNAP and Rapsearch2. Assembly results also generated coverage stats. Proper accession numbers were assigned to the retrieved bacterial species. Finally, annotation was done for non-host FASTA with revised Taxonomy IDs after the BLAST-based match refinement.

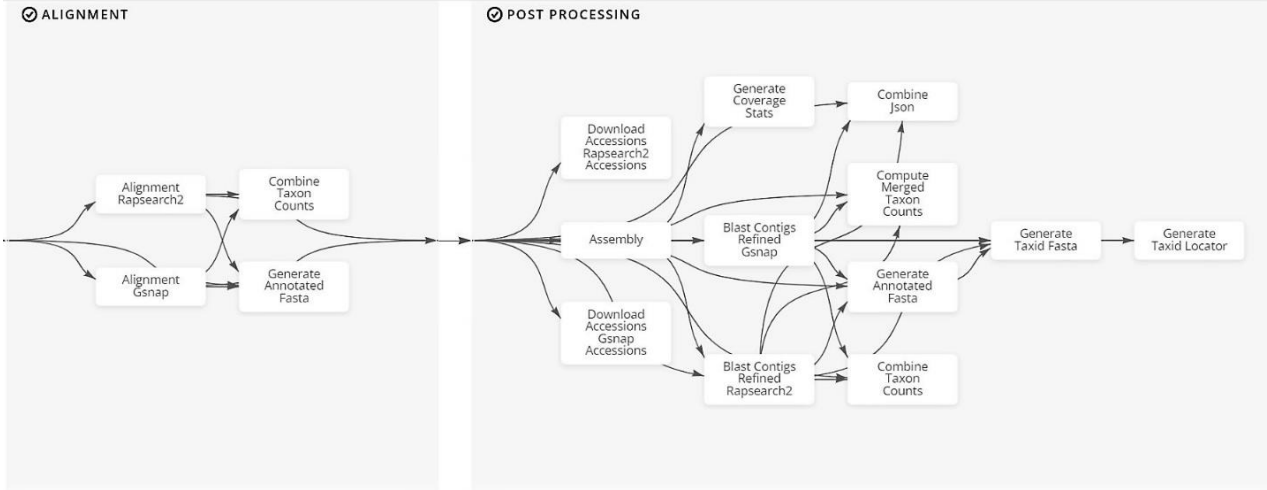


Figure 2.8: Schematic diagram of the programs used in alignment and post-processing in CZID. Flowchart illustrating a bioinformatics pipeline for sequence alignment and post-processing, including taxon count generation, annotated FASTA creation, contig refinement, coverage analysis, and taxonomic localization.

2.5.2 Kraken-based pipeline

2.5.2.1. Taxonomic profiling

The profiling procedure began with a search for bacterial and archaeal species in each raw metagenomic sample read with Kraken2 and the k2_standard database (Wood et al., 2019). Non-bacterial hits were removed from the results, and abundance was calculated based on the total number of classified reads. The Taxonomic Count results were used to generate Krona plots for each sample using the Krona tool (Ondov et al., 2011).

2.5.2.2 Antibiotic Resistance Profiling

Antibiotic resistance genes profiles were produced by using Bowtie2 database composed of NCBI's National Database of Antibiotic Resistant Organisms (NDARO, www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/) reference genes (Langmead & Salzberg, 2012). Each read of the metagenome sample was mapped against these genes using Bowtie2, and the output was then converted and sorted by SAMtools (H. Li et al., 2009). Finally, for each gene found, depth and coverage were calculated by using SAMtools pileup script.

2.5.2.3 Virulence Factors Profiling

Virulence factor profiles were produced by using Bowtie2 database composed of reference factors obtained from the Virulence Factors of Pathogenic Bacteria (VFDB) database (Langmead & Salzberg, 2012; Liu et al., 2019). Each read of the metagenome sample was mapped against these virulence factors using Bowtie2, and the output was then converted and sorted by SAMtools (H. Li et al., 2009). For each virulence factor found, SAMtools pileup script was used to calculate depth and coverage. A Venn diagram of common virulence factors was created using the FunRich program (Pathan et al., 2015).

2.5.2.4 Diversity Analysis

Alpha diversity statistics, including Chao1 diversity, Shannon diversity, and richness, were calculated using the R package Vegan and phyloseq on the taxonomic count profiles of each sample (McMurdie & Holmes, 2013; Oksanen et al., 2012). Taxonomic counts were rarefied using the R rarefy vegan function to the minimum classified total reads of the samples, then a distance matrix was calculated using the vegdist vegan function with the Jaccard distance metric (Oksanen et al., 2012).

2.6 Comparison with global data

A dataset containing HC sample data (n = 492), IBS sample data (n = 617), and total sample data- (N = 1109) was used in this current study. The dataset included 1,093 gut metagenomes from public databases across all continents, along with our own sequenced data. 16 samples were sequenced from Bangladeshi IBS and HC samples. The rest of the samples were downloaded from the NCBI sequence public repository. During the selection of the datasets, we have focused on the IBS data covering all regions of the world. In the selection process, we also tried to get more Asian samples.

To obtain a homogeneous dataset and analysis, we initially downloaded all the sequences and uploaded them to the CZID website. Here, the host sequences were removed and only the microbial parts of the sequence were retrieved. From CZID, we have prepared our taxonomic classification file where 1109 sample sequences were present. After getting the taxonomic table, we have analyzed the samples for their Alpha diversity, Beta diversity, LDA (Linear Discriminant Analysis) for distinct microbial patterns among samples between HC and IBS patients. During this

analysis, we used MicrobiomeAnalyst 2.0 web-based server (Dhariwal et al., 2017). We have also used phyloseq R packages for plotting (McMurdie & Holmes, 2013) and many other R packages during the analysis process.

Table: 2.1: Overview of NCBI Public BioProjects Utilized for Metagenomic Analysis

BioProjects	Origin of country	Platform	Instrument
PRJNA705217	USA	ILLUMINA	Illumina HiSeq 4000
PRJNA1143938	China	ILLUMINA	Illumina NovaSeq 6000
PRJNA1199203	United Kingdom	ILLUMINA	Illumina NovaSeq X Plus
PRJNA1082632	Canada & India	ILLUMINA	Illumina NovaSeq 6000
PRJNA1088844	New Zealand	ILLUMINA	Illumina NovaSeq 6000
PRJNA1269031	Denmark	ILLUMINA	NextSeq 500
PRJNA1214038	Czech Republic	ILLUMINA	NextSeq 2000
PRJDB7616	Japan	ILLUMINA	Illumina MiSeq
PRJNA1242794	South Korea	ILLUMINA	Illumina MiSeq
PRJNA876782	Bangladesh	ILLUMINA	Illumina NovaSeq 6000
PRJNA1011519	China	ILLUMINA	Illumina MiniSeq
PRJNA692652	China	ILLUMINA	Illumina MiSeq
PRJNA767207	Russia	ILLUMINA	Illumina MiSeq
PRJNA1068549	China	ILLUMINA	Illumina NovaSeq 6000
PRJEB52094	United Kingdom	ILLUMINA	Illumina NovaSeq 6000
PRJEB65418	Finland	ILLUMINA	Illumina MiSeq
PRJNA736955	USA	ILLUMINA	Illumina HiSeq 2500
PRJNA981704	Poland	ILLUMINA	Illumina MiSeq

2.7 Metagenome Assembled Genomes (MAGs) from whole metagenome sequencing data:

Metagenome-assembled genomes (MAGs) are microbial genomes reconstructed from fragmented DNA sequences found in a mixed-species sample like gut microbiome data. From the 16 whole metagenome sequencing data, in order to extract MAGs, we have used MetaBAT2 and MaxBin2 tools. The whole metagenome sequence data in fastQ format were preprocessed for quality checking and trimming. For sequence quality checking, FastQC tool was utilized. After that, the adapters and low-quality sequences were trimmed using Trimmomatic where Phred score limit

was set to 20. After the removal of adapters and low-quality sequences, the raw sequences were assembled using MEGAHIT tool. Following assemble, the trimmed raw reads and assembled metagenomic sequences were used to extract the MAGs for bacterial genomes using METABAT2 MAXBin2. The extracted MAGs from METABAT2 and MaxBin were further crosschecked to avoid redundancy and the quality of the MAGs were determined using Quast. Species level identification was carried out using KmerFinder and LinBase tools. The good quality MAGs were then annotated using Prokka and RASTtk. Virulence factor genes and antimicrobial resistance genes were analyzed using Victors/VFDB and CARD:RGI, respectively. The genome maps were constructed using Proksee. The *in-silico* pathogenicity of the predicted good quality genomes was determined using PathogenFinder tool of Center for Genomic Epidemiology (CGE).

2.8 Identification of Microbial Biomarkers via Linear discriminant analysis (LDA) Effect Size (LEfSe) and Core Microbiome Profiling:

Biomarker and core microbiome analysis was carried out using MicrobiomeAnalyst 2.0 software using default parameters (Dhariwal et al., 2017). MicrobiomeAnalyst 2.0 is a user-friendly web-based platform developed to enable comprehensive statistics, visualization, functional interpretation, and integrative analysis of microbiome data. Biomarker analysis involved LEfSe which is an algorithm for high-dimensional biomarker discovery and explanation that identifies genomic features characterizing the differences between two or more biological conditions (Segata et al., 2011).

2.9 Machine Learning based model selection and validation for IBS diagnosis

The entire Machine Learning (ML) pipeline consists of three primary categories of steps. The initial phase involves dataset preparation, during which data were gathered by analyzing two sequences and subsequently merging them according to their shared bacterial nomenclature to produce the final dataset. Upon completion of the dataset, several preprocessing stages were executed, including the elimination of missing values, data normalization, label encoding, dimensionality reduction, and feature selection. Upon obtaining the final dataset, it was divided into two subsets: training and testing datasets. The second phase involves model construction, encompassing hyperparameter tweaking to identify optimal parameters, 5-fold cross-validation to assess model performance on unseen data, and final model training utilizing the training dataset. The ultimate principal category is model evaluation. Several evaluation criteria were employed to

assess the model, including accuracy, precision, recall, F1 score, and ROC AUC. Logistic Regression (LR) had superior performance compared to the six models we analyzed.

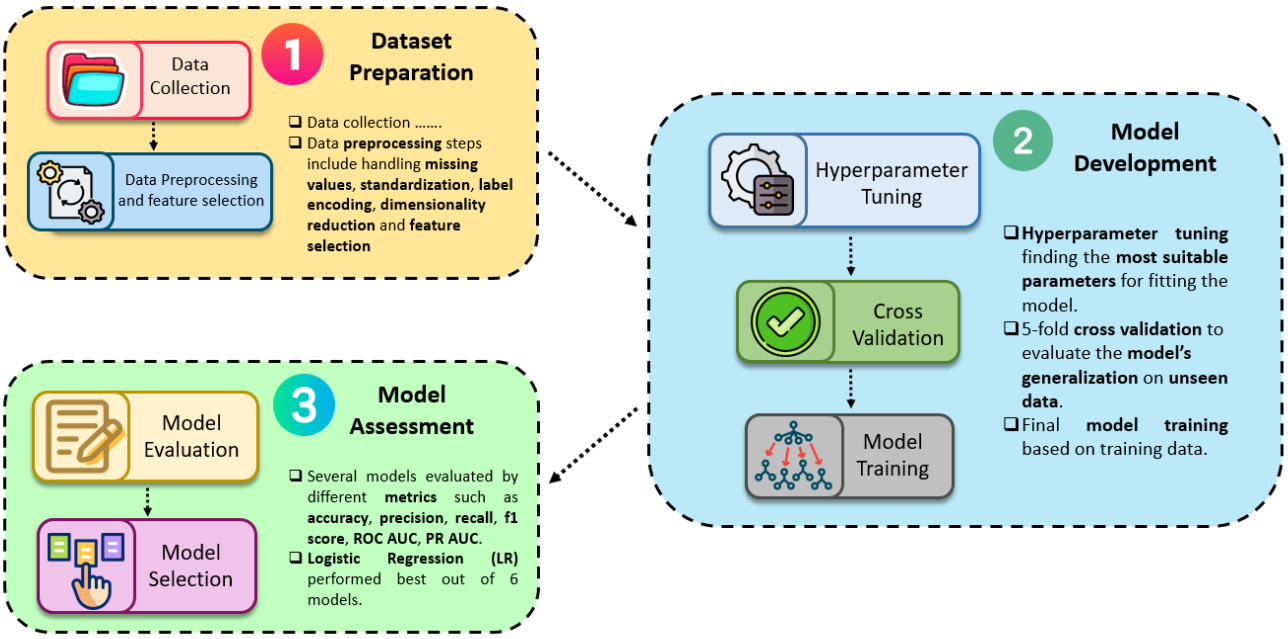


Figure 2.9: Overall pipeline of Machine Learning model development to detect IBS or healthy. Flowchart illustrating the machine learning model development pipeline, encompassing dataset preparation, model training with hyperparameter tuning and cross-validation, and final model assessment using performance metrics for selection.

2.9.1 Dataset Preparation

Dataset preparation is conventionally the initial phase in the ML pipeline. Our shotgun metagenomic sequencing analysis yielded taxonomic data from Bangladeshi samples. We utilized publicly available data from NCBI to integrate our data into the ML model. In our ongoing development of machine learning models, in addition to sequencing 16 shotgun metagenomic datasets, we acquired 594 shotgun metagenomic datasets from NCBI. We additionally obtained 599 16S sequencing datasets from NCBI to improve the precision and reliability of our machine learning model. After preparing all the raw data, it was modified to be suitable for input into a machine learning model for training. Data was collected by examining two sequences: the first sequence comprised 30,753 bacteria from 510 samples, while the second sequence included 19,625

bacteria from 599 samples. The features represent the quantity of each bacterium, resulting in two datasets with 30,753 and 19,625 dimensions, respectively. The ultimate dataset prior to preprocessing involved the amalgamation of these two high-dimensional datasets into a singular entity by identifying the shared features. The resulting dataset had 5,679 attributes and 1,109 samples before preprocessing. One of the columns is called "label" and shows whether each sample is from a person with IBS or a healthy person.

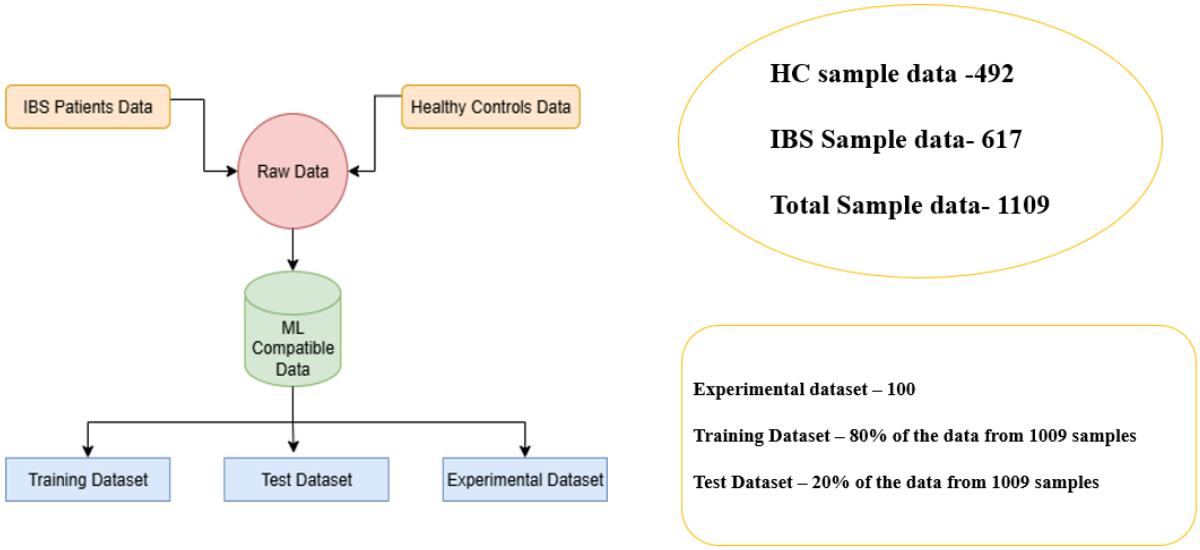


Figure 2.10: Data Collection and Organization for ML approach. Flowchart illustrating the data processing pipeline for machine learning analysis, showing the transformation of raw IBS and healthy control data into ML-compatible datasets, followed by their distribution into training, test, and experimental subsets. Sample statistics and dataset proportions are also summarized.

The preparation steps are very important since the dataset needs to be in the right format for the machine learning models. The first phase of preprocessing was to deal with missing values. This wasn't very strict, though, because only one item had a missing value, therefore the whole row was removed. After removing any missing values, we used the formula (1) to scale the data normally, setting the mean to 0 and the standard deviation to 1. In this formula, x^((i)) is the original value, μ_x is the mean, and σ_x is the standard deviation. It usually makes machine learning models work better and lessens the effect of specific features that are too big. One column shows whether a sample is from a person with IBS or a healthy person, as was said before. We turned the words into numbers by label encoding, giving IBS a 0 and healthy a 1. This is because machine learning

models can only work with numbers. The number of features in a dataset has a big effect on the output of machine learning models. Training is harder with high-dimensional data because it has too many dimensions, which makes it harder to understand and leads to overfitting. So, we need to reduce the number of dimensions and pick important aspects to build strong models. We used the Analysis of Variance (ANOVA) F-statistics (Siraj et al., 2022) to do feature selection. The formula (2) shows that n_i is the frequency of class x_i in the dataset, x_i is the mean of the class, x is the overall mean of the feature, and k is the total number of classes. The Select Best module in Python was used to choose the best 1400 features. The final dataset has 1107 samples and 1400 features. We used t-SNE (t-Distributed Stochastic Neighbor Embedding) to show the high-dimensional dataset. t-SNE makes it easier to move high-dimensional data into a lower-dimensional space (tSNE_component_1 and tSNE_component_2) while retaining the dataset's original qualities (Maaten & Hinton, 2008). In figure 1, each point of data is a sample. The red points are IBS and the blue points are Healthy controls. There are several locations where samples of the same class cluster, even if there is some overlap between the IBS and Healthy categories. This suggests that the characteristics give useful information.

$$x_{std}^{(i)} = \frac{x^{(i)} - \mu_x}{\sigma_x} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

$$\sigma_{cl}^2 = \frac{\Sigma(\underline{x_i} - \underline{x})^2 n_i}{(k - 1)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

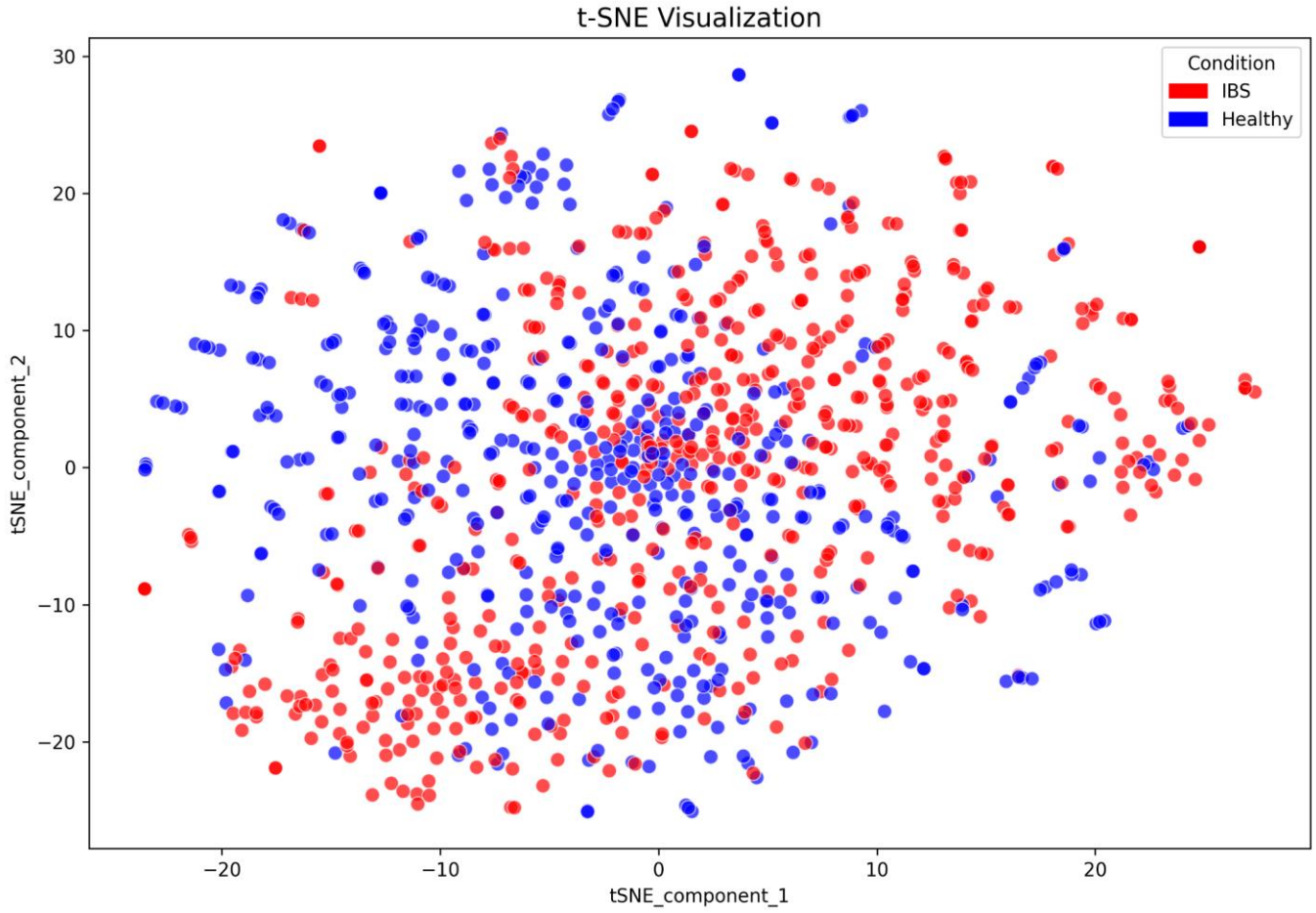


Figure 2.11: t-SNE representation to visualize the clusters of IBS patient and healthy sample data. t-SNE plot visualizing dimensionality-reduced metagenomic data, showing clustering patterns between IBS patients (red) and healthy controls (blue), highlighting group separation and potential microbial signature differences.

2.9.2 Model Development

The model development phase commences with hyperparameter optimization (HPO). Hyperparameters are parameters that govern the learning process of a machine learning model and must be established prior to model training. Therefore, hyperparameters have to be tuned to produce the optimal outputs of the model (Probst, 2019). The objective of HPO is to attain the optimum of an unknown black box function, which can be evaluated for any element of. Consequently, the hyperparameter space (Cho et al., 2020) will be the locus. Numerous optimization strategies encompass Random Search (Bergstra & Bengio, 2012), Grid Search (Belete & Huchaiah, 2022), Bayesian Optimization (J. Wu et al., 2019), gradient-based

optimization (Bengio, 2000), and Particle Swarm Optimization (Kennedy & Eberhart, 1995), among others. This research employs Grid Search and Bayesian Optimization with Optuna (Akiba et al., 2019). Grid Search looks at all the possible combinations of hyperparameters from a given grid.

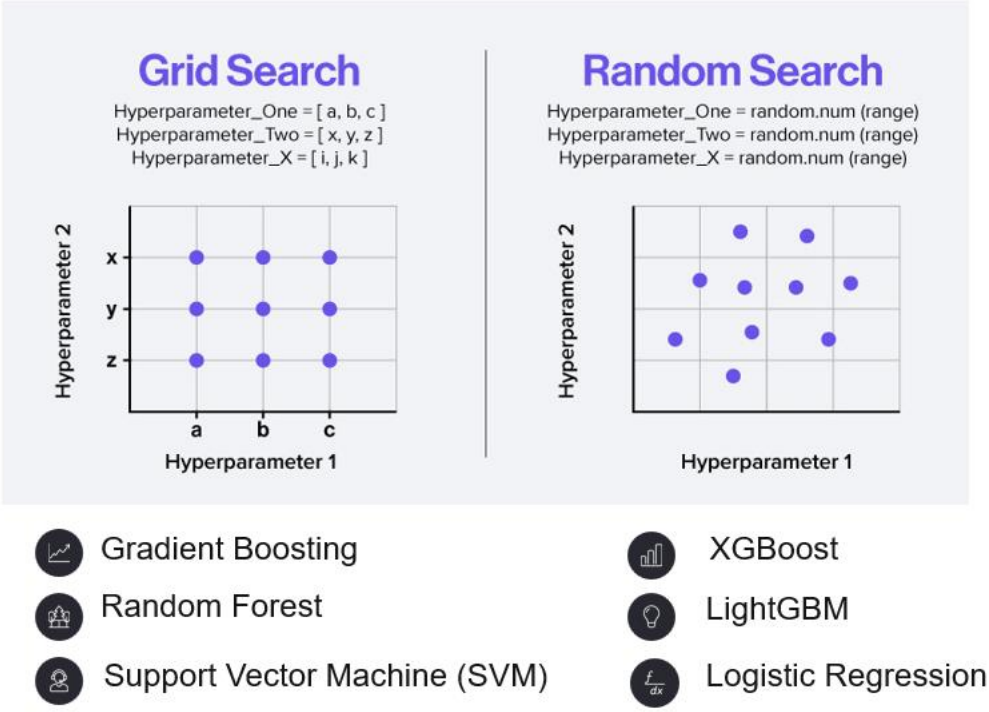


Figure 2.12: Hyperparameter tuning by grid search. Comparative illustration of hyperparameter optimization techniques, showing systematic grid search versus stochastic random search across parameter spaces, with applicable machine learning models listed below.

This costs a lot of processing power and time, especially in search spaces with many dimensions. Optuna, on the other hand, uses a Bayesian optimization method called the Tree-structured Parzen Estimator. This method avoids having to look at all possible hyperparameter combinations in a large hyperparameter space, which greatly improves the performance of the model. Optuna, on the other hand, uses a Bayesian optimization method called the Tree-structured Parzen Estimator. This method doesn't require testing every possible combination of hyperparameters in a huge hyperparameter space, which greatly improves the performance of the model. Consequently, the hyperparameters were optimized by Optuna, taking into account the final dataset, and the resultant

values were subsequently employed for the models' further training. Hyperparameters for over ten models were optimized; however, this work presents six models in **Table 2.2**

Table 2.2: Optimized hyperparameters for all 6 models using Optuna

Model Name	Optimized Hyperparams
Gradient Boosting	n_estimators=271, max_depth=6, min_samples_split=9, min_samples_leaf=4, subsample=0.88932, learning_rate=0.183158
XGBoost	n_estimators=220, colsample_bytree=0.6776, learning_rate=0.08692, max_depth=5, subsample=0.6705, gamma=1.1496297, min_child_weight=1
Random Forest	n_estimators=64, max_depth=15, max_featurs='sqrt', min_samples_split=7, min_samples_leaf=1, bootstrap=False
LightGBM	colsample_bytree=0.9441, learning_rate=0.1596, max_depth=3, min_child_samples=6, n_estimators=152, n_jobs=-1, reg_lambda=0.152087, subsample=0.83876, reg_alpha=0.3124, num_leaves=46
Support Vector Machine	C=9.78914, gamma='scale', kernel='rbf'
Logistic Regression	C=0.0457897, penalty='l2', solver='saga', max_iter=100

After HPO, we utilized cross-validation (CV) to assess how well the model performed on new data, or how well it could generalize. We did 5-fold cross-validation, which meant that in each iteration, 80% of the whole dataset was used for training and 20% for testing. This was different from just one train-test pair. Gradient Boosting, XGBoost, Random Forest, LightGBM, Support Vector Machine (SVM), and Logistic Regression (LR) are the six models that were used for hyperparameter tuning and five-fold cross-validation. We found the average accuracy of all six models in each fold to see how well they worked. After cross-validation, the final model was trained. We first split the entire dataset into two parts: 80% of the data (885 samples) for training and 20% (222 samples) for testing. The six models were then trained on the training dataset, which helped them uncover patterns and relationships in the data.

2.9.3 Model Assessment

Model assessment is a key part of a machine learning pipeline since it tells you how well the models work and how they might perform on new data. This work shows how to use the most common metrics in classic machine learning workflows, such as accuracy, precision, recall, F1 score, and ROC AUC. Accuracy is the number of times that the model accurately predicted an

instance (either positive or negative) out of the total number of instances. Precision is the number of true positives (TP) predictions divided by the total number of true positives (TP) and false positives (FP). Recall, on the other hand, shows how well the model finds the positive cases among true positives (TP) and false negatives (FN). The F1 score is the average of precision and recall. The area under the Receiver Operating Characteristics (ROC) curve shows how well a model does by displaying the true positive rate (TPR) versus the false positive rate (FPR) at different thresholds (Gönen, 2006). The Matthews Correlation Coefficient (MCC) is another technique to measure how well a model works that takes into consideration all of the true positives, true negatives, false positives, and false negatives. It works better in bioinformatics and delivers a more useful answer than other measurements (Chicco et al., 2021). We looked at how well all six models did and used the test data to choose the one that did the best.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad \ldots\ldots (3)$$

$$Precision = \frac{TP}{TP + FP} \quad \ldots\ldots (4)$$

$$Recall\ (or, TPR) = \frac{TP}{TP + FN} \quad \ldots\ldots (5)$$

$$F1\ score = 2 \times \frac{precision \times recall}{precision + recall} \quad \ldots\ldots (6)$$

$$False\ Positive\ Rate, FPR = \frac{FP}{FP + TN} \quad \ldots\ldots (7)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad \ldots\ldots (8)$$

Chapter 03

Results

3. Results

3.1 Demographic information

A total of sixteen stool samples were taken, of which ten of them were confirmed IBS patients, and six were healthy controls (HC). The age range of the participants was from 22 to 70 years. None of the individuals had a history of IBS diagnoses. (**Table 3.1**).

Table 3.1: Demographic information of study participants

Samples	Age	Sex	Hometown	Marital Status	Lactose Intolerance	Stool Type	Family history	Education	Groups
HC1	35	Male	Noakhali	Married	Yes	Normal	No	18	HC
HC2	27	Male	Dhaka	Married	Yes	Normal	No	17	HC
HC3	30	Male	Noakhali	Married	No	Normal	No	18	HC
HC4	27	Male	Dhaka	Unmarried	No	Normal	No	17	HC
HC5	40	Female	Dhaka	Married	No	Mixed	No	14	HC
HC6	34	Female	Dhaka	Married	No	Normal	No	12	HC
IBS1	70	Male	Tangail	Married	Yes	Diarrhea	No	8	IBS
IBS2	27	Male	Dhaka	Unmarried	Yes	Alternating	No	13	IBS
IBS3	26	Male	Dhaka	Married	Yes	Mixed	No	12	IBS
IBS4	25	Male	Borguna	Married	Yes	Alternating	Yes	16	IBS
IBS5	25	Male	Cumilla	Unmarried	Yes	Alternating	No	16	IBS
IBS6	20	Male	Gazipur	Unmarried	Yes	Mixed	No	12	IBS
IBS7	26	Male	Dhaka	Unmarried	No	Alternating	No	16	IBS
IBS8	22	Female	Cumilla	Married	Yes	Diarrhea	No	12	IBS
IBS9	39	Male	Dhaka	Married	Yes	Mixed	No	5	IBS
IBS10	23	Male	Dhaka	Unmarried	Yes	Constipation	Aunt	16	IBS

HC = Healthy Control, IBS = Irritable Bowel Syndrome

3.2 Summary results of metagenomic data

At first, all the samples were underwent for host filtering and quality trimming to get good quality reads. Among all the samples, IBS4 had a greater portion of human DNA which was also successfully removed by utilizing Bowtie2. After cleaning and host filtering, we obtain around 5-6% of clean DNA sequences that represent all the microorganisms' sequences. Both healthy control and IBS samples passed quality control parameters. (**Figure 3.1**) Analysis of the quality-controlled sequences graph revealed that out of 16 samples in our current investigation, 10 samples have over 120 million reads, demonstrating the high quality and depth of the sequence data.

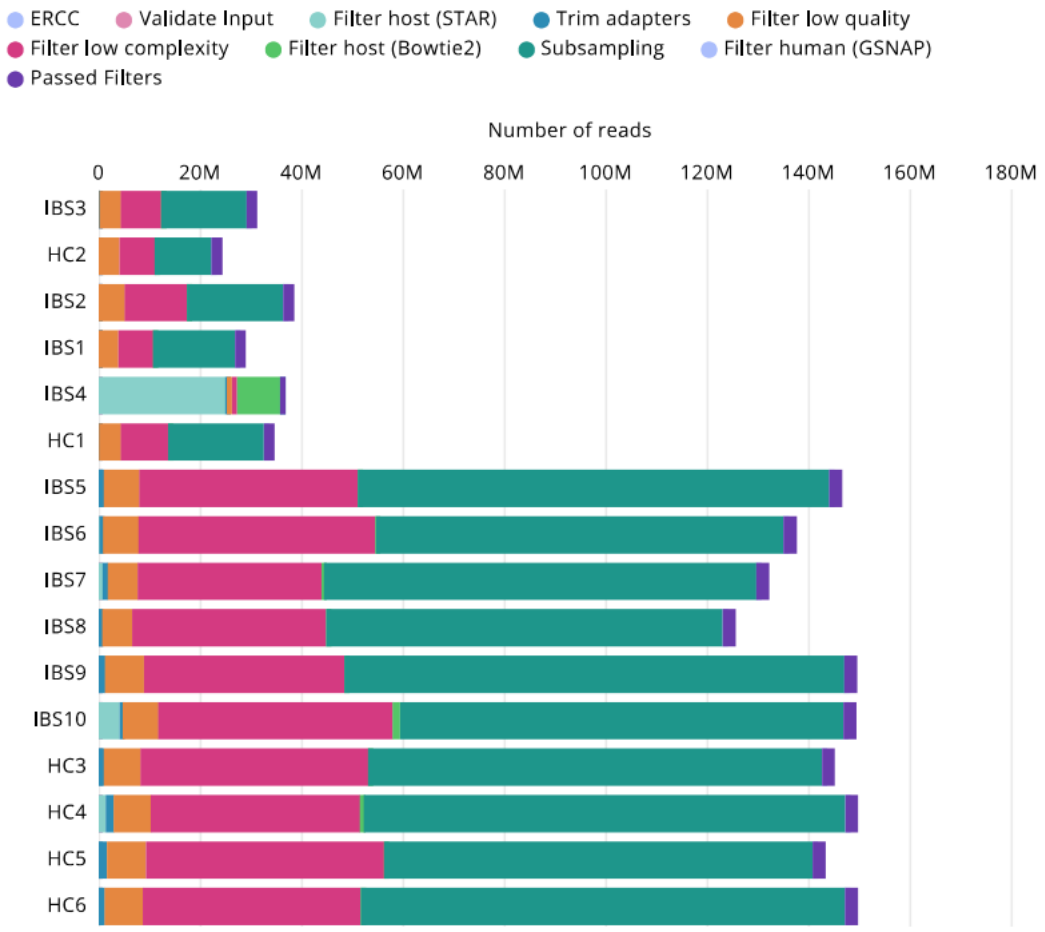


Figure 3.1: Quantity of reads from our sequenced samples. Horizontal bar chart displaying read count distribution across IBS and healthy control samples, segmented by preprocessing steps including host filtering, adapter trimming, quality control, and final read retention.

3.3 Quality of sequencing

In the metagenomic DNA sequences, the percentage of GC content fluctuated between 50% to 75%. In sequence duplication, the rate varied from 5% to 25%. Phred score >30 (S. Zhang et al., 2017), which indicates the base quality in the DNA sequence was excellent across all the sixteen samples (**Figure 3.2**).

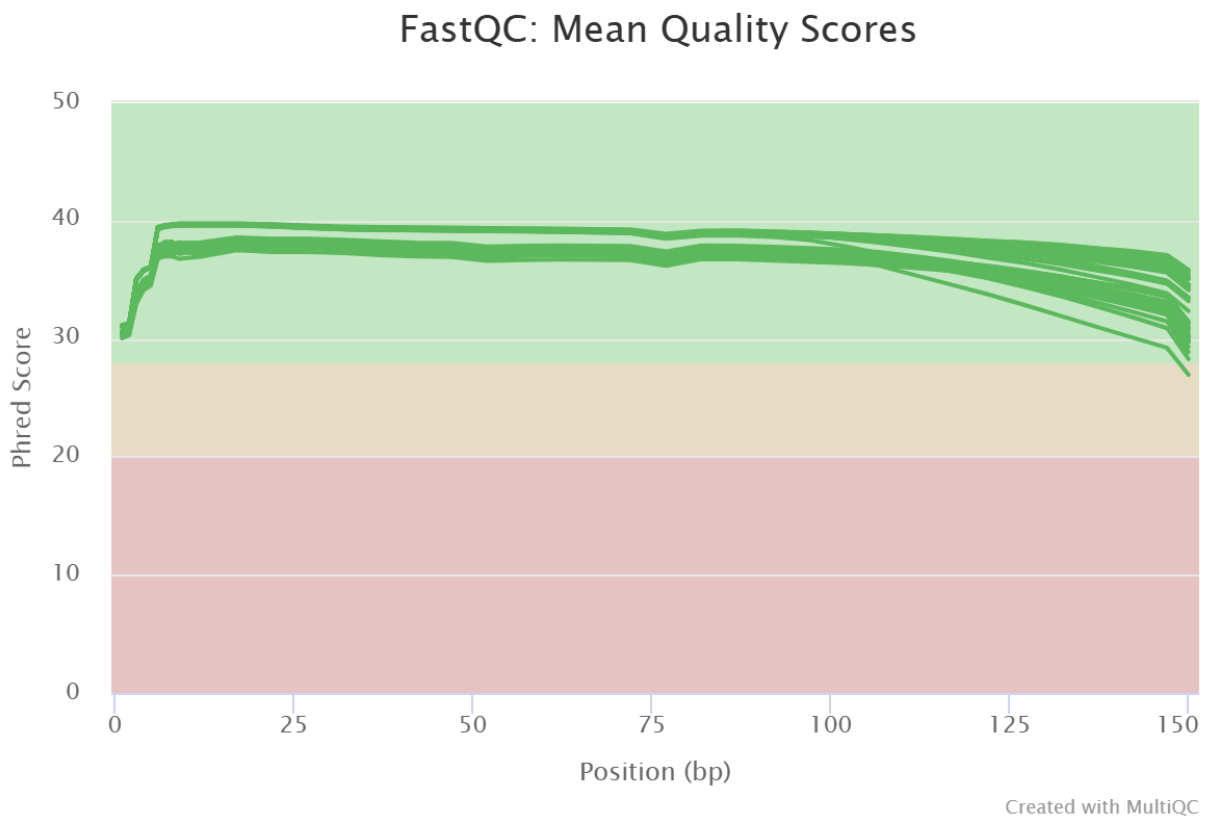


Figure 3.2: Representative figure of quality scores (FastQC) of sequenced data. FastQC plot showing mean Phred quality scores across sequencing read positions, indicating consistently high base call accuracy throughout most of the read length, with a slight decline toward the end.

3.4 Bacterial composition in the gut microbiome

3.4.1 Microbial diversity in the gut of IBS patients revealed dysbiosis:

Following whole metagenome sequence analysis, significant dysbiosis was observed in the gut microbiome of patients with IBS. An investigation into phylum-based classification unveiled a substantial decrease in the abundance of Actinobacteria in individuals with IBS compared to healthy controls. Control 3, 4, 5, and 6 exhibited relatively elevated levels of Actinobacteria. Conversely, all IBS patients displayed markedly reduced levels of Actinobacteria. Notably, IBS1 exhibited no detectable Actinobacteria following sequence analysis. When compared to healthy controls, the gut microbiota of IBS patients showed a significant lack or almost complete absence of *Bifidobacterium* species. Additionally, the gut microbiomes of IBS patients showed a notable

decrease or lack of these species, according to an abundance summary of 25 different *Bifidobacterium* species.

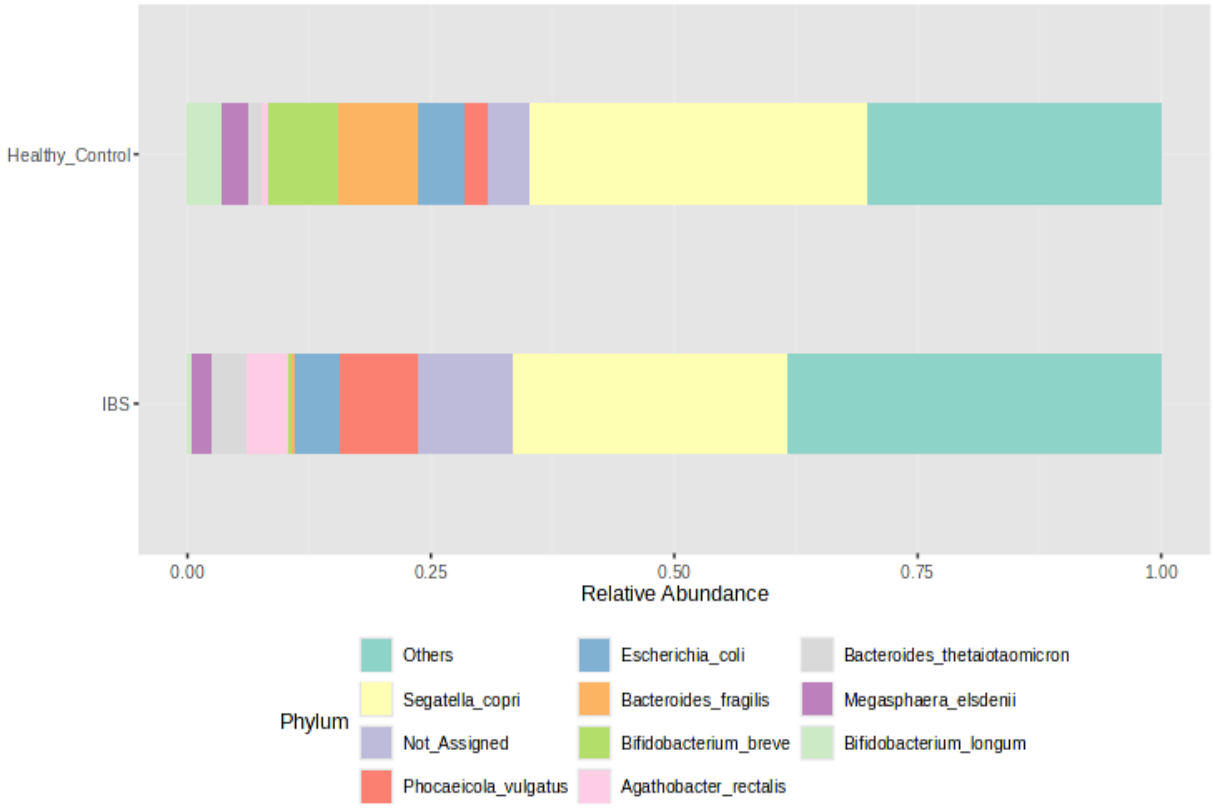


Figure 3.3: Comparative bacterial abundance between individuals with IBS and healthy controls. Stacked bar chart comparing the relative abundance of bacterial species between IBS patients and healthy controls, highlighting compositional differences in gut microbiota across groups.

Notably, *Bifidobacterium breve*, *B. longum* subsp. *suillum*, *B. bifidum*, *B. longum* subsp. *Suis*, *B. longum* subsp. Infantis, *B. longum* subsp. *Kashiwanohense*, *B. catenulatum*, *B. pseudocatenulatum* were detected to have highest distinct abundance in healthy controls and were significantly lower in IBS patients. Furthermore, *B. cebidarum*, *B. felsineum*, *B. reuteri*, *B. scaligerum*, *B. moukalabense*, *B. myosotis* and *B. leontopitheci* were notably elevated in individuals with Irritable Bowel Syndrome (IBS) compared to healthy controls. A significant absence of *Bifidobacterium* was observed in control group 1 and control group 2, both of which were characterized by lactose intolerance.

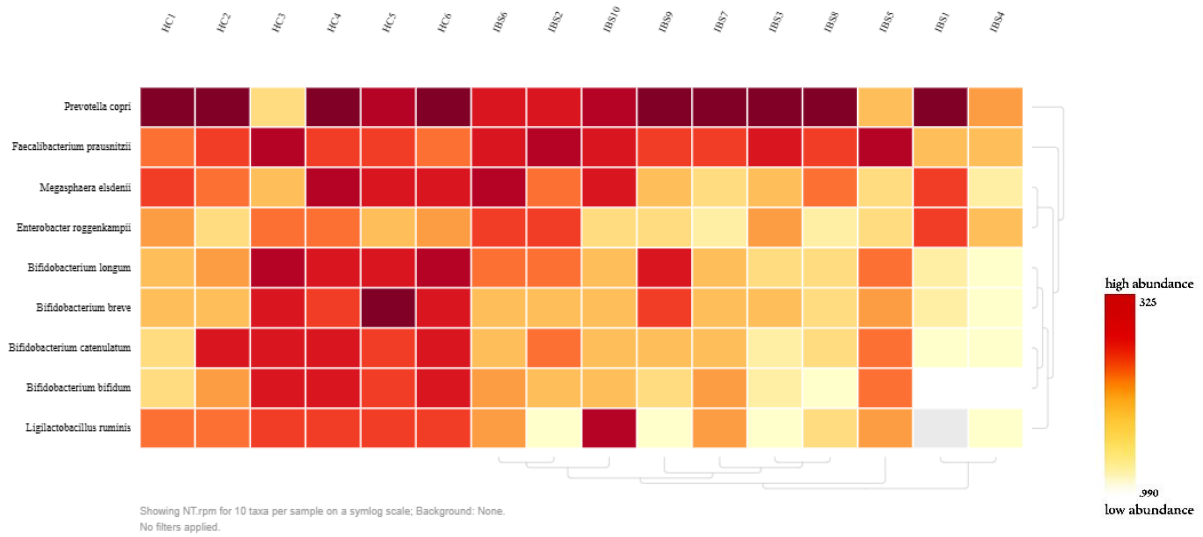


Figure 3.4: Heatmap of the most abundant bacterial taxa. Heatmap showing normalized relative abundance of 10 bacterial taxa across multiple samples, with hierarchical clustering highlighting patterns of microbial distribution and potential group-specific associations.

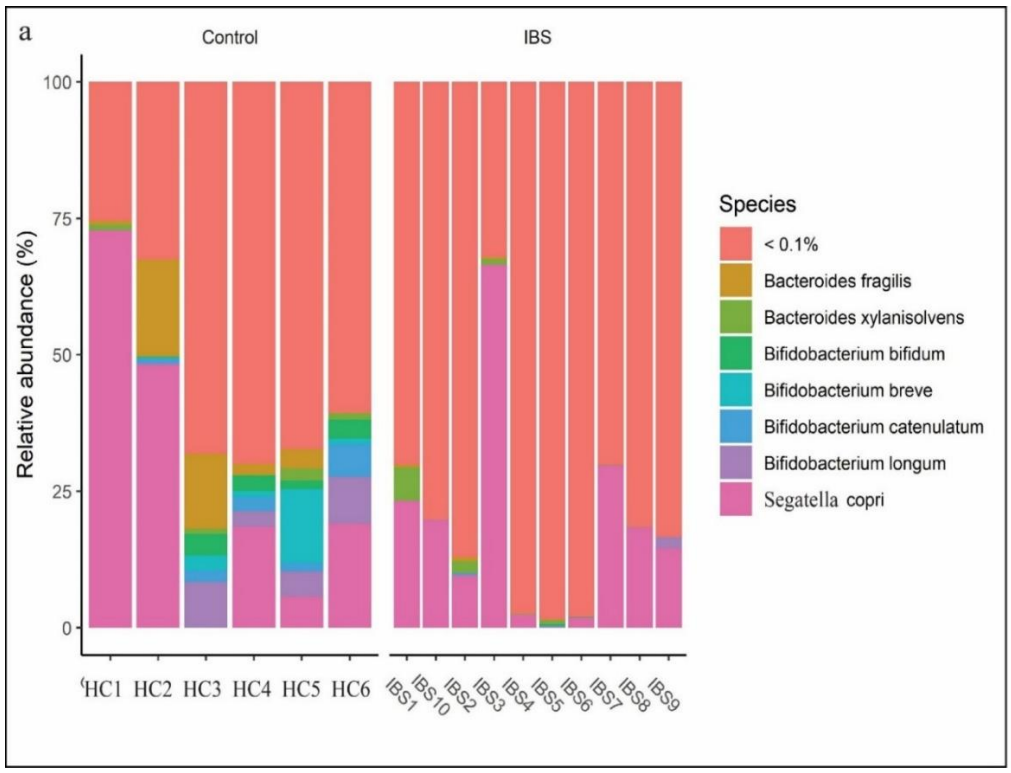


Figure 3.5: Relative abundance of bacterial taxa across all bacterial taxa. Bar chart illustrating the relative abundance of selected bacterial species across individual IBS and healthy control

samples, highlighting compositional differences and dominant taxa such as *Segatella copri* and Bacteroides fragilis.

Furthermore, IBS patients had a higher abundance of Firmicutes in their gut microbiota compared to healthy controls. Notably, *Faecalibacterium prausnitzii* was discovered to be considerably more abundant in IBS patients. Healthy controls, on the other hand, had larger abundances of *Megasphaera elsdenii*, *Megasphaera indica*, and *Lactobacillus luminis*.

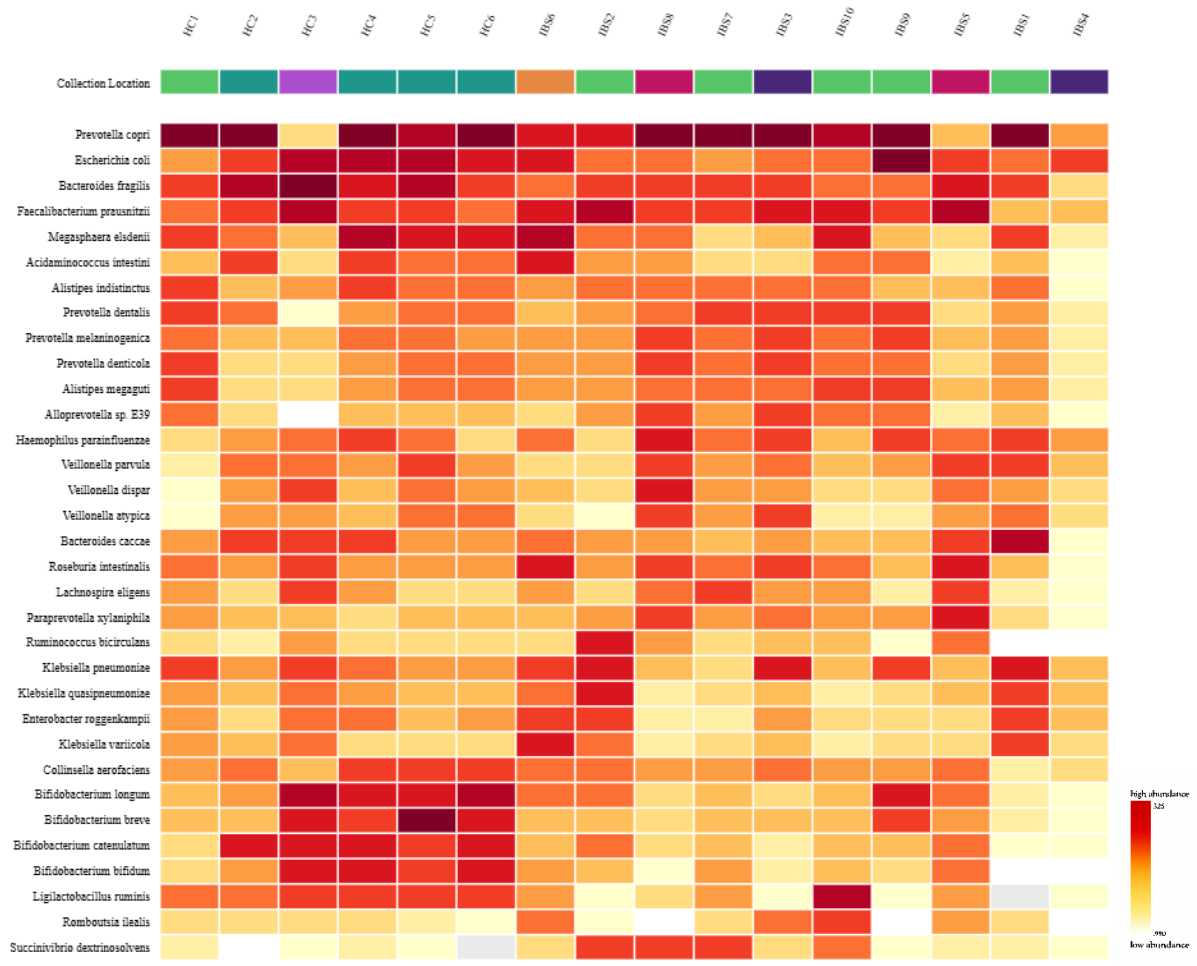


Figure 3.6: The most prevalent species displayed in the heatmap. Heatmap displaying the relative abundance of microbial taxa across samples from different collection locations, with hierarchical clustering revealing patterns of taxonomic distribution and site-specific microbial profiles.

The study of Bacteroidetes found a consistent pattern of relative abundance in the gut microbiota of both IBS patients and healthy people. Bacteroidetes emerged as the most common phylum in both groups, with some deviations in IBS4, IBS6, and Control 3. In IBS4 and IBS6, the dominating phyla were Proteobacteria and Firmicutes, respectively. Among the species within Bacteroidetes, *Segatella copri* exhibited higher abundance levels in both IBS patients and healthy controls.

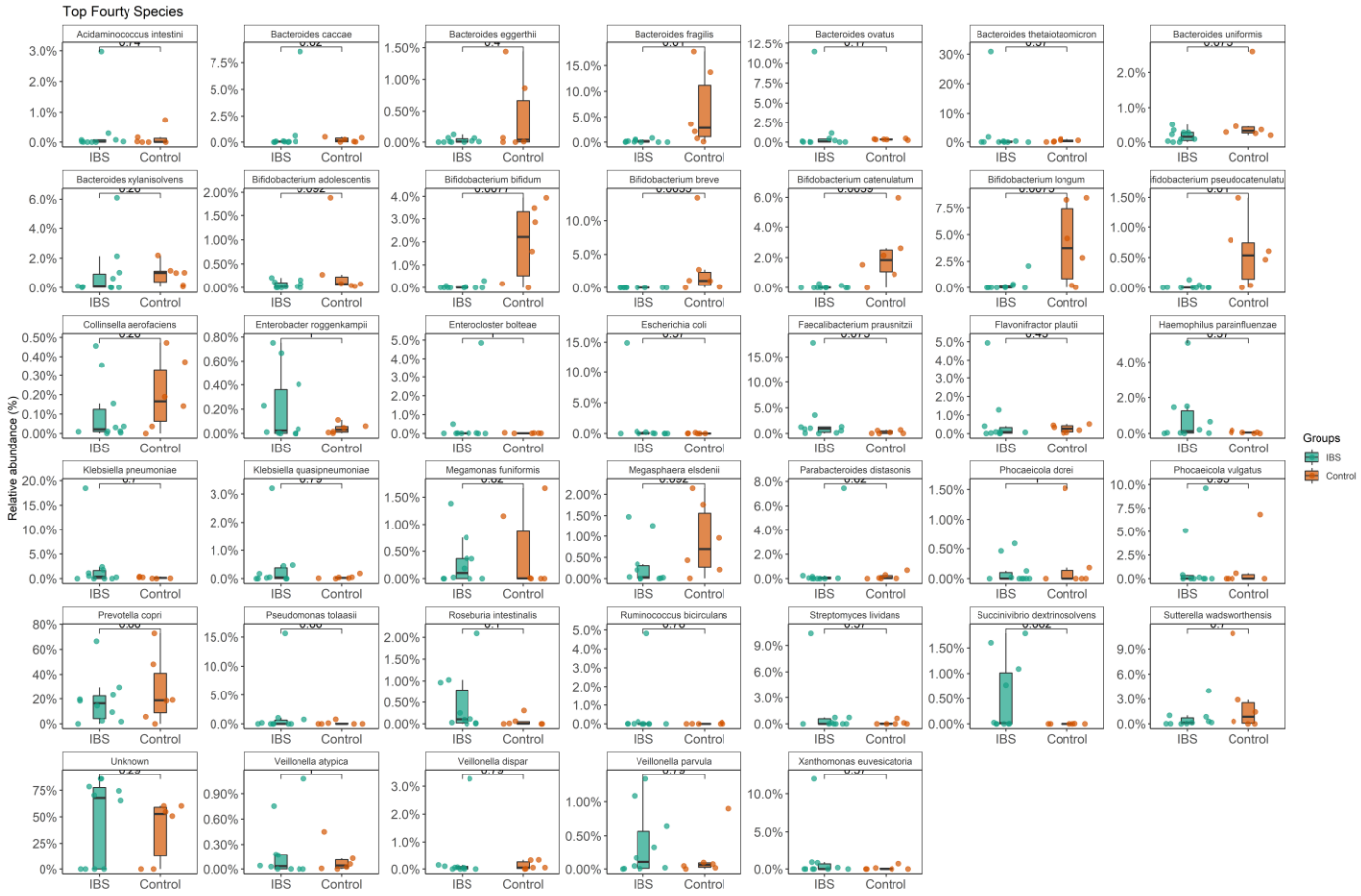


Figure 3.7: Top most 40 organisms present in current study. Box plots comparing the relative abundance of the top 40 bacterial species between IBS patients and healthy controls, highlighting group-specific microbial variations.

Segatella copri was identified as the most abundant bacterium in the gut microbiota of both groups, except for IBS5 and Control3, within the scope of this study. The abundance of Bacteroides fragilis was notably diminished in IBS patients compared to healthy controls. Among the Proteobacteria, *Klebsiella pneumoniae* and *Haemophilus parainfluenzae* were found to be comparatively more prevalent in IBS patients than in healthy controls. Overall abundance analysis also revealed a

relatively high abundance of particular species, such as *Enterobacter roggenkampii*, *Roseburia intestinalis*, *Succinivibrio dextrinosolvens*, and *Veillonella parvula*, in individuals with IBS compared to healthy controls.

3.5 Core Microbiome analysis between Healthy Control and IBS samples

The most common bacterial taxa for Healthy Controls are *Segatella copri*, *Bacteroides fragilis*, *Bacteroides ovatus*, *Bifidobacterium catenulatum*, *Bifidobacterium longum*, and a number of others, all of which have high prevalence rates. The majority of these taxa show a frequency of almost 1.0, suggesting that they are highly prevalent in healthy persons.

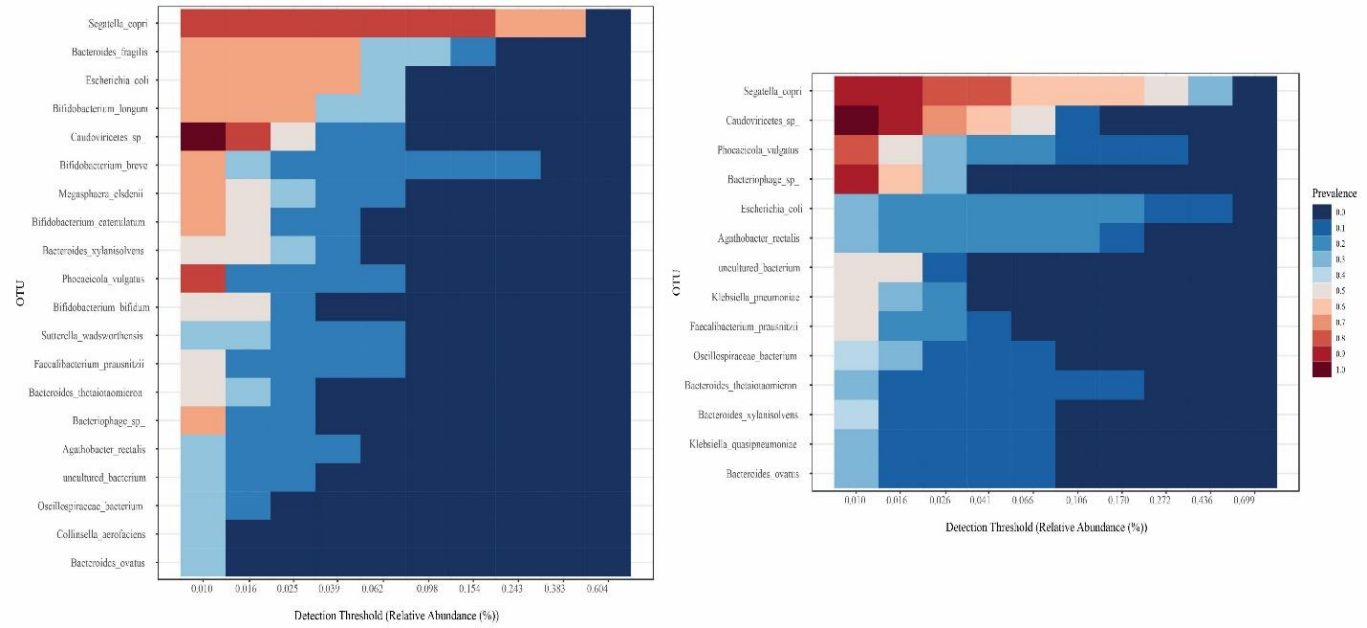


Figure 3.8: Heatmap of Core Microbiome analysis between Healthy Control and IBS samples. Dual heatmaps illustrating the prevalence of selected OTUs (shotgun metagenomics) across varying detection thresholds of relative abundance, with color gradients indicating shifts in microbial representation and clustering patterns.

IBS sufferers, on the other hand, have a very distinct microbiome composition. Compared to the Healthy Control group, taxa including *Bacteroides fragilis*, *Bacteroides ovatus*, and *Segatella copri* are significantly less common. Additionally, IBS patients have a considerably greater prevalence of the Not Assigned taxa, which may suggest a more varied or poorly defined microbiome makeup in this population. Furthermore, IBS patients have higher levels of other

bacteria, such as *Sutterella wadsworthensis* and *Klebsiella pneumoniae*, which may indicate a shift in the microbial community toward taxa associated with inflammation and dysbiosis. By identifying specific bacterial taxa that may promote or delay disease development, this differential microbiome landscape suggests a link between IBS pathophysiology and the composition of the gut microbiota. The reduced frequency of beneficial taxa (such as some species of *Bifidobacterium*) in IBS patients raises the possibility that microbial imbalances may be a factor in the gastrointestinal symptoms of IBS.

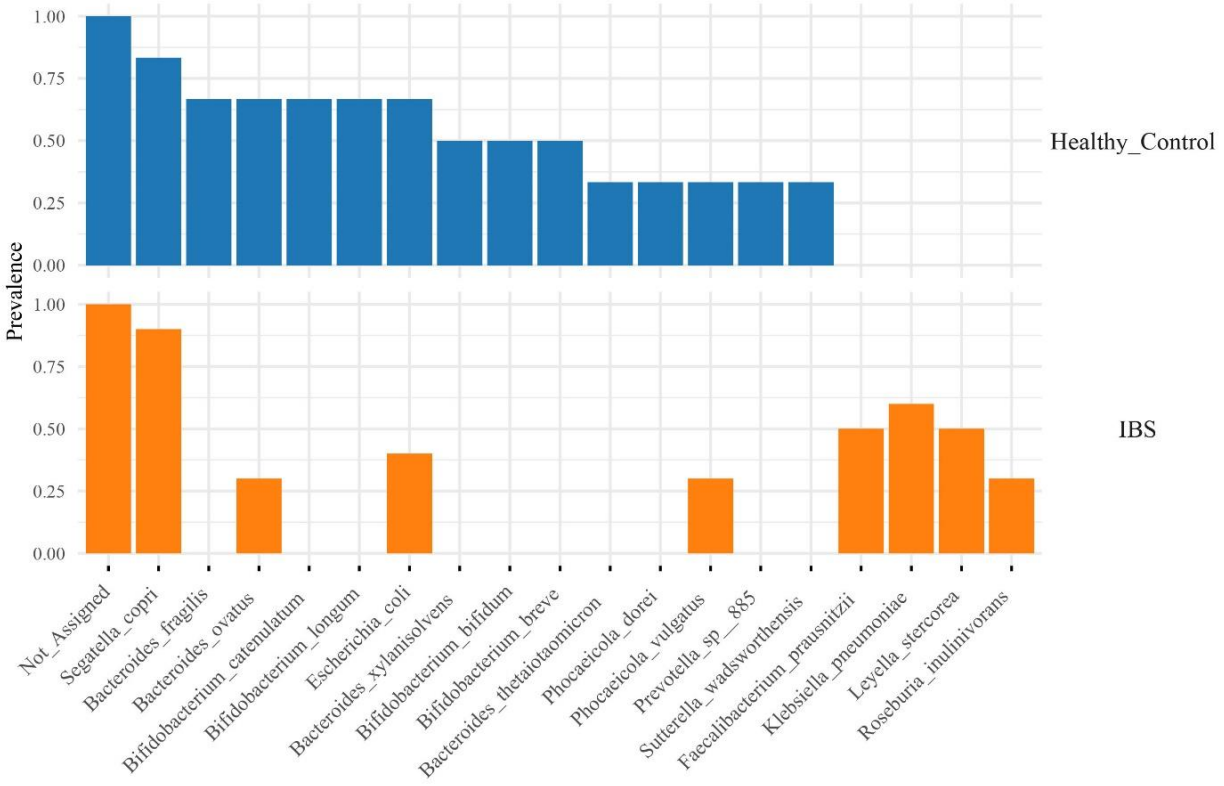


Figure 3.9: Core microbiome analysis between Healthy Controls and IBS patients. Bar chart comparing the prevalence of selected bacterial species between IBS patients and healthy controls, illustrating group-specific microbial patterns and differential species representation.

3.6 Profile of known pathogenic organisms

3.6.1 Pathogenic strains in IBS patients

Analysis of taxonomic abundance, conducted with a 95% confidence limit (Z Score = 1.96), revealed a comparatively elevated presence of pathogenic bacterial strains in IBS patients compared to the healthy control group **(Figure 3.10).**

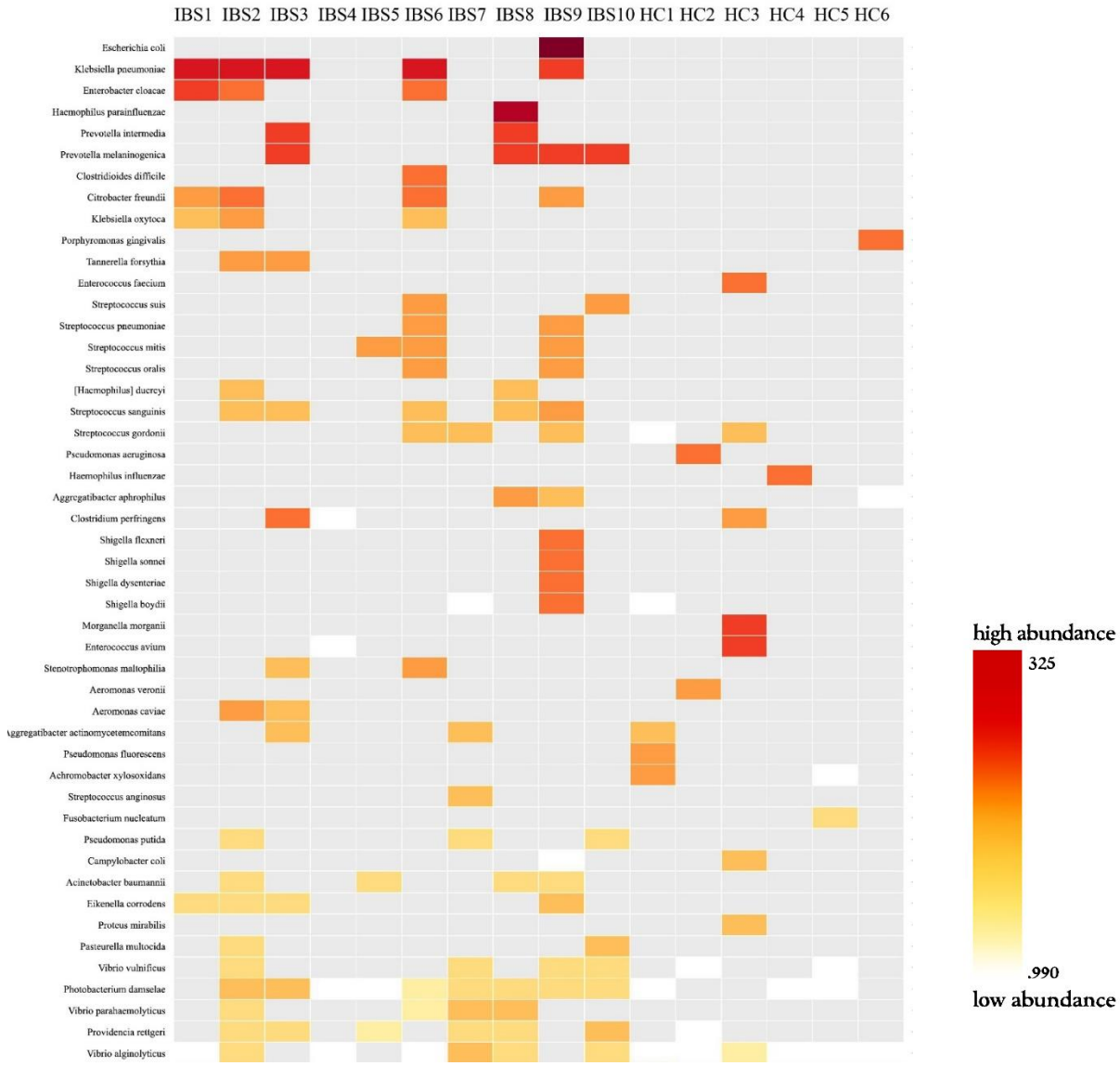


Figure 3.10: Heatmap of the comparative analysis of known pathogenic organisms. Heatmap showing the relative abundance of bacterial species across IBS and healthy control samples, with color intensity indicating species-level variation and potential group-specific microbial signatures.

Pathogenic *bacteria, including Klebsiella pneumoniae*, *Enterobacter cloacae*, *Prevotella melaninogenica*, *Citrobacter freundii*, and *Streptococcus sanguinis*, were found to be distinctly abundant in the gut of IBS patients. Furthermore, *Acinetobacter baumannii*, *Photobacterium damselae*, *Providencia rettgeri*, *Nocardia brasiliensis*, and *Bordetella bronchiseptica* exhibited relatively greater abundance in IBS patients compared to the healthy control group. Within the *Vibrio genus*, *Vibrio vulnificus*, *Vibrio parahaemolyticus*, and *Vibrio alginolyticus* were prevalent in the gut microbiota of IBS patients.

3.7 Relative Abundance of Bacteriophage

3.7.1 Relative abundance of Bacteriophage in IBS patients and Healthy controls

The relative abundance of bacteriophage sequences in the gut of IBS patients was found to be significantly lower than that of healthy controls. *Salmonella* phage, *Escherichia* phage, and *Erwinia* phage were shown to be more abundant in healthy controls than in IBS patients. Interestingly, compared to healthy controls, the gut microbiota of IBS patients showed a comparatively increased abundance of the *Faecalibacterium* virus. The *Faecalibacterium virus Lugh*, the *Faecalibacterium virus Taranis*, the *Faecalibacterium virus Oengus*, and the *Faecalibacterium virus Toutatis* were notable members of the *Faecalibacterium* virus group. In addition, several additional viruses and phages were found in the gut microbiota of healthy controls and IBS patients (**Figure 3.11**).

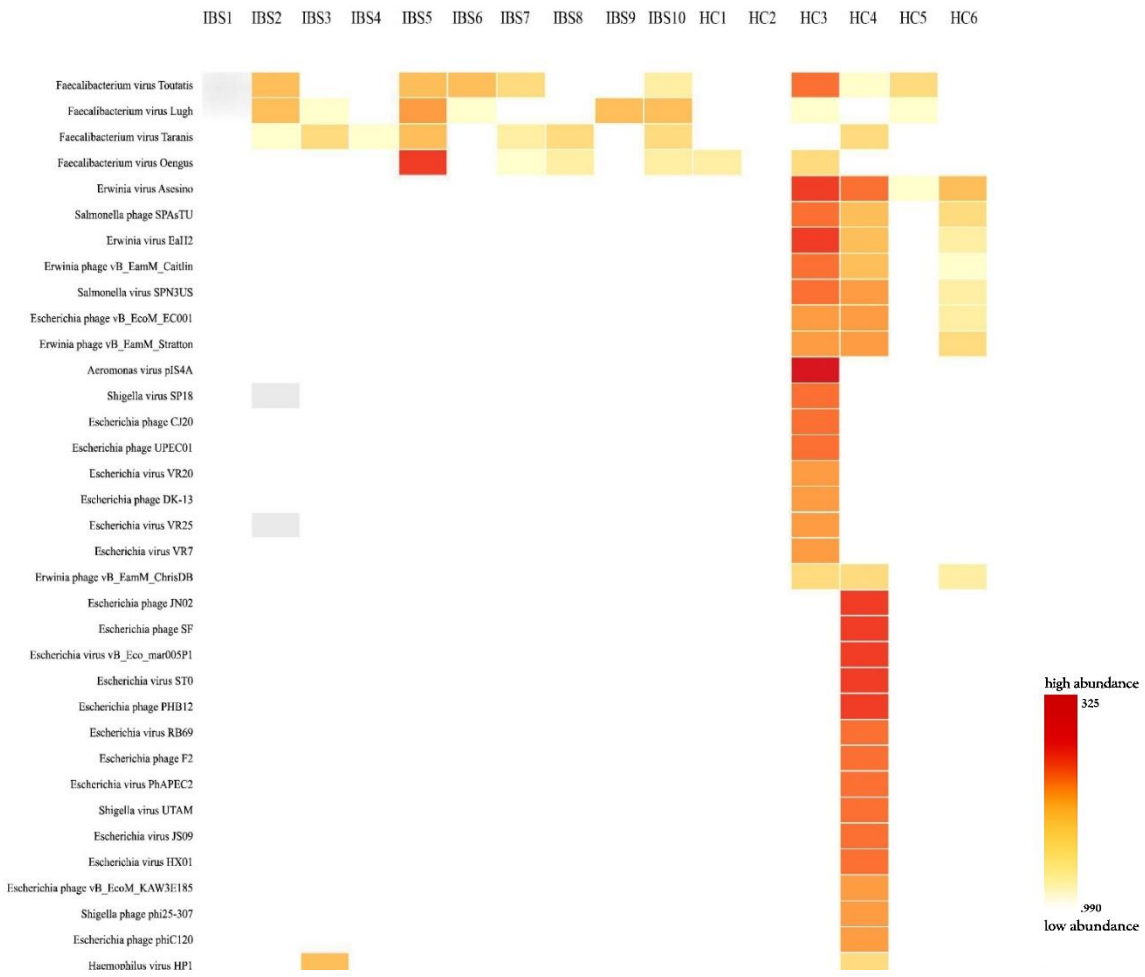


Figure 3.11: Abundance of Bacteriophages. Heatmap comparing the relative abundance of bacteriophages across IBS and healthy control samples, with color intensity indicating species prevalence and highlighting group-specific microbial patterns.

3.8 Alpha diversity among patients and the control group

Important microbiome characteristics between the two study groups were identified by the alpha diversity index analysis. Alpha diversity (Xia & Sun, 2023a) metrics between the Control and IBS groups are compared in the figure. Each of the alpha diversity metrics Observed, Chao1 (Chao, 1987), ACE (Buja & Kass, 1985), Shannon (Konopiński, 2020), and Simpson (Somerfield et al., 2008) indices is shown in a distinct boxplot (**Figure. 3.12)**. The observed index's median value for the Control and IBS groups is comparable, suggesting a similar level of species richness. The two groups' interquartile ranges (IQRs) are thus comparable, however Control has somewhat greater

variability, as seen by the slightly wider box. By taking into consideration the quantity of uncommon species in the sample, the Chao1 value obtained for species richness takes unseen species into account. Comparable species richness is indicated by the similar median value of the observed index for the Control and IBS groups. The two groups' interquartile ranges (IQRs) are thus comparable, however Control exhibits somewhat greater variability, as shown by the slightly bigger box. The number of uncommon species in the sample is taken into account when estimating the Chao1 value for species richness. The number of uncommon species in the sample is taken into account when estimating the Chao1 value for species richness.

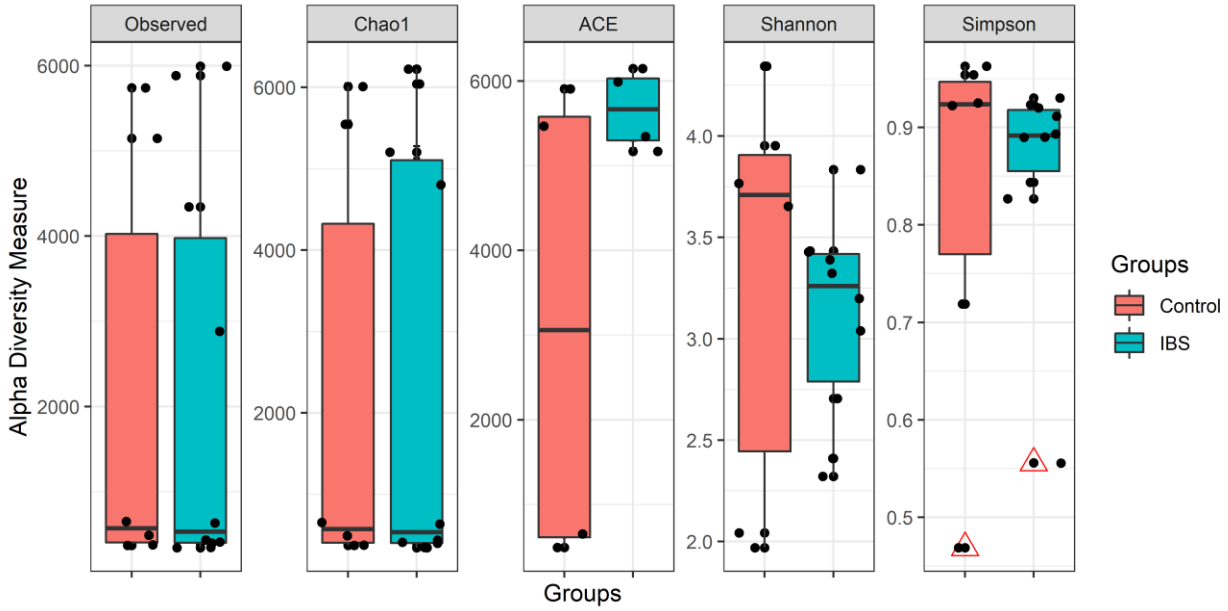


Figure 3.12: Alpha diversity among IBS patients and the Healthy Control group. Boxplots comparing alpha diversity indices (Observed, Chao1, ACE, Shannon, Simpson) between IBS and control groups, illustrating differences in microbial richness and evenness.

Both the control and IBS groups have similar chao1 median values, indicating a similar estimated species richness. The Control group, on the other hand, exhibits more variability, as indicated by a larger interquartile range (IQR) and a higher number of outliers. The Control group had a higher median in abundance-based Coverage Estimator (ACE) value analysis than the IBS group, indicating better richness in abundant species. The interquartile range (IQR) of the IBS group is narrower, indicating less diversity in species richness among abundant species. The Control group has several outliers, indicating that the data is more variable. The Shannon index, which takes into account both species richness and evenness, found that the Control group had a higher median.

Similarly, the Simpson index, which highlights common or dominant species, indicated that the Control group had a higher median than the IBS group.

3.9 Beta diversity among patients and control group

The Principal Coordinates Analysis (PCoA) plot (Figure) depicts the beta diversity of microbial communities among samples, with Axis 1 and Axis 2 accounting for 24.6% and 19.1% of the variance, respectively. The results reveal a distinct difference between the microbial populations of individuals with IBS and healthy controls. In terms of group clustering, the control samples tend to cluster together, indicating a relatively homogeneous gut microbiota composition among healthy individuals. In contrast, the IBS samples reveal a distinct and independent clustering pattern, indicating that the gut microbiota composition of IBS patients differs significantly from that of healthy controls. The ellipses surrounding the groups represent 95% confidence intervals, highlighting the differences in microbial community structure between IBS patients and healthy controls. Notably, several samples from various areas show closeness, indicating that, while geographic location influences microbiota composition, IBS has a greater impact.

Non-metric multidimensional scaling (NMDS) analysis was used to visualize the changes in microbial communities between control and irritable bowel syndrome (IBS) groups in various locations. The NMDS plot (**Figure 3.13**) shows an apparent clustering of samples depending on their health state (Healthy Control vs. IBS) and geographical location. The NMDS1 and NMDS2 axes explain the variety in the data, while the confidence ellipses indicate the variability within the groups. The microbial communities of Dhaka and Gazipur clearly overlap, but those of Borguna, Cumilla, Noakhali, and Tangail exhibit more pronounced differences. IBS samples exhibit greater dispersion than control samples, indicating increased heterogeneity in the microbial communities associated with IBS.

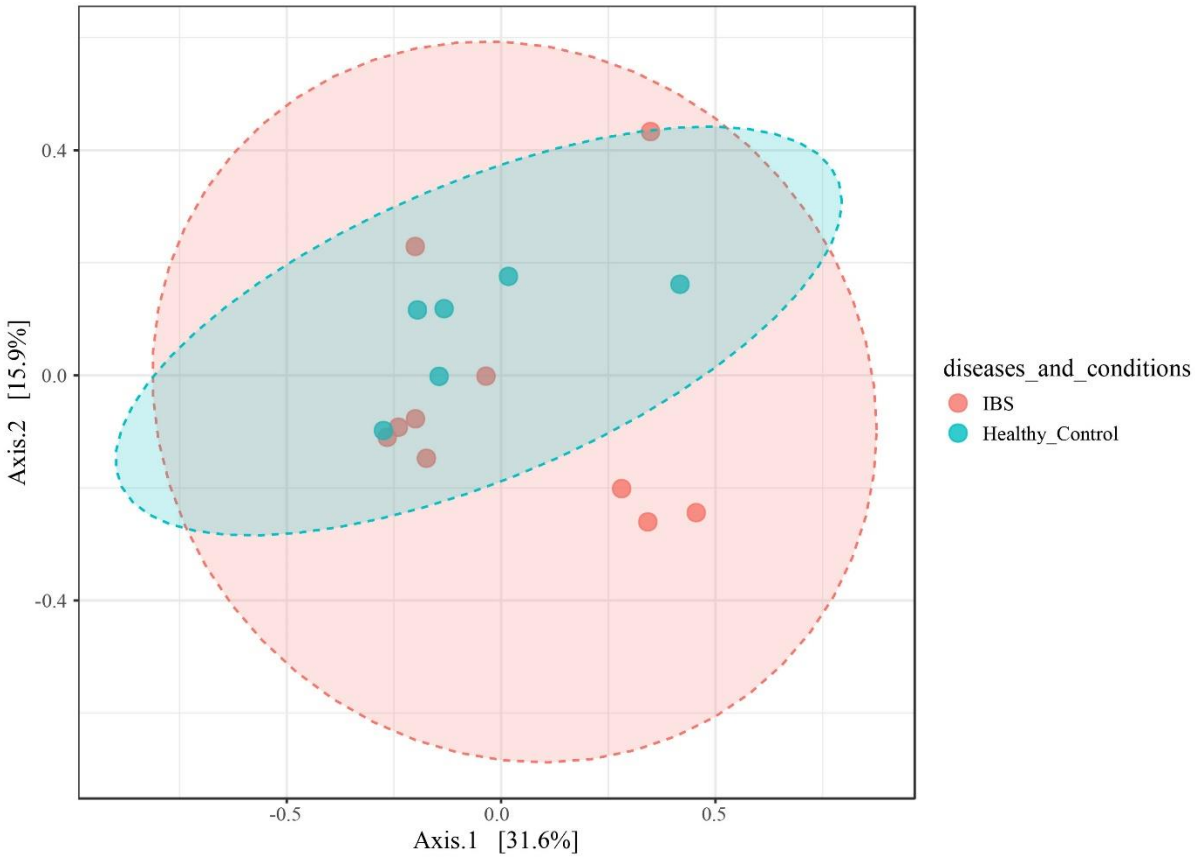


Figure 3.13: Beta diversity among IBS patients and the Healthy Control group. Principal Coordinates Analysis (PCoA) based on Bray-Curtis dissimilarity reveals distinct clustering of gut microbiome profiles between IBS patients (red) and healthy controls (teal). Each group is enclosed within a 95% confidence ellipse, indicating separation along Axis.1 (31.6%) and Axis.2 (15.9%), which together explain 47.5% of the total variance in microbial community composition.

3.10 Heatmap of the top-most abundant AMR genes

Antibiotic resistance gene analysis demonstrated the prevalence and quantity of antibiotic resistance among intestinal bacteria. An in-depth study identified the most significant number of resistance genes (**Figure 3.14**). Tetracycline was shown to have the highest baseline abundance of antimicrobial resistance genes. Following that, genes implicated in the macrolide, aminoglycoside, and β-lactam antibiotic classes were prevalent. There were no substantial variations in antimicrobial resistance gene classes between the guts of IBS patients and healthy controls. *TetQ*, *tetW*, and *tetM* were the most common tetracycline-related genes discovered. Furthermore, in the

macrolide class, ermF, *ermX*, and *mphA* genes were shown to be the most abundant. On the other hand, the *cfxA6* gene was shown to be significantly prevalent among beta-lactam classes.

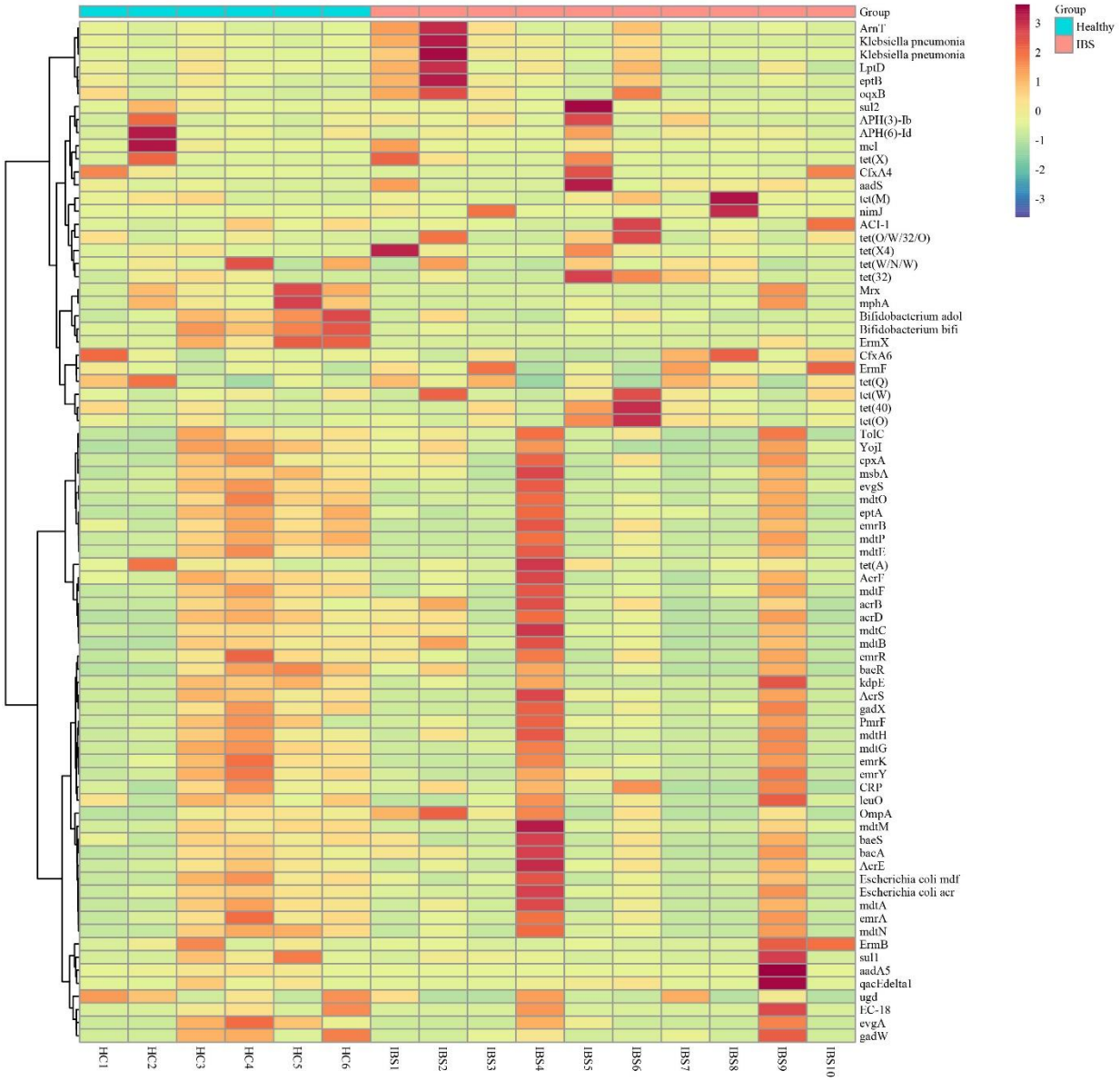


Figure 3.14: Heatmap of the top-most abundant AMR genes distributed among all samples. Heatmap showing the relative abundance of antimicrobial resistance genes (ARGs) across IBS and healthy control samples. Rows represent individual ARGs and taxa, while columns correspond to samples grouped by disease status. Color intensity reflects normalized abundance (Z-score), with red indicating higher and blue indicating lower expression. Hierarchical clustering reveals distinct ARG enrichment patterns in IBS-associated microbiomes.

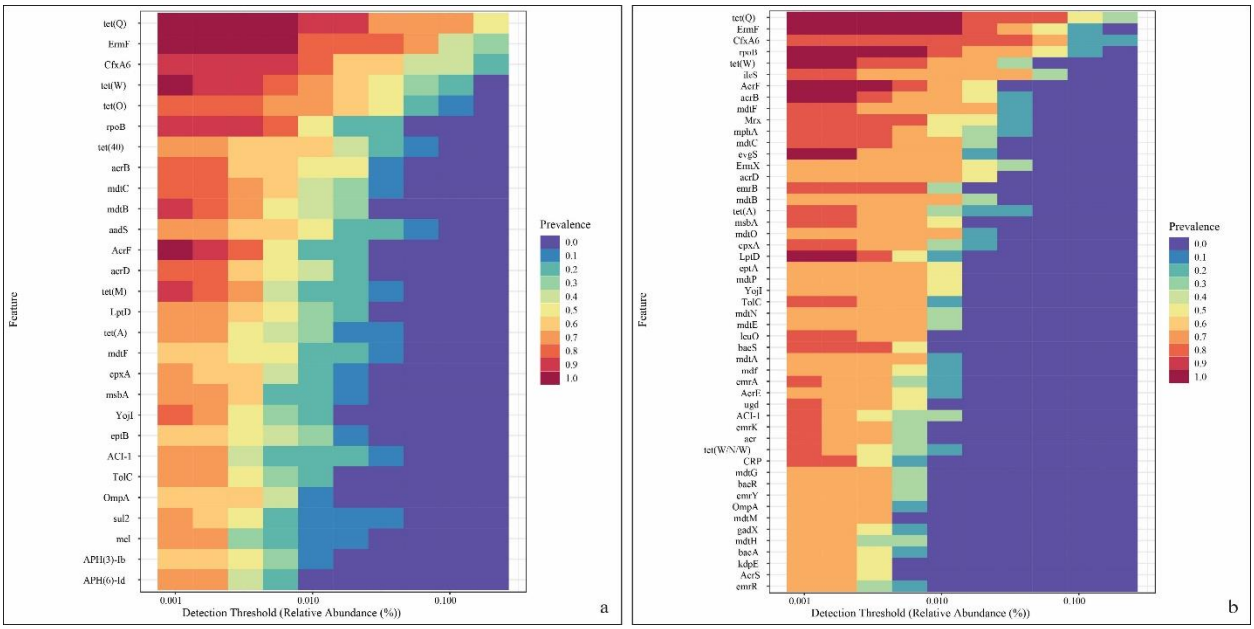


Figure 3.15: Top-most abundant AMR genes of HC(a) and IBS(b). Heatmaps illustrating feature prevalence across varying detection thresholds of relative abundance (%). Panel (a) and (b) display distinct sets of microbial or genetic features, with color gradients representing prevalence from low (blue) to high (red). Increasing detection thresholds reveal differential persistence of features, highlighting robust biomarkers and threshold-sensitive patterns relevant to disease stratification or microbial profiling.

The two heatmaps show a striking difference in antimicrobial resistance genes (ARGs) patterns between healthy persons **Fig. 3.15a** and those with IBS **Fig. 3.15b.** This shows a rather consistent and predictable resistome, most likely influenced by a healthy gut flora. In contrast, IBS samples have a substantially larger and more irregular ARGs profile. This variability suggests a disordered microbial ecology in which temporary or opportunistic bacteria carry resistance genes. The presence of these low-abundance ARGs markers in IBS could indicate enhanced microbial turnover or selective pressure. Research on antibiotic resistance genes has revealed the prevalence and quantity of antibiotic resistance in gut bacteria. A thorough investigation revealed the presence of the most significant number of resistance genes. Tetracycline was found to have the highest baseline abundance of antimicrobial resistance genes. Genes related to macrolides, aminoglycosides, and β-lactam antibiotics were prominent. There were no significant differences in antimicrobial resistance gene classes between the stomachs of IBS patients and those of healthy controls.

The most prevalent tetracycline-related genes found were *TetQ*, *TetW*, and *TetM*. Furthermore, within the macrolide class, the *ermF*, *ermX*, and *mphA* genes were found to be the most abundant (**Figure 3.16)**. The *cfxA6* gene, on the other hand, is much more common across beta-lactam classes. Our current study suggests that IBS is associated with a diverse and unstable resistome. It also emphasizes the necessity of selecting proper detection thresholds when evaluating microbiome data.

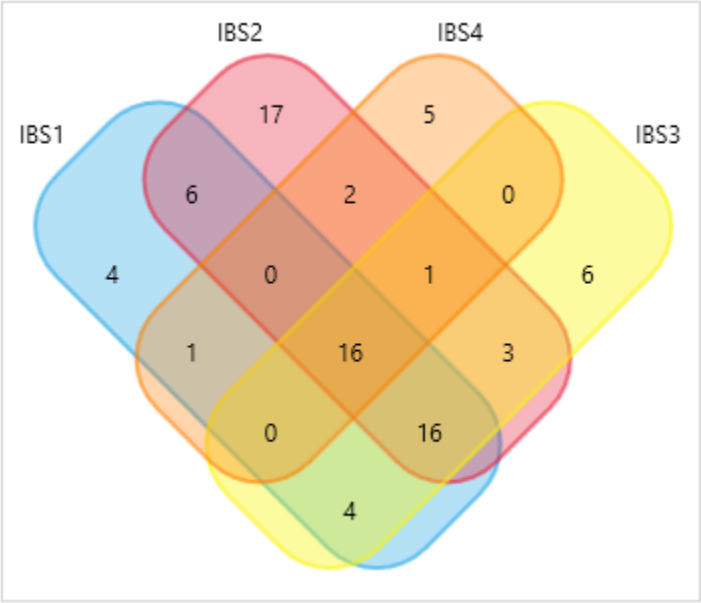


Figure 3.16: Common AMR genes present in all IBS patients. Four-set Venn diagram illustrating the shared and unique microbial features across IBS subtypes (IBS1–IBS4). Colored ellipses represent each subtype, with overlapping regions indicating feature co-occurrence. The central intersection highlights 16 features common to all subtypes, while peripheral zones reveal subtype-specific and pairwise-shared elements, supporting stratification of IBS-associated microbial signatures.

3.11 Comparative Antibiotic Resistance Profiles in IBS vs Healthy Controls

When comparing the antibiotic resistance profiles of the IBS and HC groups, the prevalence of certain antibiotic classes, specifically macrolide, tetracyclines, beta-lactam inactivation, and aminoglycoside inactivation, was marginally greater in the IBS group than in the HC group. Common AMR Genes present in all IBS samples are *tet(A), tet(Q), tet(O), cfxA, sul1, sul2, blaSHV, aadE, aph(6)-Id, cfxA6, blaTEM, qnrS,), oqxA, erm(F), oqxB, aph(3'')-Ib* (**Figure 3.17)**. On the

other hand, fluoroquinolone efflux is present in higher numbers in the case of the HC groups. This may occur due to the tendency to use antibiotics more frequently among IBS patients during their suffering days.

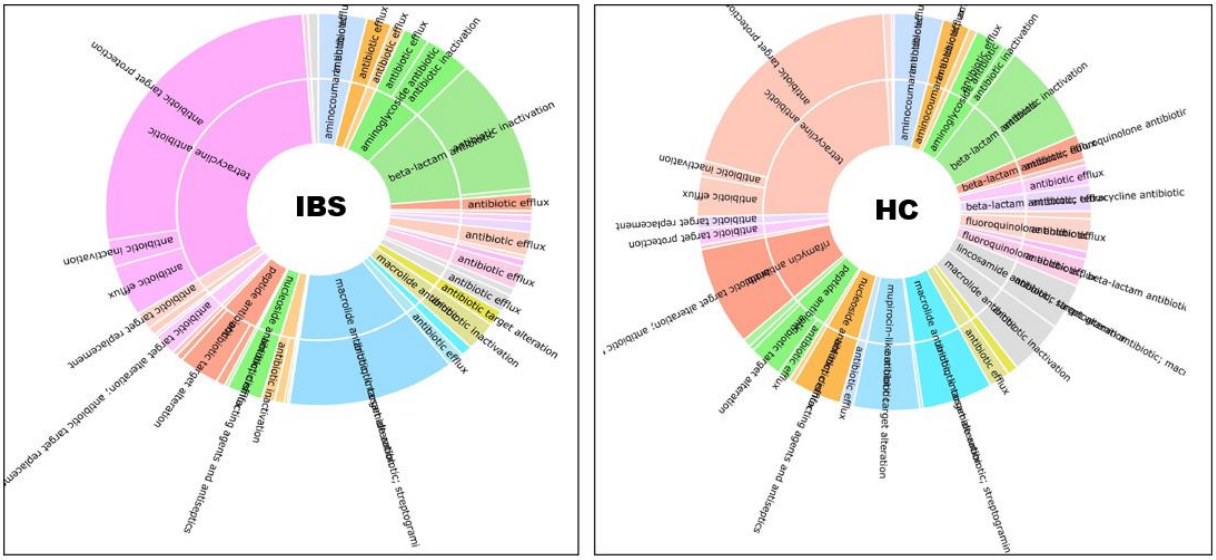


Figure 3.17: Antibiotic Resistance Profiles in IBS vs Healthy Controls. Circular taxonomic composition plots comparing gut microbiome profiles between IBS patients (left) and healthy controls (right). Segments represent bacterial taxa or resistance genes, with inner rings denoting higher-level classifications and outer rings showing specific features. Visual differences highlight shifts in microbial diversity and functional potential associated with IBS.

3.12 Functional Genomics and Virulence Factor Genes Analysis

Functional genomics and virulence factor investigation revealed a number of pathways and genes having probability to contribute in the gut condition of IBS. The relative abundance of functional pathways involved in metabolism, genetic information processing, environmental information processing, cellular processes, human diseases and organismal systems was determined. In depth analysis further revealed relative abundance of pathways and genes involved in digestive system, immune system, signaling molecules and interaction, environmental adaptation, carbohydrate metabolism, amino acid metabolism, biosynthesis of secondary metabolites, glycan biosynthesis and metabolism, endocrine and metabolic disease. Several genes such as *ure*A, *ure*B, *das*A, *das*B, *das*C, *alg*L, *acn*R, *hxl*A, *fit*A, *dev*R, *cef*D, *arg*H, *arg*A were detected to be less abundant in the gut

of IBS patients than healthy control. Virulence factor genes analysis also revealed several genes and proteins having virulent property playing potential role in gut environment.

3.13 KEGG Pathway and Network Analysis

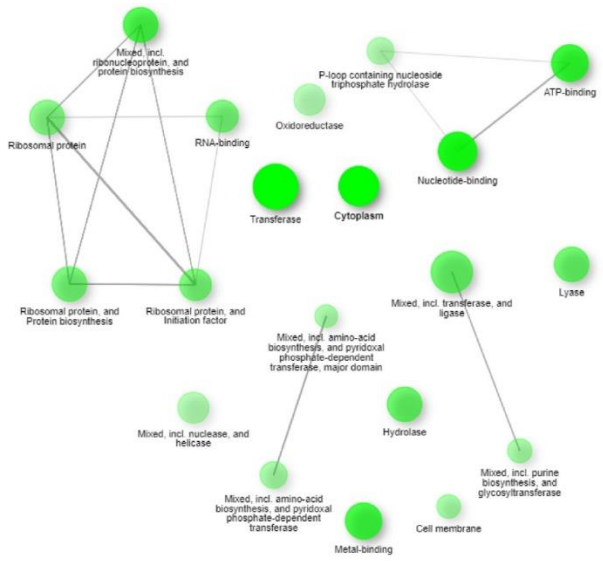


Figure 3.18: HC1 network analysis from KEGG pathway. Functional interaction network highlighting enriched molecular activities and biological processes in the dataset. Nodes represent annotated functions such as ribosomal proteins, RNA-binding, and ATP-binding, with edge connections indicating functional associations. Node size and color intensity reflect relative significance or abundance, revealing clustered domains of protein synthesis, nucleotide metabolism, and cellular localization.

A comparative analysis of KEGG pathway and Gene Ontology (GO) networks between HC and IBS samples uncovers distinct functional and structural patterns indicative of microbial dysbiosis in IBS. HC networks generally demonstrate a wider variety of paths and exhibit more modular interconnections. This indicates that microbial metabolism and homeostatic biological activities, including nutrition synthesis and immunological control, are in equilibrium. IBS networks, conversely, generally exhibit hubs that are either fragmented or excessively interconnected, along with an increased number of pathways associated with oxidative stress, sulfur metabolism, and immunological signaling, including responses mediated by Toll-like receptors and cytokines. GO annotations in IBS samples underscore the amplification of inflammatory processes, the disruption

of the epithelial barrier, and stress-induced enzymatic activity (**Figure 3.18**). HC samples, on the other hand, have stable enzymatic activity and keep biological components intact. These inconsistencies highlight the functional disruptions and host-microbe interactions related to IBS pathogenesis.

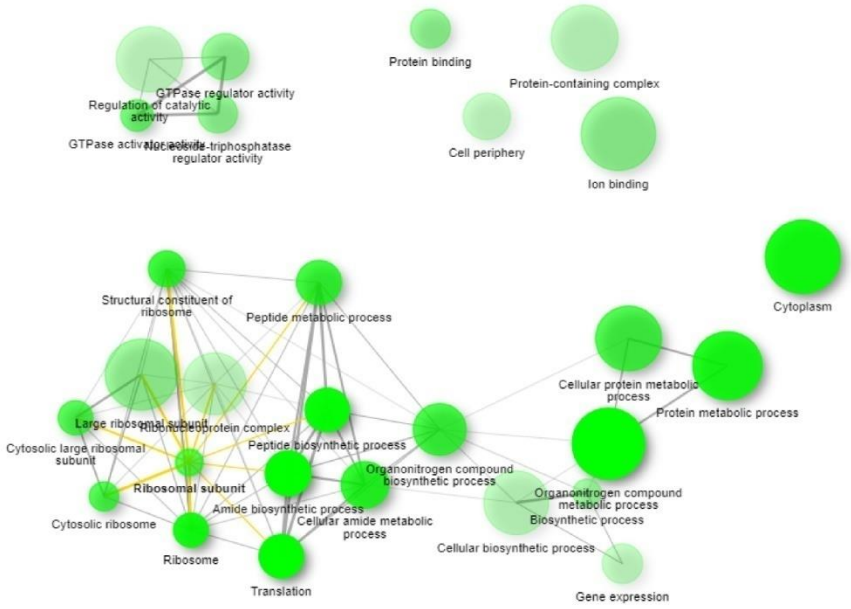


Figure 3.19: HC2 network analysis from KEGG pathway. Gene ontology network depicting enriched biological processes and molecular functions in the dataset. Nodes represent GO terms such as translation, ribosome, peptide metabolic process, and ATP-binding, with edge connections indicating functional relationships. Node size and color intensity reflect enrichment significance, highlighting clusters involved in protein synthesis, metabolic activity, and cellular localization.

In this investigation, we employ ShinyGO v0.741 (Ge et al., 2020) to examine the gene networks between the HC and IBS cohorts. The metabolic pathway network for the IBS group exhibits a highly intricate interconnection. Conversely, the HC has little connections among several metabolic activities. The network analysis revealed a higher prevalence of various metabolic processes in IBS patients (**Figure 3.19, 3.20, 3.21, 3.22, 3.23**). This may be ascribed to the presence of unique microbial communities and their corresponding metabolic activity in the gut microbiomes of IBS patients, which markedly differ from those in healthy control groups.

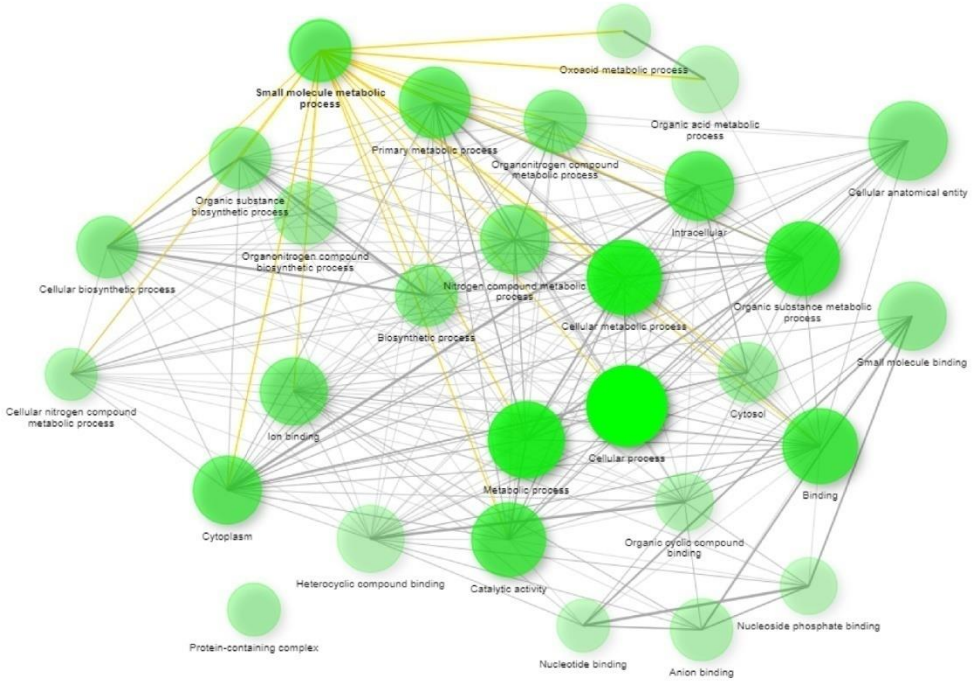


Figure 3.20: IBS1 network analysis from KEGG pathway. Gene ontology network illustrating enriched biological processes and molecular functions associated with IBS-related microbial activity. Nodes represent GO terms such as small molecule metabolism, nucleotide binding, and catalytic activity, with edge connections indicating functional relationships. Node size and color intensity reflect enrichment significance, revealing interconnected pathways in cellular metabolism, biosynthesis, and molecular binding.

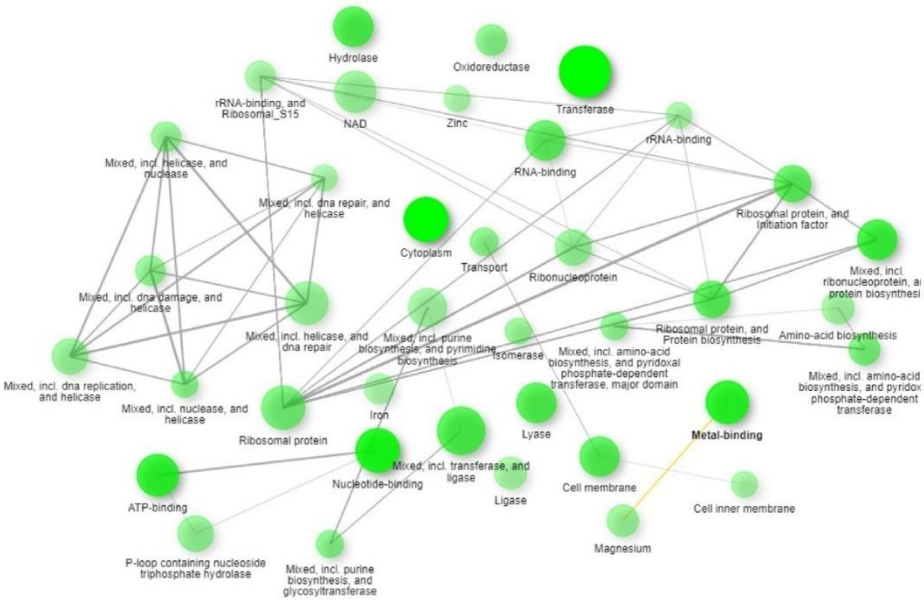


Figure 3.21: IBS2 network analysis from KEGG pathway. Network diagram illustrating enriched molecular functions and biological activities associated with IBS-related microbial signatures. Nodes represent functional categories such as hydrolase, oxidoreductase, RNA-binding, and ATP-binding, with edges denoting functional associations. Node size and color intensity reflect enrichment significance, while highlighted nodes (e.g., metal-binding, cell inner membrane) indicate key functional hubs relevant to microbial adaptation and host interaction.

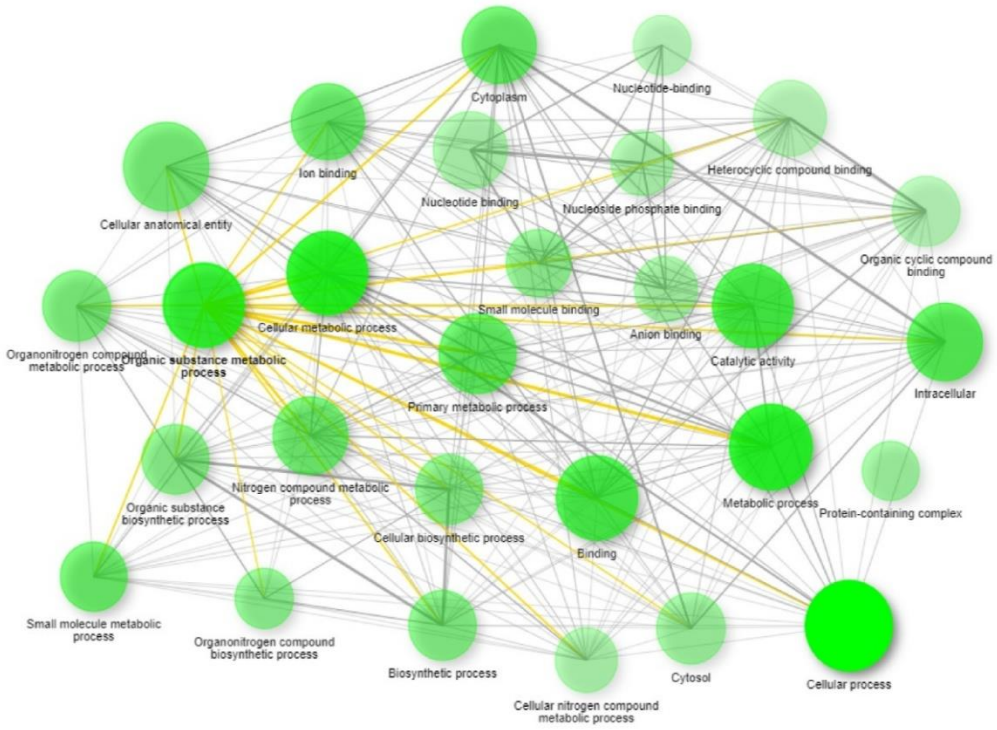


Figure 3.22: IBS3 network analysis from KEGG pathway. Gene ontology network illustrating enriched biological processes and molecular functions in IBS-associated microbial communities. Nodes represent GO terms such as cellular process, catalytic activity, and ion binding, with edges indicating functional relationships. Node size and color intensity reflect enrichment significance, while yellow edges highlight stronger associations, revealing interconnected pathways in metabolism, biosynthesis, and molecular binding.

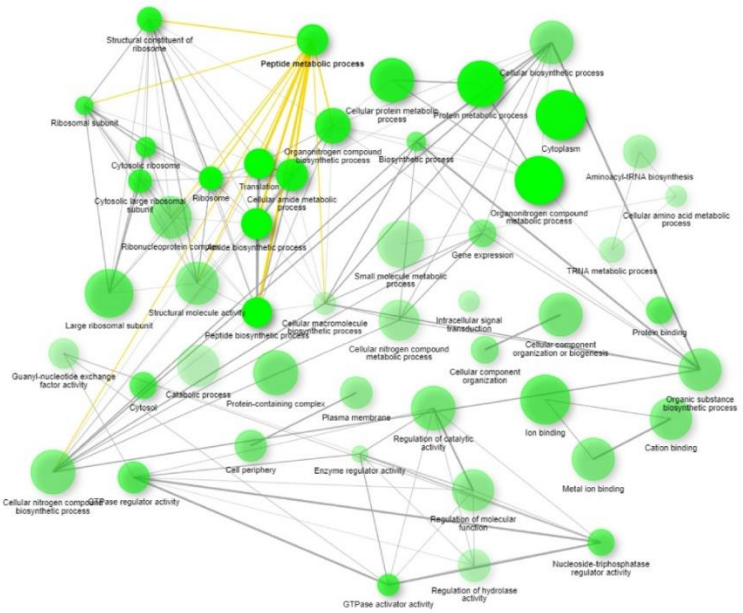


Figure 3.23: IBS4 network analysis from KEGG pathway. Gene ontology network illustrating enriched biological processes and molecular functions in IBS-associated microbial datasets. Nodes represent terms such as translation, gene expression, and peptide metabolic process, with edges indicating functional relationships. Node size and green color intensity reflect enrichment significance, while yellow edges highlight stronger associations, revealing coordinated activity in protein synthesis and cellular metabolism.

3.14 Most Relevant Metabolic Pathway Analysis

The bar chart depicts the comparative prevalence of ten significant microbial metabolic and biosynthetic pathways among six sample groups, comprising four IBS subtypes (IBS1-IBS4) and two healthy controls (HC1, HC2). The data reveal distinct patterns of microbial functional activity that may reflect underlying dysbiosis in IBS.

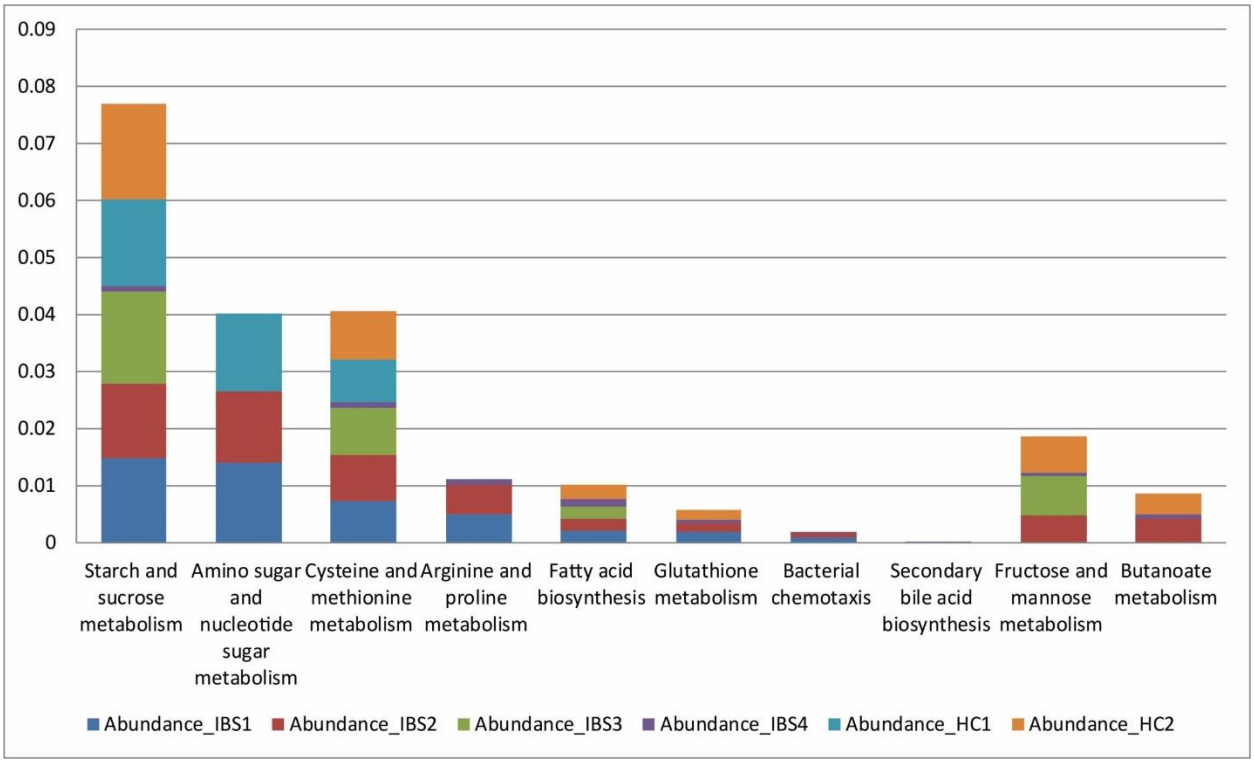


Figure 3.24: Elevated pathways in representative IBS samples. Bar chart comparing the relative abundance of key microbial metabolic pathways across IBS subtypes (IBS1–IBS4) and healthy controls (HC1–HC2). Pathways include carbohydrate, amino acid, and lipid metabolism, as well as bacterial chemotaxis and bile acid biosynthesis. Colored segments represent sample groups, revealing functional shifts in microbial activity associated with IBS phenotypes.

In IBS samples (especially IBS1 and IBS2), starch and sucrose metabolism had the highest cumulative abundance among all metabolic responses. This suggests that microbial carbohydrate digestion is enhanced in IBS, potentially resulting in altered fermentation and gas production. Amino sugar and nucleotide sugar metabolism are also significantly higher in IBS groups, particularly IBS2 and IBS3 (**Figure 3.24)**. This could indicate increased mucin breakdown or host glycan intake, which is typically associated with gut barrier disruption. IBS samples have moderately higher levels of cysteine and methionine metabolism, as well as arginine and proline metabolism, indicating a disruption in the sulfur and nitrogen cycles. These pathways participate in both oxidative stress responses and microbial-host interactions.

3.15 Modified Microbial Redox and Metabolic Pathways Fundamental IBS Symptoms and Dysbiosis

Glutathione metabolism, a marker of redox balance, is low in IBS, indicating a decrease in microbial antioxidant ability. Bacterial chemotaxis and secondary bile acid synthesis are underrepresented, although slightly more common in HC samples. This could indicate improved microbial motility and bile acid transformation in the controls, both of which are frequently disrupted in IBS. Fructose, mannose, and butanoate metabolism is more uniformly distributed, with HC samples (particularly HC2) showing slightly greater amounts. Butanoate (butyrate) synthesis is essential for colonic health, and its presence in HC may imply increase epithelial support. Bloating and pain are frequent symptoms of IBS. These symptoms are brought on by enhanced glucose and amino sugar metabolism. Changes in amino acid pathways may affect microbial signaling and host immune regulation, resulting in decreased bile acid and butyrate-related activity, indicating poor microbial homeostasis and epithelial support. Dysbiosis traits include increased fermentative activity, disruption of the mucosal barrier, and a decrease in the synthesis of anti-inflammatory metabolites. Additional taxonomic and statistical validation would strengthen these conclusions.

3.16 Comparison with global data

A large dataset of 1109 people comprising 617 IBS patients and 492 healthy individuals were analyzed in our current study where Sixteen samples from our sequenced samples from Bangladeshi people. The evaluation of microbial diversity by 16S rRNA sequencing depends on sequencing depth. A sample with increased sequencing depth is more likely to exhibit greater diversity than one with less sequencing depth. To address this problem rarefaction is the process which was utilized to make harmonized samples from wide range of sequencing data. As we are analyzing large number of global data and the sequencing depth are varied in different sequencing machines and time frame. Before going into the core analysis rarefaction was performed to get uniform data. The figure below shows the rarefied data for both IBS and HC samples.

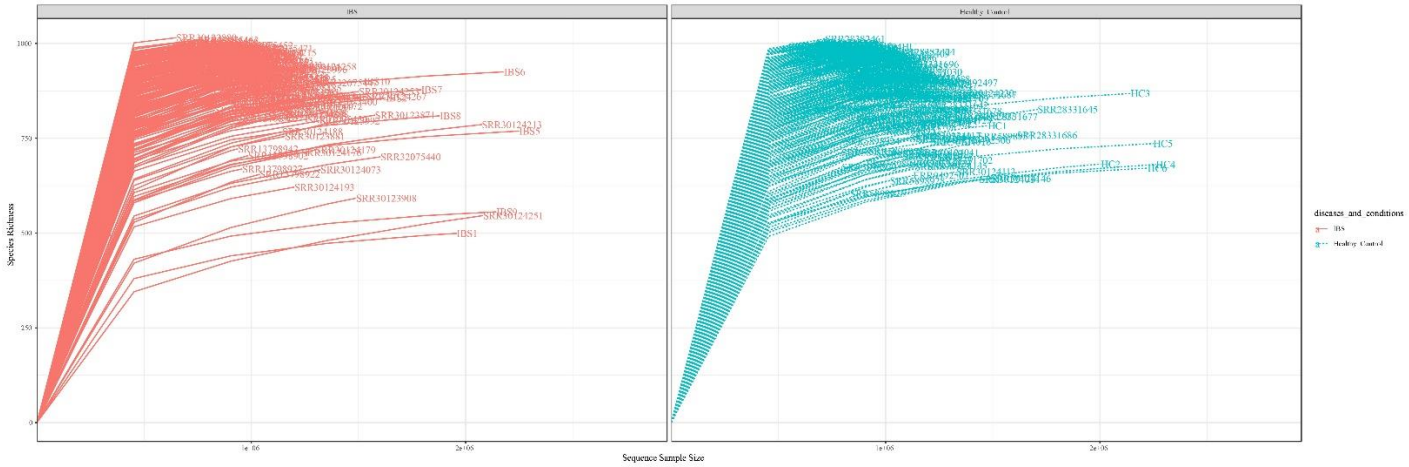


Figure 3.25: Rarefaction curve of IBS and HC samples from global data. Line plots comparing sequencing depth across sample sizes for IBD (left, red) and multi-control (right, blue) groups. Each line represents an individual sample, showing the relationship between sequence sample size and total sequence count. The steeper curve in IBD samples suggests higher microbial richness or sequencing saturation compared to controls, highlighting group-specific diversity dynamics.

After that, we have observed the Alpha diversity of the total samples to get insight about the sample's richness (the total number of different species present) and evenness (how evenly individuals are distributed among those species). During the Alpha diversity measures we have used different diversity methods to compare our data and all the methods shows similar diversity pattern. In our analysis we have utilized the Chao1, Shannon, Simpson, Observed methods of alpha diversity analysis programs (**Figure 3.26)**. From our study we have found that, the HC group is well distributed and the samples were evenly distributed in the diversity plot created by using R programing. As all we know the high alpha diversity indicates a healthy ecosystem by its nature. From the Principal Component Analysis (PCA) which is the method of simplifying complex datasets for making efficient and easier interpretation, we can see that the HC and IBS samples clustered distinctly in the distribution plot (**Figure 3.27)**. It indicates the gut microbiome pattern of IBS and HC group is distinctly differed from each other among the population resides all over the world.

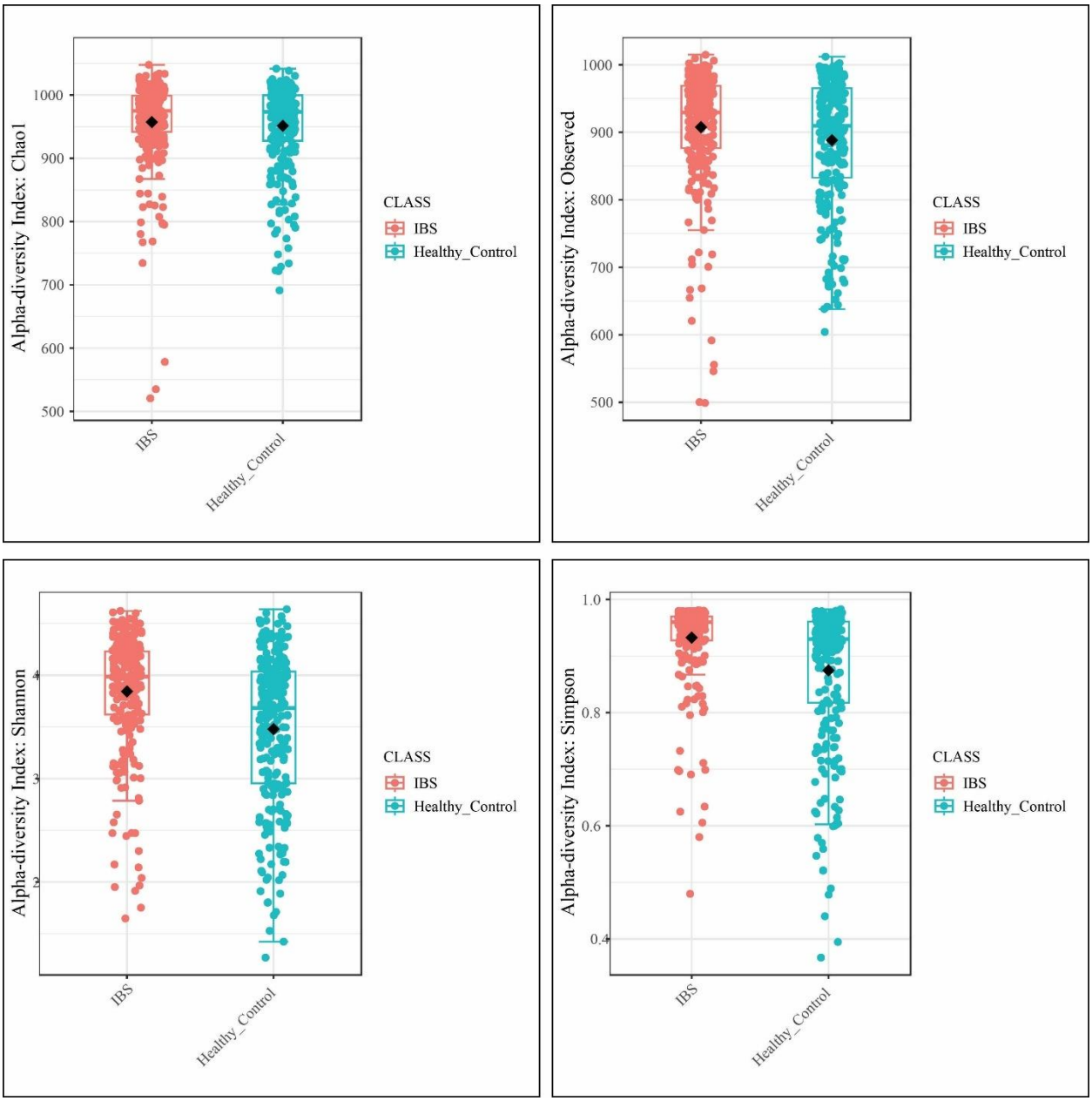


Figure 3.26: Alpha diversity plot of global dataset. Box plots comparing alpha-diversity indices (Chao1, Observed, Shannon, Simpson) between IBS patients (red) and healthy controls (blue). Each plot summarizes microbial richness and evenness across groups, with black diamonds indicating mean values. Reduced diversity in IBS samples suggests altered gut microbial ecology potentially linked to disease pathology.

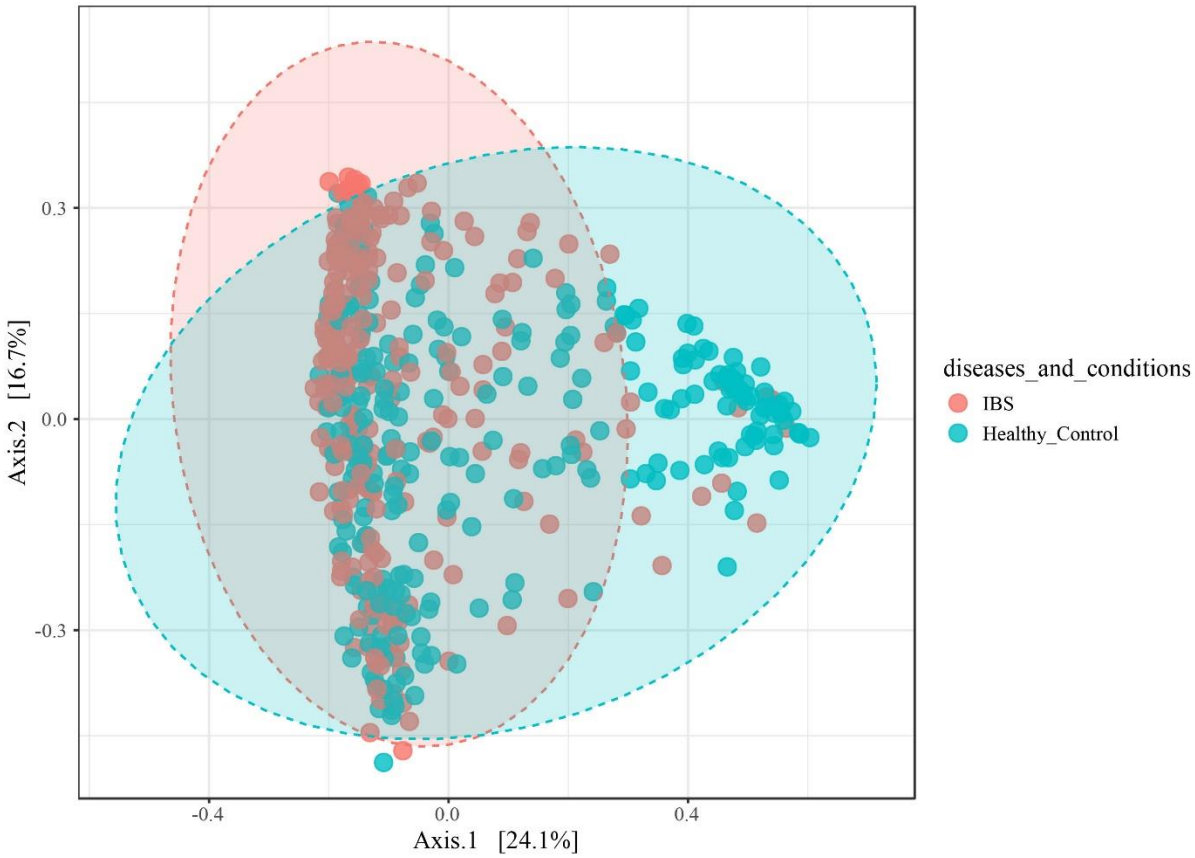


Figure 3.27: PCoA analysis between IBS and HC from global dataset. Principal Coordinates Analysis (PCoA) based on microbial community composition reveals distinct clustering between IBS patients (red) and healthy controls (teal). Axis.1 and Axis.2 explain 24.1% and 16.7% of the total variance, respectively. Dashed ellipses indicate group dispersion, highlighting compositional shifts in gut microbiota associated with IBS.

3.17 Core Microbiome Analysis of Global population:

The core microbiome is the group of bacteria that are always present in a certain environment or host across several samples. This notion is vital in microbial ecology as it aids in identifying the fundamental microbial communities that enhance the stability and functionality of ecosystems. Core microbiome analysis is a useful way to learn about microbial communities and how they fit into the ecosystem. Researchers may learn a lot about how microbial ecosystems work and how they affect health and the environment by using established methods and strict statistical analysis. From core microbiome analysis between IBS and HC samples, we can see that *Segatella copri and Bacteroides xylanisolvens* are the core bacteria present in the HC groups among the global

population taxonomic class. These organisms are statistically significant in number, and their presence contributes to the stability of the gut environment in healthy individuals. On the contrary, *Anaerobutyricum halli, Anaerostipes hadrus, Blautia obeum, Blautia wexlerae, Collinsella aerofaciens, and Roseburia intestinalis* are organisms that are present in large numbers in IBS patients' samples, as analyzed in a large dataset of 1109 samples from around the world (**Figure 3.28)**. This indicates that the presence of these particular species may facilitate gut dysbiosis by encouraging the proliferation of pathogenic bacteria and impeding the growth of good bacteria.

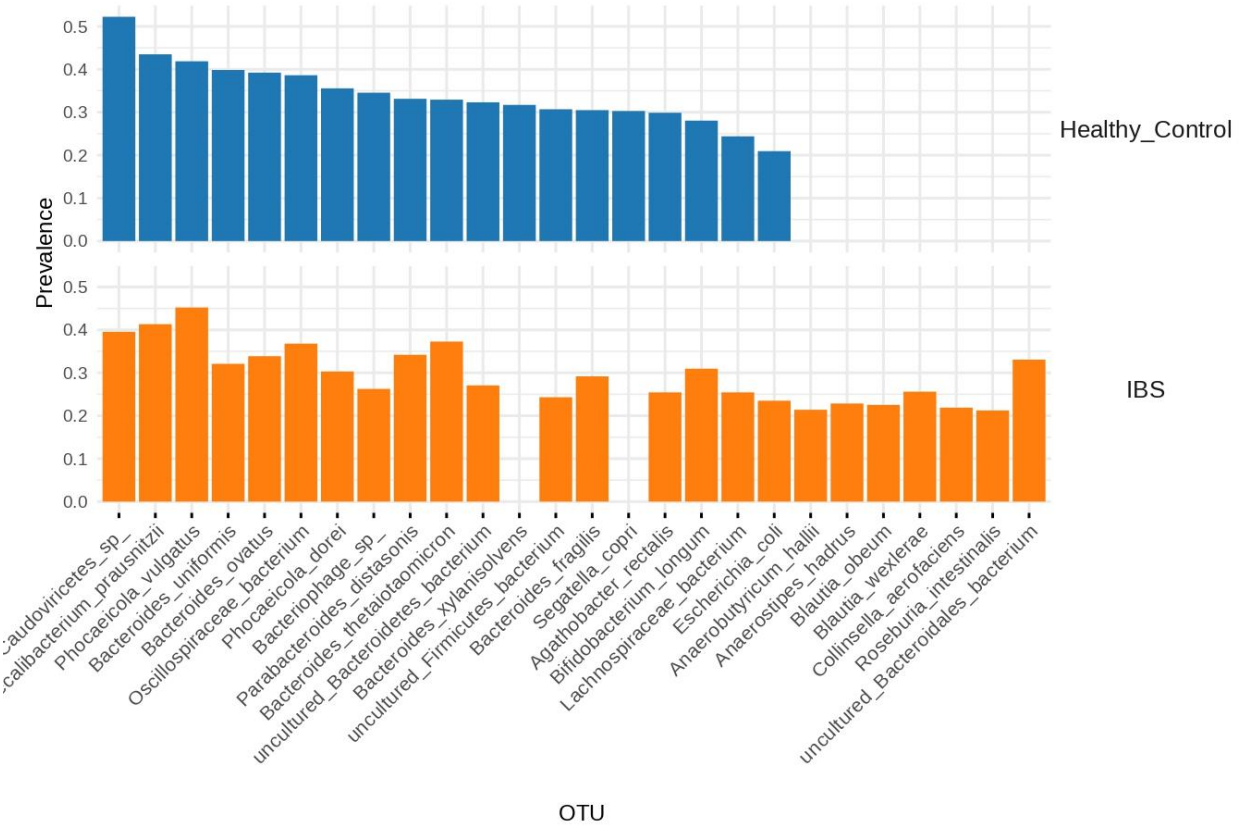


Figure 3.28: Core Microbiome Analysis of Global population. Bar chart comparing the prevalence of key bacterial OTUs (shotgun metagenomics) between healthy controls (blue) and IBS patients (orange). Each OTU is labeled along the x-axis, with bar height reflecting relative prevalence. Notable shifts in taxa such as *Bacteroides fragilis*, *Parabacteroides merdae*, and *Escherichia coli* highlight microbial compositional differences associated with IBS.

3.18 Results of MAGs

3.18.1 Sequencing and Assembly

Each sample generated 20 to 150 million paired-end reads. Following trimming and host removal, only microbial sequence reads remained. The Phred score and FastQC results verified that the data obtained from our sequencing were all of good quality. After evaluating the read quality, we compiled our sequences for the assembled genome.

3.18.2 MAG Recovery and Quality

A total of 392 MAGs were reconstructed: 153 from HC and 329 from IBS samples. In our recent investigation, we discovered that *Segatella copri* MAG is detected in both HC and IBS samples in higher numbers. It is present in 15 MAGs, followed by *Megasphaera elsdenii* (14), *Escherichia coli* (13), *Flavonifractor plautii* (11), and *Faecalibacterium* sp. (11). These are the majority of the species recovered in our current investigation. *Megasphaera elsdenii*, *Escherichia coli*, *Ligilactobacillus ruminis*, *Sutterella wadsworthensis*, and *Collinsella aerofaciens* were also observed to be recovered in a larger ratio in HC than in IBS samples. *Bifidobacterium longum*, *Bifidobacterium breve*, *Bacteroides fragilis*, *Campylobacter upsaliensis*, *Streptococcus infantarius*, and *Alistipes finegoldii* are the only MAGs found from an HC sample. In IBS samples, however, the presence of *Klebsiella pneumoniae*, *Roseburia intestinalis*, *Oscillibacter* sp., and *Bacteroides uniformis* is more common. *Bifidobacterium adolescentis*, *Klebsiella quasipneumoniae*, *Sutterella megalosphaeroides*, *Bacteroides ovatus*, and *Parabacteroides merdae* are the only MAGs found from IBS samples. The recovery of these MAGs corresponds to the results of the taxonomy classification comparison. This may have occurred in a manner where the recovery of MAGs is dependent on the sequential repetition of the samples. During this comparison, each MAG occurrence is not counted separately. This systematic count enables a precise, quantitative comparison of microbial composition between the two groups (HC and IBS).

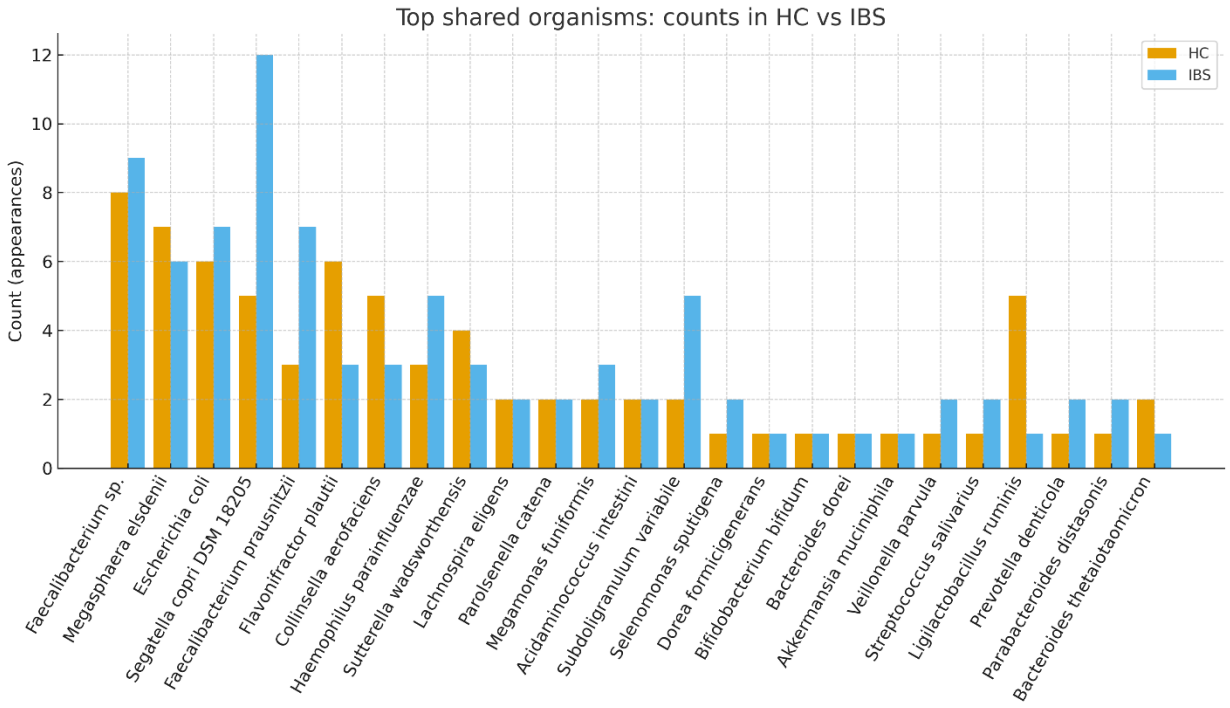


Figure 3.29: Top-most abundant MAGs recovered between HC and IBS groups. Bar chart displaying counts of top shared microbial organisms between healthy controls (orange) and IBS patients (blue). Each organism is represented along the x-axis, with bar height indicating detection frequency. Differential abundance of taxa such as *Faecalibacterium prausnitzii*, *Escherichia coli*, and *Collinsella aerofaciens* highlights compositional shifts in gut microbiota associated with IBS.

Forty-two met high-quality standards, while 83 met medium-quality standards. The average completeness was 87.4% (SD = 5.6), and the contamination rate was 2.3%. HC samples produced more varied MAGs, indicating increased microbial biodiversity.

Table 3.2: MAGs recovered from HC and IBS groups

Organism	Count in HC	Count in IBS	Total Count
Segatella copri DSM 18205	7	8	15
Megasphaera elsdenii	9	5	14
Escherichia coli	8	5	13
Flavonifractor plautii	5	6	11

Organism	Count in HC	Count in IBS	Total Count
Faecalibacterium sp.	7	4	11
Collinsella aerofaciens	6	3	9
Sutterella wadsworthensis	6	2	8
Ligilactobacillus ruminis	6	1	7
Haemophilus parainfluenzae	3	4	7
Subdoligranulum variabile	3	4	7
Klebsiella pneumoniae	1	5	6
Megamonas funiformis	3	3	6
Faecalibacterium prausnitzii	2	4	6
Bacteroides fragilis	3	0	3
Streptococcus infantarius	3	0	3
Veillonella parvula	2	1	3
Acidaminococcus intestini	2	1	3
Bifidobacterium longum	3	0	3
Bifidobacterium breve	3	0	3
Campylobacter upsaliensis	2	0	2
Parolsenella catena	2	1	3
Selenomonas sputigena	2	1	3
Catenibacterium mitsuokai	1	2	3
Succinivibrio dextrinosolvens	1	2	3
Roseburia intestinalis	1	3	4

Organism	Count in HC	Count in IBS	Total Count
Oscillibacter sp.	1	3	4
Bacteroides uniformis	1	3	4
Bifidobacterium adolescentis	0	3	3
Parabacteroides distasonis	1	2	3
Klebsiella quasipneumoniae	0	2	2
Sutterella megalosphaeroides	0	2	2
Bacteroides ovatus	0	2	2
Parabacteroides merdae	0	2	2
Alistipes finegoldii	2	0	2
Lachnospira eligens	2	1	3
Prevotella denticola	2	1	3
Bifidobacterium bifidum	1	1	2
Eubacterium eligens	1	1	2
Veillonella sp.	1	0	1
Ruminococcus gnavus	1	0	1
Bacteroides thetaiotaomicron	2	1	3
Bacteroides dorei	1	1	2
Akkermansia muciniphila	1	1	2
Enterococcus faecalis	2	0	2
Streptococcus salivarius	1	1	2
Dorea formicigenerans	1	1	2

Organism	Count in HC	Count in IBS	Total Count
Eubacterium ventriosum	0	1	1
Roseburia hominis	0	2	2
Butyrivibrio crossotus	0	1	1
Granulicatella adiacens	0	1	1
Dorea longicatena	0	1	1
Paeniclostridium sordellii	0	1	1
Acidaminococcus fermentans	0	1	1
Ruminococcus bicirculans	0	1	1
Ruthenibacterium lactatiformans	0	1	1
Veillonella atypica	0	1	1
Dysosmobacter welbionis	0	1	1
Alistipes shahii	0	1	1
Alistipes onderdonkii	0	1	1
Lactobacillus mucosae	1	0	1
Eubacterium callanderi	1	0	1
Morganella morganii	1	0	1
Alistipes indistinctus	1	0	1
[Clostridium] innocuum	1	0	1
Alistipes communis	0	1	1
Veillonella dispar	0	1	1
Paraprevotella xylaniphila	0	1	1

Organism	Count in HC	Count in IBS	Total Count
Sutterella sp	0	1	1
Streptococcus sp.	0	1	1

3.18.3 Taxonomic Patterns among Curated MAGs

An initial comparative assessment of metagenome-assembled genomes (MAGs) derived from the same species in IBS patients and healthy controls revealed several notable differences. For *Segatella copri*, two MAGs were selected: one from a healthy control sample (HC2) and another from an IBS patient sample (IBS1). The comparative analysis between these two genomes highlighted distinct genomic features. *S. copri* HC2 and IBS1 consisted of 76 and 436 contigs, respectively (**Table 3.3 A**). Although neither genome contained any major virulence factors, both carried antimicrobial resistance determinants. Notably, the HC2 strain encoded 21 antimicrobial resistance genes, whereas the IBS1 strain harbored nearly double that number, with 41 such genes identified.

Similarly, as *S. copri*, two MAGs of *Megasphaera elsdenii* were selected for comparative analysis. One from a healthy control sample (HC2) *Megasphaera elsdenii* HC2 and another from an IBS patient sample (IBS2) *Megasphaera elsdenii* IBS2. The comparative analysis between these two genomes highlighted distinct genomic features. *M. elsdenii* HC2 and IBS1 consisted of 379 and 328 contigs, respectively (**Table 3.3 B**). The genome quality was good for both genomes. The number of hypothetical CDS of *M. elsdenii* HC2 and IBS1 were 414 and 441 respectively. Although neither genome contained any major virulence factors, both carried antimicrobial resistance determinants. Notably, the HC2 strain encoded 17 antimicrobial resistance genes, whereas the IBS2 strain harbored 23 such genes identified.

Similarly, in case of the MAGs of *E. coli*, the MAGs reassembled from the sequences of IBS patients were predicted to be more virulent than the MAG strains reassembled from healthy controls. For instance, the comparative analysis in between two *E. coli* strains, one from healthy control 4 (*E. coli* HC4) and another from IBS patient 9 (*E. coli* IBS9) showed striking features. Interestingly, Virulence factor genes analysis showed that *E. coli* HC4 harbored 144 virulence factor genes, on the other hand *E. coli* IBS9 harbored 267 virulence factor genes (**Table 3.3 C**). In

case of antimicrobial resistance genes, *E. coli* HC4 had no significant antimicrobial resistance genes in it. On the other hand, *E. coli* IBS9 was found to have the presence of 62 antimicrobial resistance genes in it's genome. In case of *Faecalibacterium* sp. HC4, only 1 antimicrobial resistance gene was found and in *Faecalibacterium* sp. IBS9, 23 antimicrobial resistance genes were detected.

In case of the MAGs reassembled of *Streptococcus salivarius* (*S. salivarius*), S. *salivarius* HC3 (MAG from healthy control) harbored 38 and 19 virulence factor genes and antimicrobial resistance genes respectively. On the other hand, in case of *S. salivarius* IBS5 (MAG from IBS patient) harbored 54 virulence factor genes and 31 antimicrobial resistance genes (**Table 3.3 H**). Other than these MAGs, *Prevotella denticola, Lachnospira eligens, Ligilactobacillus ruminis* showed striking features which indicate toward the contrasting features between the MAGs assembled from IBS patients and healthy controls.

Table 3.3 (A): Comparative genomic features of MAGs of *Segatella copri*

RAST Annotation	***Segatella copri* HC2**	***Segatella copri* IBS1**
Genome Statistics		
Contigs	76	436
Genome Length	3329822	6601142
GC Content	45.334106	46.11987
Contig L50	15	63
Contig N50	68478	33516
Annotation Statistics		
tRNA	65	119
rRNA	4	12
CDS	2844	5792
CDS Ratio	0.8540997	0.87742394
Hypothetical CDS	1293	2687
Hypothetical CDS Ratio	0.5196906	0.52537984
Genome Quality		
Coarse Consistency	99	96.9

Fine Consistency	98.5	70.2
CheckM Completeness	100	100
CheckM Contamination		50.8
Genome Quality	Good	Poor
Virulence Factor (Victors)	0	0
Antibiotic Resistance (PATRIC)	21	41

Table 3.3 (B): Comparative genomic features of MAGs of *Megasphaera elsdenii*

RAST Annotation	***Megasphaera elsdenii* HC2**	***Megasphaera elsdenii* IBS2**
Genome Statistics		
Contigs	379	328
Genome Length	2015033	2080339
GC Content	51.92982	51.85078
Contig L50	93	82
Contig N50	6809	8133
Annotation Statistics		
tRNA	29	34
rRNA	3	
CDS	2027	2048
CDS Ratio	1.0059389	0.98445493
Hypothetical CDS	414	441
Hypothetical CDS Ratio	0.28367046	0.29541016
PLFAM CDS	1897	1932
PLFAM CDS Ratio	0.9358658	0.9433594
Genome Quality		
Coarse Consistency	97.9	96.5
Fine Consistency	95.4	94.5
CheckM Completeness	97	97.6

CheckM Contamination	5.1	3.6
Genome Quality	Good	Good
Virulence Factor (Victors)	0	0
Antibiotic Resistance (PATRIC)	17	23

Table 3.3 (C): Comparative genomic features of MAGs of *Escherichia coli*

RAST Annotation	***Escherichia coli* HC4**	***Escherichia coli* IBS9**
Genome Statistics		
Contigs	713	487
Genome Length	6405495	5679373
GC Content	50.57595	51.009415
Contig L50	96	62
Contig N50	19396	22402
Annotation Statistics		
tRNA	98	86
rRNA	8	7
CDS	7101	6071
CDS Ratio	1.1085794	1.068956
Hypothetical CDS	974	732
Hypothetical CDS Ratio	0.28657934	0.28463185
PLFAM CDS	6720	5822
PLFAM CDS Ratio	0.94634557	0.9589853

Genome Quality		
Coarse Consistency	99.1	99.3
Fine Consistency	88.1	90.3
CheckM Completeness	99.4	99
CheckM Contamination	12.6	11.8
Genome Quality	Poor	Good
Virulence Factor (Victors)	144	267
Antibiotic Resistance (PATRIC)	0	62

Table 3.3 (D): Comparative genomic features of MAGs of *Faecalibacterium* sp.

RAST Annotation	***Faecalibacterium* sp. HC4**	***Faecalibacterium* sp.IBS9**
Genome Statistics		
Contigs	197	612
Genome Length	969192	3151376
GC Content	47.625237	58.90954
Contig L50	42	132
Contig N50	7112	6779
Annotation Statistics		
tRNA	11	55
rRNA	1	1
CDS	1342	3461

CDS Ratio	1.3846586	1.0982504
Hypothetical CDS	906	1192
Hypothetical CDS Ratio	0.72727275	0.40681884
PLFAM CDS	664	3021
PLFAM CDS Ratio	0.4947839	0.87286913
Genome Quality		
Coarse Consistency	93.9	96.1
Fine Consistency	93.9	85.8
CheckM Completeness		54.2
CheckM Contamination		13.3
Genome Quality	Good	Good
Virulence Factor (Victors)	0	0
Antibiotic Resistance (PATRIC)	1	23

Table 3.3 (E): Comparative genomic features of MAGs of *Prevotella denticola*

RAST Annotation	***Prevotella denticola* HC1**	***Prevotella denticola* IBS9**
Genome Statistics		
Contigs	112	71
Genome Length	3287733	2721865
GC Content	49.72694	51.189938
Contig L50	21	15

Contig N50	41765	52719
Annotation Statistics		
tRNA	62	57
rRNA	5	6
CDS	2620	2222
CDS Ratio	0.7969017	0.816352
Hypothetical CDS	1055	909
Hypothetical CDS Ratio	0.4744275	0.4788479
PLFAM CDS	2290	1935
PLFAM CDS Ratio	0.8740458	0.8708371
Genome Quality		
Coarse Consistency	98.3	98.2
Fine Consistency	94.3	97.7
CheckM Completeness	97.6	97.6
CheckM Contamination	5.5	1
Genome Quality	Good	Good
Virulence Factor (Victors)	0	0
Antibiotic Resistance (PATRIC)	23	19

Table 3.3 (F): Comparative genomic features of MAGs of *Ligilactobacillus ruminis*

RAST Annotation	***Ligilactobacillus ruminis* HC4**	***Ligilactobacillus ruminis* IBS10**
Genome Statistics		
Contigs	77	67
Genome Length	1982485	1979686
GC Content	44.007244	43.76987
Contig L50	17	12
Contig N50	45376	54985
Annotation Statistics		
tRNA	26	39
rRNA	3	3
CDS	2146	2174
CDS Ratio	1.0824798	1.098154
Hypothetical CDS	691	693
Hypothetical CDS Ratio	0.3863001	0.31876725
PLFAM CDS		
PLFAM CDS Ratio		
Genome Quality		
Coarse Consistency	99.1	
Fine Consistency	98.7	
CheckM Completeness	100	

CheckM Contamination	4	
Genome Quality	Good	Good
Virulence Factor (Victors)	0	0
Antibiotic Resistance (PATRIC)	21	25

Table 3.3 (G): Comparative genomic features of MAGs of *Lachnospira eligens*

RAST Annotation	***Lachnospira eligens* HC3**	***Lachnospira eligens* IBS7**
Genome Statistics		
Contigs	24	246
Genome Length	2326881	2630751
GC Content	35.888298	37.803997
Contig L50	4	49
Contig N50	159054	15957
Annotation Statistics		
tRNA	24	37
rRNA		
CDS	2092	2639
CDS Ratio	0.89905757	1.0031356
Hypothetical CDS	760	804
Hypothetical CDS Ratio	0.42638624	0.37438422
PLFAM CDS	1904	2194

PLFAM CDS Ratio	0.91013384	0.83137554
Genome Quality		
Coarse Consistency	92.1	98.3
Fine Consistency	91.4	92.2
CheckM Completeness	75.8	97.7
CheckM Contamination	1.5	14.3
Genome Quality	Good	Good
Virulence Factor (Victors)	1	3
Antibiotic Resistance (PATRIC)	20	23

Table 3.3 (H): Comparative genomic features of MAGs of *Streptococcus salivarius*

RAST Annotation	***Streptococcus salivarius* HC3**	***Streptococcus salivarius* IBS5**
Genome Statistics		
Contigs	212	418
Genome Length	1376650	1910296
GC Content	41.471615	40.5634
Contig L50	31	101
Contig N50	11160	5588
Annotation Statistics		
tRNA	7	35

rRNA	1	1
CDS	1448	2071
CDS Ratio	1.0518287	1.0841252
Hypothetical CDS	258	403
Hypothetical CDS Ratio	0.26104972	0.27184933
PLFAM CDS	1369	1960
PLFAM CDS Ratio	0.94544196	0.9464027
Coarse Consistency	87.4	97.1
Fine Consistency	83.1	90.8
CheckM Completeness	87.8	91.6
CheckM Contamination	20	19
Genome Quality	Good	Good
Virulence Factor (Victors)	38	54
Antibiotic Resistance (PATRIC)	19	31

3.18.3 Genome Visualization

The genetic differences were depicted by Proksee maps of representative MAGs are shown below (**Figure 3.30 A, Figure 3.30 B, Figure 3.31 A, Figure 3.31 B, Figure 3.32 A, Figure 3.32 B)**.

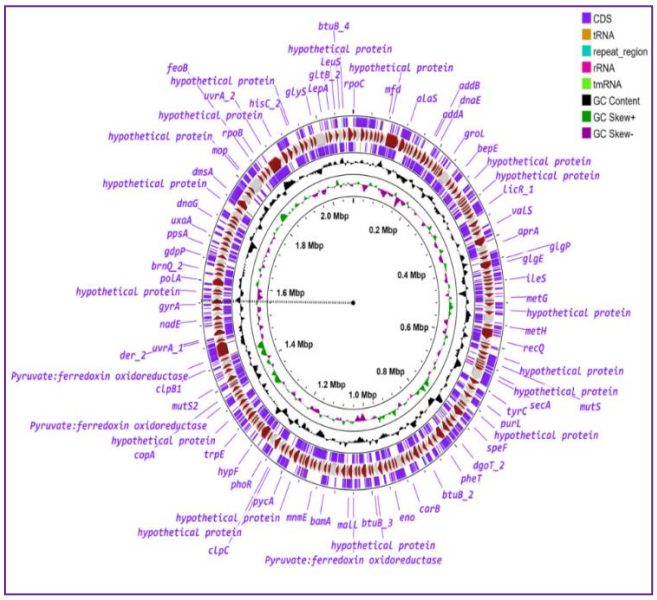


Figure 3.30: A. ***Megasphaera elsdenii*_HC2**

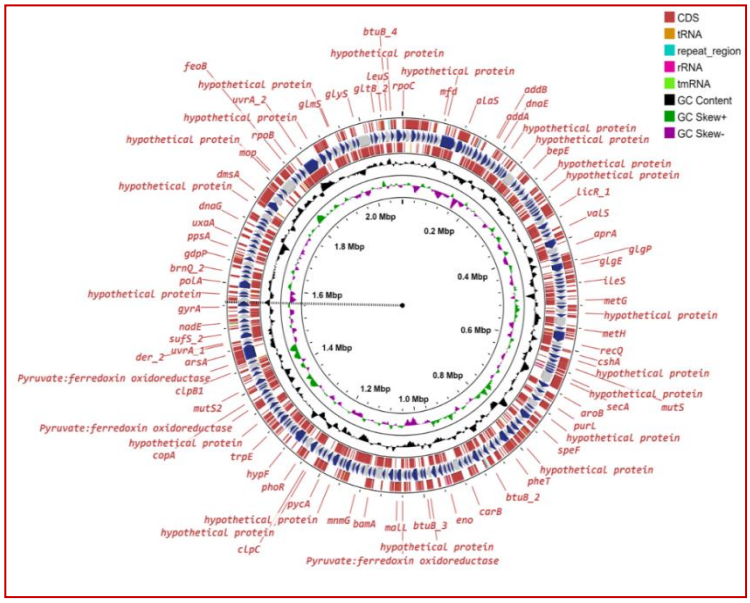


Figure 3.30: B. ***Megasphaera elsdenii*_IBS2**

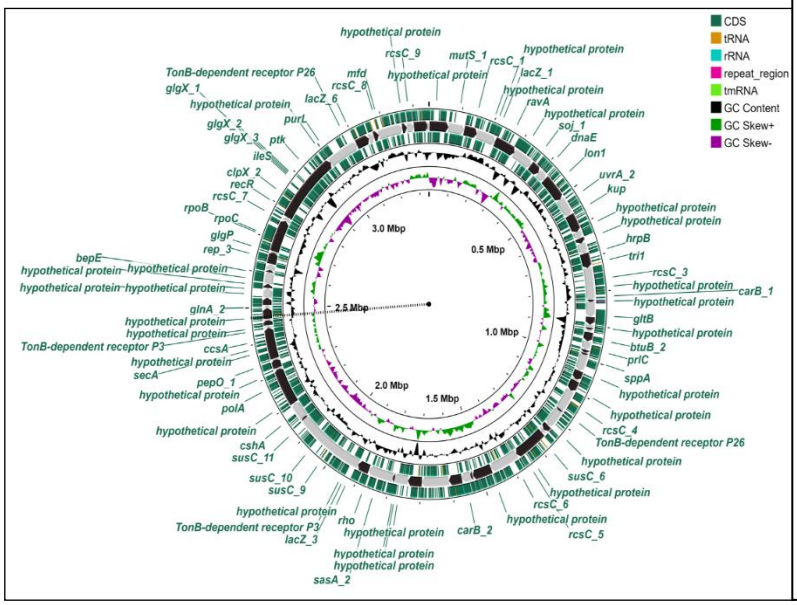


Figure 3.31: A. ***Segatella copri*_HC2**

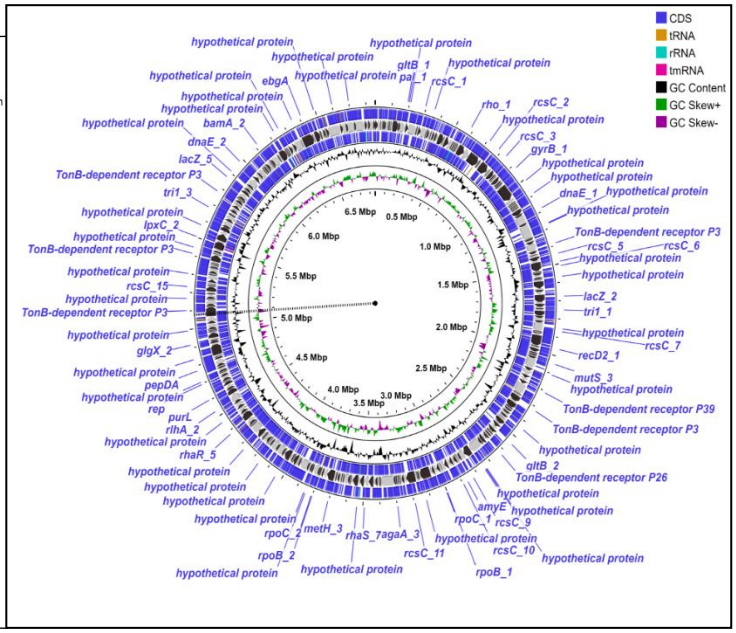
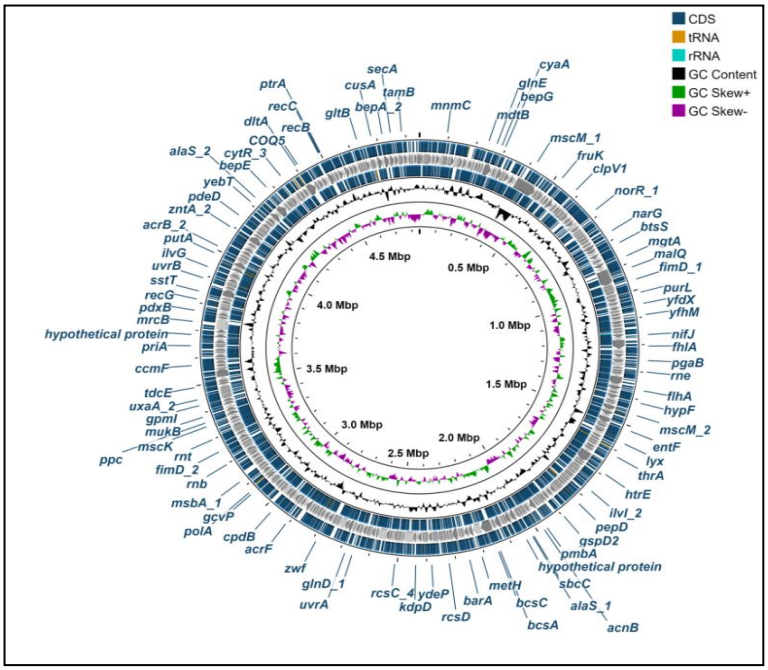


Figure 3.31: B. ***Segatella copri*_IBS1**



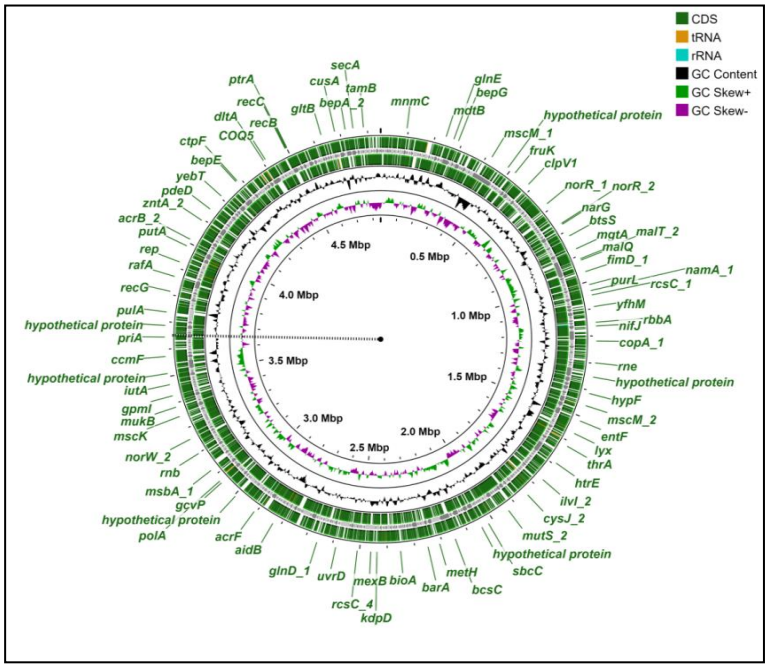


Figure 3.32: A. ***Klebsiella pneumoniae*_HC2** **Figure 3.32:** B. ***Klebsiella pneumoniae* _IBS3**

3.19 Biomarker investigation through LEfSe analysis

The LEfSe analysis (LDA score threshold: 2.5, $p < 0.05$, FDR-adjusted) revealed significant taxonomic differences between IBS and healthy groups (**Figure 3.33**). Healthy controls were enriched in Bacteroides species (*B. stercoris, B. uniformis*), consistent with their role in gut homeostasis. Also, *Bifidobacterium catenulatum* was detected to be a potential marker of healthy gut. In contrast, IBS patients showed higher abundances of *Klebsiella*, *Roseburia* and *Faecalibacterium* species. Notably, *Collinsella aerofaciens, Klebsiella pneumoniae, Streptococcus salivarius* and *Clostridium* species was uniquely associated with IBS.

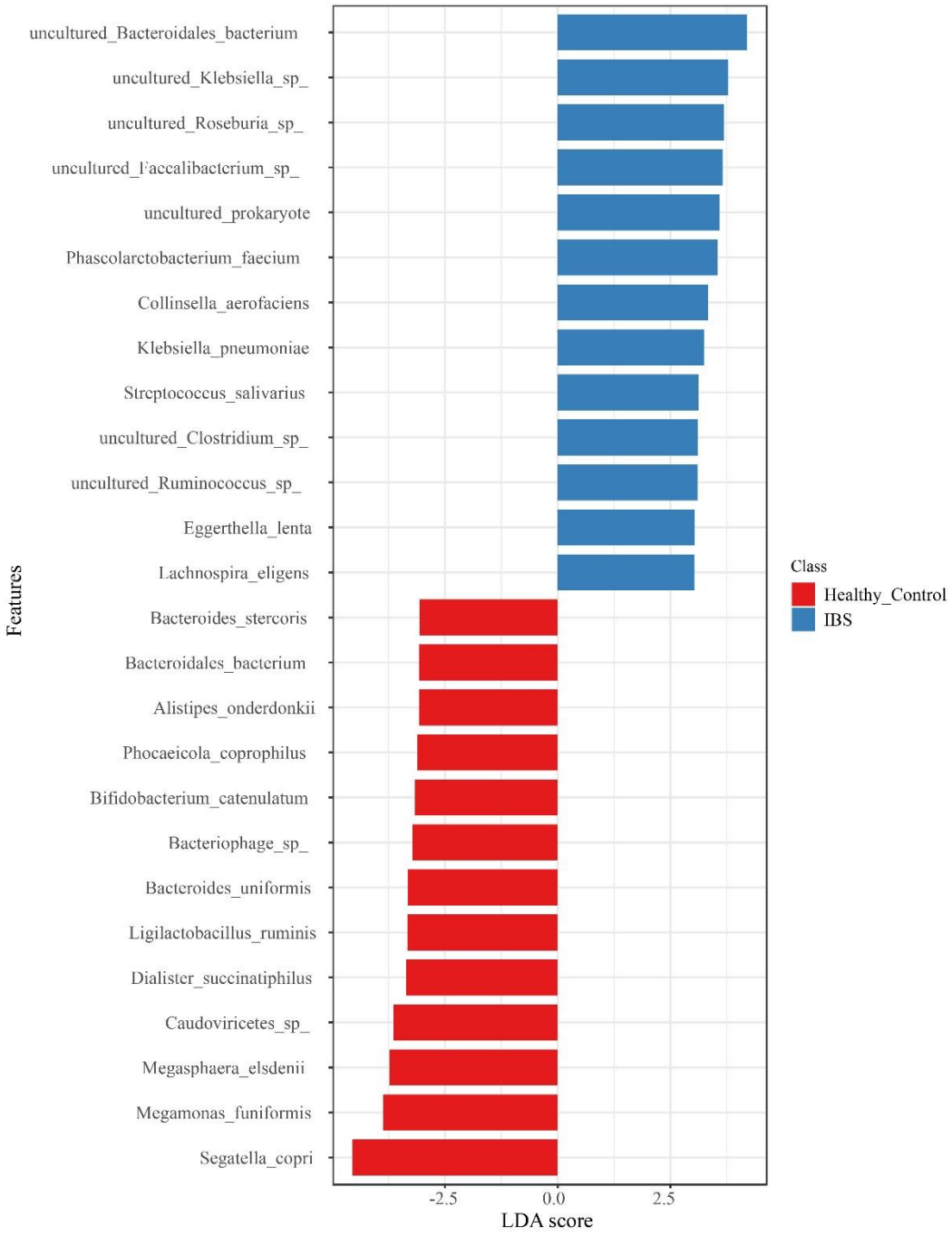


Figure 3.33: Microbial Biomarkers Distinguishing IBS from Healthy Gut Profiles using LEfSe analysis. This figure presents microbial features significantly associated with either IBS or healthy controls, identified through Linear Discriminant Analysis (LDA). Blue bars indicate taxa enriched in IBS, while red bars represent those enriched in healthy individuals. The length of each bar reflects the magnitude of the LDA score, highlighting the discriminatory power of each feature in distinguishing between the two groups.

3.20 Result of Machine learning for IBS and HC detection

In this section, the results will be presented and discussed from the experimented models of methodology. Firstly, we performed the 5-fold cross validation utilizing the whole dataset to see the generalized behavior of the model. In **figure 3.34**, it can be demonstrated that, except gradient boosting and XGBosst, CV scores exhibit substantial variability which may occur for the high dimensional data. But among all, logistic regression showed the largest mean CV scores.

After cross validation, we split our dataset into 2 parts: 885 samples for training and the rest 224 to evaluate the model. The performance of each model was assessed using seven evaluation metrics accuracy, precision, recall, F1 score, Matthew's correlation coefficient (MCC), and the areas under the ROC and PR curves which are presented in table 2. Logistic Regression outperformed the other models, achieving 0.80 in both accuracy and F1 score, and 0.88 in ROC-AUC. If we analyze the confusion matrix for Logistic Regression, we observe that the model correctly predicted 103 IBS samples, while 23 were misclassified as healthy. Similarly, out of 96 healthy samples, 74 were correctly identified, whereas the rest 22 were inaccurately classified. The ROC and PR curves (**Figures 3.35 and 3.36**, respectively) clearly show that Logistic Regression outperformed the other models, as represented by the violet line.

We can visualize how Logistic Regression can differentiate and make clusters between IBS and healthy by the t-SNE graph. It shows that the model can distinguish between two classes, IBS (red samples) and healthy (blue samples), into partially separable clusters. However, there is some overlap between the two classes, which represents the misclassification as we depicted in the confusion matrix.

The ROC (Receiver Operating Characteristic) curve is a graph that illustrates the performance of a binary classifier system as its discrimination threshold is varied. Although it is a relatively simple linear model, logistic regression achieves the best AUC. A line could separate the data. Complicated models like XGBoost, LightGBM, and Random Forest don't make a big difference, either because they overfit or are too complicated. Gradient Boosting, XGBoost, LightGBM, and Random Forest are all examples of ensemble models that are good at finding non-linear correlations. Their AUCs (0.85–0.86) are similar, indicating that they work well; however, it's unclear which one is superior in this case. To get the most out of them, you may need to do feature

engineering or hyperparameter optimization. The Support Vector Machine (SVM) performs well (AUC = 0.85), but not quite as well as Logistic Regression **(Figure 3.37)**.

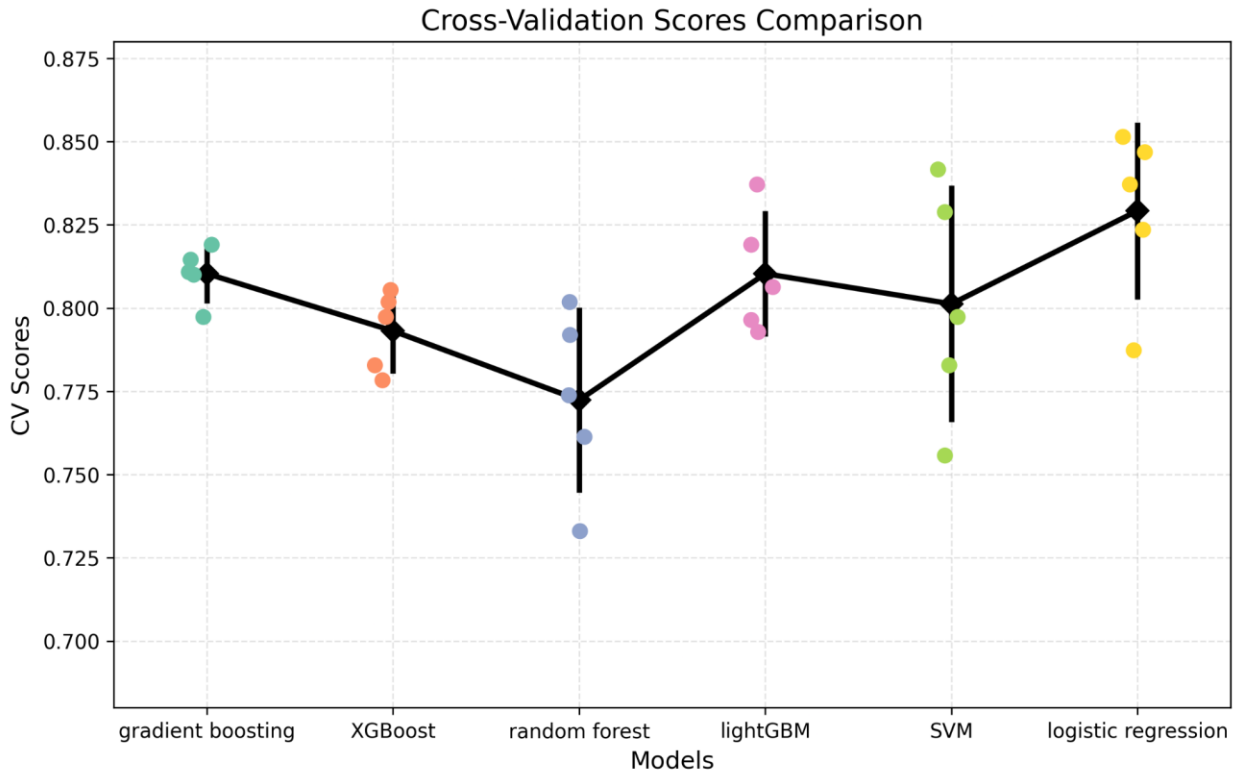


Figure 3.34: Cross validation of different ML models. Scatter plot comparing cross-validation scores across six machine learning models: gradient boosting, XGBoost, random forest, LightGBM, SVM, and logistic regression. Colored dots represent individual scores, while black lines with error bars indicate mean performance and variability. The visualization highlights model robustness and predictive accuracy for IBS-related classification tasks.

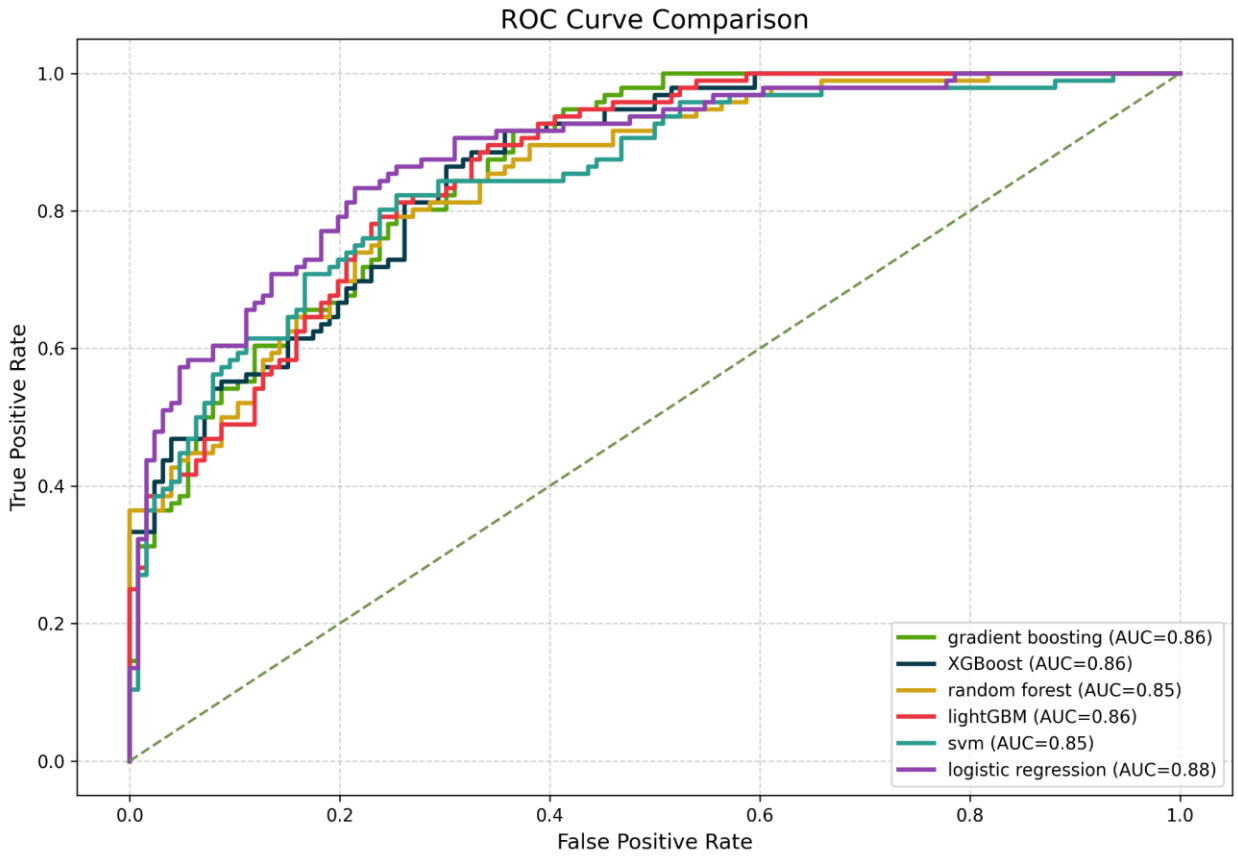


Figure 3.35: ROC curve comparison of the ML approach used in current study. ROC curve comparison of six machine learning models logistic regression, gradient boosting, XGBoost, random forest, LightGBM, and SVM used for IBS classification. Curves represent model sensitivity versus specificity, with AUC values ranging from 0.85 to 0.88. Logistic regression achieved the highest AUC (0.88), indicating superior discriminatory performance among tested models.

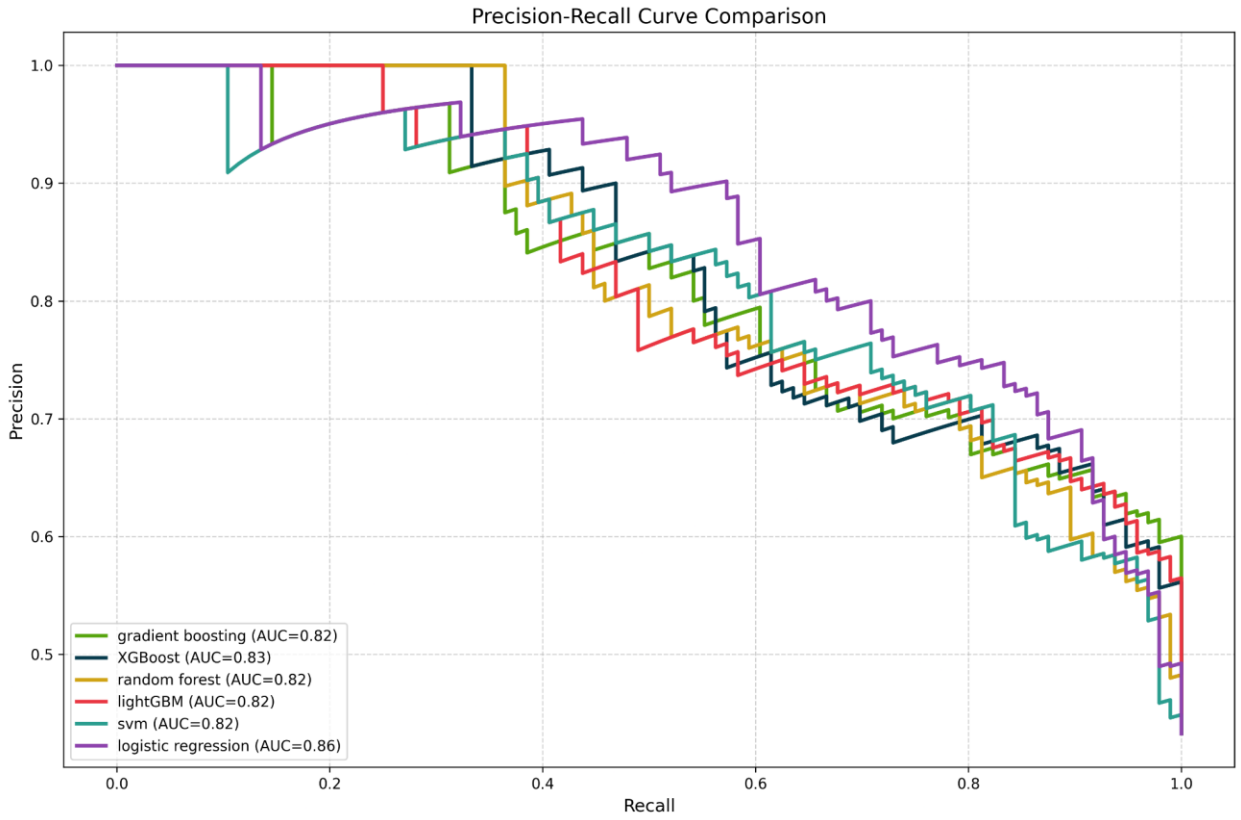


Figure 3.36: Precision Recall curve comparison. Precision–recall curve comparison of six machine learning models logistic regression, gradient boosting, XGBoost, random forest, LightGBM, and SVM used for IBS classification. Curves illustrate trade-offs between precision and recall, with logistic regression achieving the highest AUC (0.86), indicating superior performance in identifying IBS-related patterns with minimal false positives.

This dataset is better for logistic regression since it is easier to understand, simpler, and has a higher AUC. In biological or microbiological contexts (e.g., IBS versus healthy classification), a greater AUC indicates enhanced sensitivity and specificity, which is essential for reducing incorrect diagnoses. For dealing with imbalanced datasets, use ROC analysis along with precision-recall curves, confusion matrices, and calibration plots.

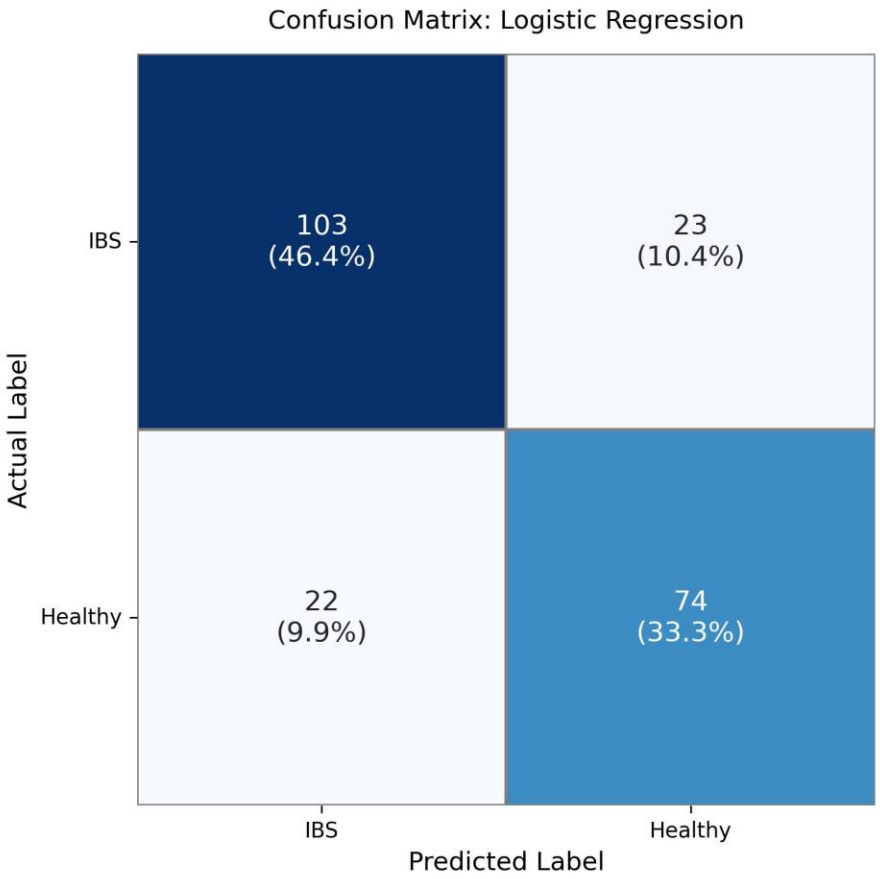


Figure 3.37: Confusion Matrix of Logistic Regression. Confusion matrix summarizing logistic regression model performance in classifying IBS versus healthy individuals. True positives (IBS = 103) and true negatives (Healthy = 74) indicate correct predictions, while false negatives (23) and false positives (22) reflect misclassification rates. Overall distribution highlights model accuracy and diagnostic reliability for microbiome-based IBS detection.

The provided confusion matrix quantitatively assesses the effectiveness of a logistic regression classifier in distinguishing individuals diagnosed with Irritable Bowel Syndrome (IBS) from healthy controls. The matrix is set up as a 2×2 grid that illustrates the relationship between the actual and anticipated class labels. True Positives correctly identify 103 instances (46.4%) of IBS. This cell demonstrates the model's ability to accurately identify IBS cases, which is the most significant aspect of the dataset. It means that it is moderately sensitive. IBS was incorrectly labeled as Healthy 23 times (10.4%), which is referred to as a False Negative. The errors indicate that the model failed to identify IBS in many individuals who had it, which could render it less effective for diagnostic screening. On the other hand, 22 healthy people (9.9%) were wrongly

diagnosed with IBS. The misclassification rate indicates that the algorithm tends to overpredict IBS, potentially leading to unnecessary anxiety or follow-up care in healthy individuals. True Negatives correctly predicted Healthy in 74 cases (33.3%). This cell illustrates the model's specificity, which refers to its ability to distinguish between healthy and sick individuals.

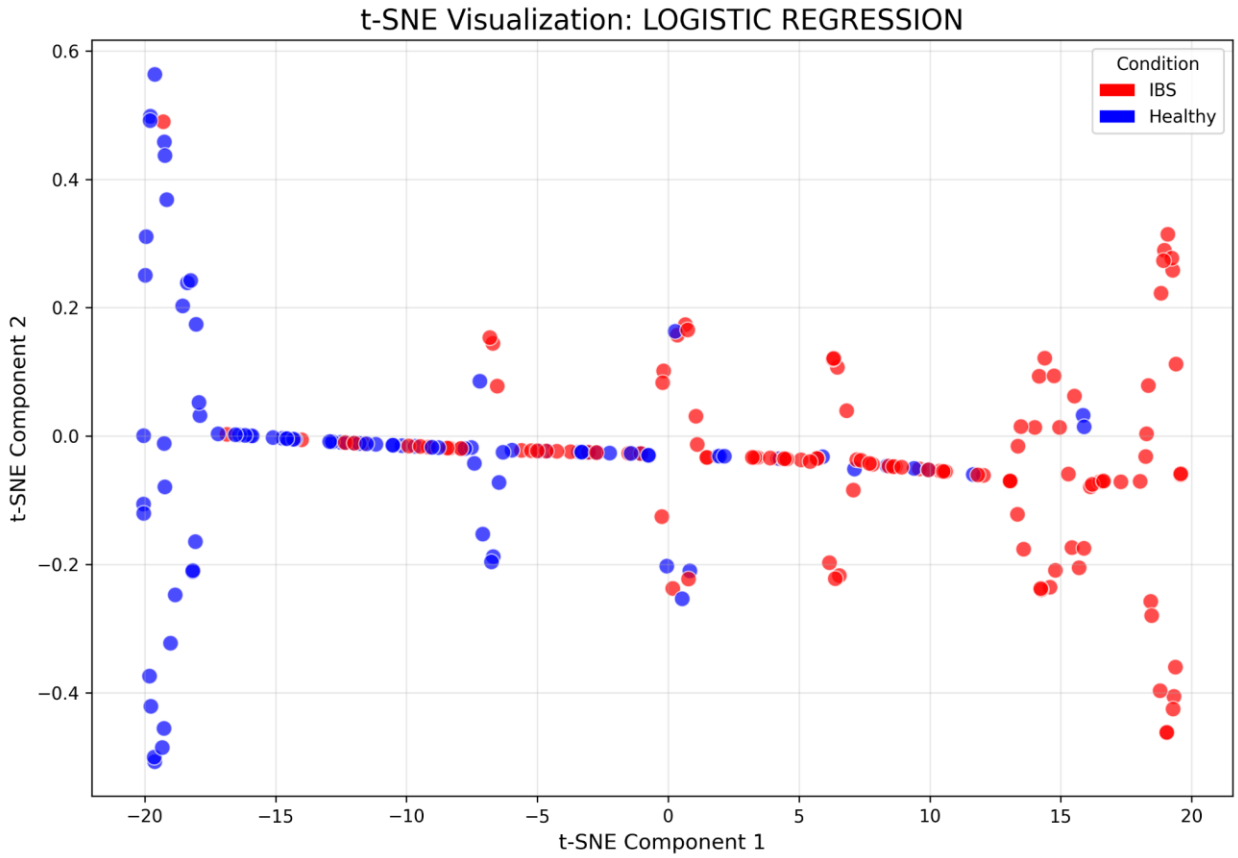


Figure 3.38: t-SNE visualization of logistic regression model. t-SNE plot illustrating dimensionality reduction of microbiome-based logistic regression classification. Data points represent IBS (red) and healthy (blue) individuals, mapped onto two components. Distinct clustering patterns suggest effective separation between groups, highlighting latent structure in microbial features relevant to IBS diagnosis.

To make high-dimensional data easier to see, this scatter plot projects it into two dimensions using t-distributed Stochastic Neighbor Embedding (t-SNE). Red spots represent individuals with IBS, whereas blue points represent healthy controls. The blue (healthy) and red (IBS) dots are mostly grouped on the opposing sides of the plot, with the blue on the left and the red on the right **(Figure 3.38)**. This implies that the logistic regression model has learned features that successfully

differentiate between the two circumstances. High categorization confidence is indicated by minimal mixing between red and blue dots. When it comes to differentiating between those with IBS and healthy people, the model probably has good sensitivity and specificity. Samples with comparable feature profiles such as microbial abundance or the presence of AMR genes are grouped, and notable differences are observed between IBS and healthy individuals in these profiles.

Table 3.4: Comparative performance of different machine learning models based on test data.

model	accuracy	precision	recall	f1_score	MCC	ROC_ AUC	PR_AUC
gradient boosting	0.7523	0.7543	0.7523	0.7529	0.4989	0.86	0.82
XGBoost	0.7387	0.7387	0.7387	0.7387	0.4678	0.86	0.83
random forest	0.7568	0.7567	0.7568	0.7536	0.5	0.85	0.82
lightGBM	0.7658	0.7683	0.7658	0.7665	0.5271	0.86	0.82
svm	0.7613	0.7644	0.7613	0.762	0.5189	0.85	0.82
logistic regression	**0.7973**	**0.7976**	**0.7973**	**0.7974**	**0.5876**	**0.88**	**0.86**

This radar chart compares six machine learning models (Gradient Boosting, XGBoost, Random Forest, LightGBM, SVM, and Logistic Regression) to six performance indicators. To assess how well each model adapted, we examined test accuracy, Precision, PR AUC, Recall, F1 Score, and ROC AUC (**Table 3.4**). A colored polygon represents each model, making it easy to compare them across several dimensions visually. Logistic Regression is better than all other models on every metric, which is impressive given its simplicity compared to ensemble methods. Ensemble models (Gradient Boosting, XGBoost, Random Forest, and LightGBM) exhibit similar performance metrics, indicating that adding more complexity to these models in this dataset does not yield significant improvements. SVM shows average performance, but it does not fare well in any one statistic. This could be because of kernel restrictions or how it handles feature scaling. The fact that non-logistic models are so close together suggests that hyperparameter tuning or feature engineering may be necessary to enhance their performance further.

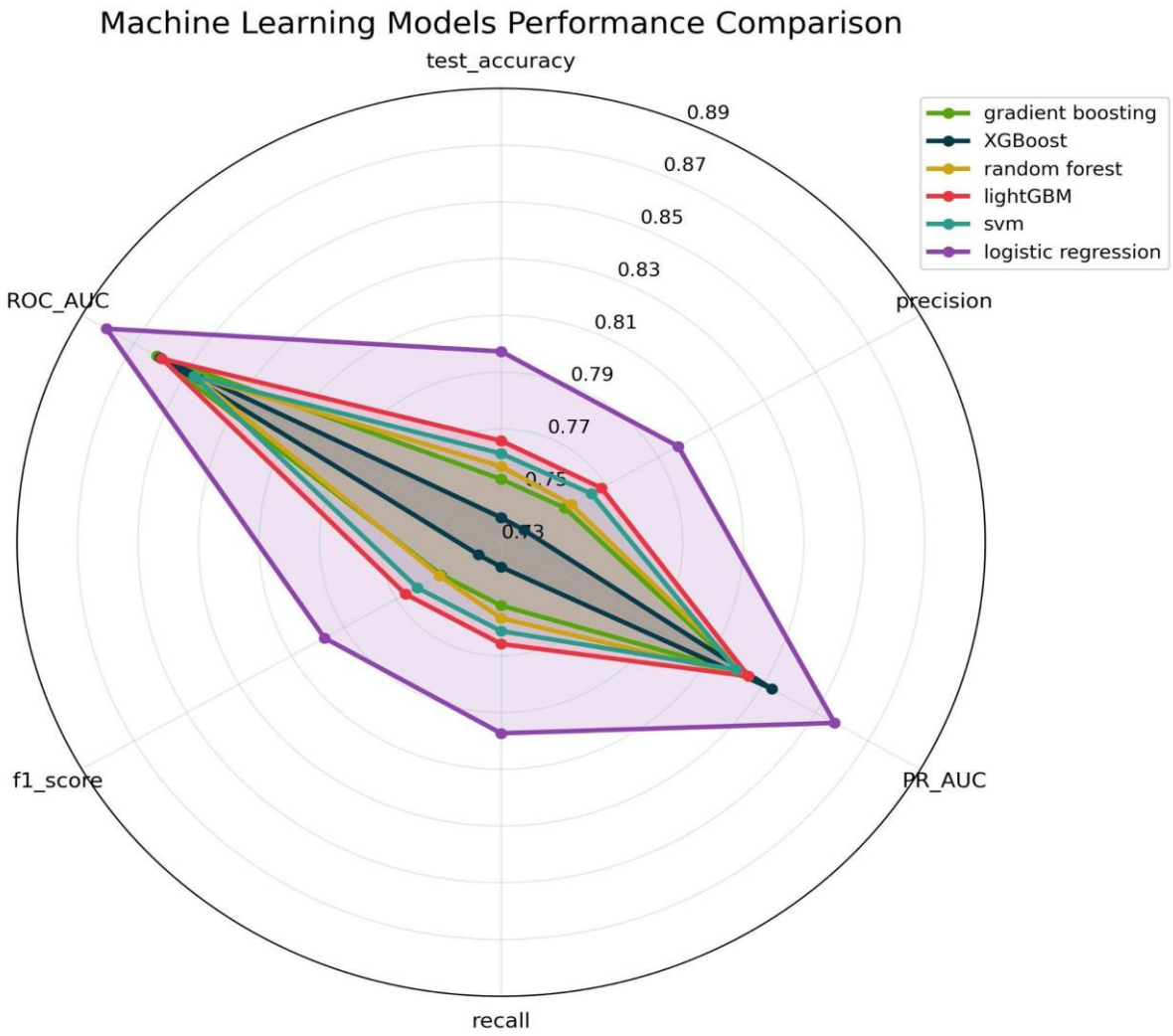


Figure 3.39 Radar chart to compare six machine learning models. Radar chart comparing six machine learning models logistic regression, gradient boosting, XGBoost, random forest, LightGBM, and SVM across six evaluation metrics: test accuracy, precision, recall, F1 score, ROC_AUC, and PR_AUC. Logistic regression demonstrates consistently high performance across all metrics, indicating strong predictive capability for IBS classification.

This dataset is better for logistic regression since it is easier to understand, simpler, and has a higher AUC. In biological or microbiological contexts (e.g., IBS versus healthy classification), a higher AUC signifies improved sensitivity and specificity, which is crucial for minimizing erroneous diagnoses. To address imbalanced datasets, utilize ROC analysis in conjunction with precision-recall curves, confusion matrices, and calibration plots.

Chapter 04

Discussion

4. Discussion

Due to the enormous complexity of the gut microbiota, our understanding of the human gut microbiome is still in its infancy. Metagenomics emerged as a modern approach for studying "microbial dark matter". Using traditional culture methods, it was previously impossible to analyze those dark matters. The gut microbiota of IBS patients from Bangladesh was examined in this study utilizing an advanced metagenomic method with two distinct pipelines. This is the first study from Bangladesh, and it serves as a foundation for future in-depth investigations.

The demographic and clinical attributes of the study participants indicate notable differences between the healthy control (HC) group and the irritable bowel syndrome (IBS) group. The HC group (n = 6) was mostly made up of men (4 out of 6), and their ages ranged from 27 to 40 years. Most participants were married, and none reported diabetes, hypertension, or family history of gastrointestinal disorders. The stool type was consistently normal. Education levels were generally higher in HC participants, with most reporting 17–18 years of education. In contrast, the IBS group (n = 10) demonstrated greater heterogeneity in age, sex, stool type, and comorbidities. Ages ranged from 20 to 70 years, with a predominance of younger adults (20–27 years). Males were more frequent, though females were also represented. Lactose intolerance was highly prevalent among IBS participants (9/10), underscoring its potential role as a contributing factor. Stool types varied widely, including diarrhea, constipation, mixed, and alternating patterns, reflecting the clinical diversity of IBS. Unlike the HC group, several IBS participants reported family history of gastrointestinal conditions and antibiotic consumption, which may indicate environmental or genetic predispositions. After metagenomic DNA extraction, the concentration and purity of all the DNA samples were of high quality.

4.1 Metagenomic sequencing quality and depth

The preprocessing processes, including quality trimming and host DNA removal, were completed successfully. Especially for sample IBS4, which included a high proportion of human DNA. The use of STAR followed by Bowtie2 is a reliable pipeline for host filtering that is widely used in metagenomic workflows (Mandal et al., 2015). The retention of 5-6% clean reads after filtration is uniform with expectations for human gut metagenomic samples, where host DNA can dominate

raw reads. Processing 120 million reads from ten samples shows a lot of sequencing depth, which makes it easier to figure out what species are present and how they work (Xu et al., 2022).

Gut microbiota typically have GC content ranging from 50-75%, which comprises a wide range of bacterial taxa with varied genomic compositions. Significant redundancy, which might be brought on by PCR bias or the overrepresentation of prominent species, is indicated by sequence duplication rates between 5 and 25%. The high Phred scores (>30) across all sixteen samples revealed excellent base calling accuracy. It supports the virtue of downstream analysis (Lo & Chain, 2014).

4.3 Bacterial Composition in the Gut Revealed Striking Dysbiosis

The present study provides an extensive analysis of the gut microbiome in patients with IBS, highlighting significant dysbiosis when compared to healthy controls. The observed dysbiosis corresponds with prior studies demonstrating modified gut microbiota compositions in IBS patients, which may contribute to the condition's pathogenesis (Gun-Ha et al., 2012).

One of the key findings of this study is the substantial decrease in the abundance of Actinobacteria in IBS patients. This is consistent with previous reports which have shown reduced levels of Actinobacteria, particularly *Bifidobacterium* species, in IBS patients (Jeffery et al., 2012). *Bifidobacterium*, a genus within Actinobacteria, is known for its beneficial roles in gut health, including maintaining gut barrier function and modulating immune responses (Callaghan & Sinderen, 2016). The significant reduction or absence of *Bifidobacterium* species such as *B. breve*, *B. longum*, and *B. bifidum* in IBS patients supports the hypothesis that these bacteria play a protective role in gut health, and their depletion may contribute to IBS pathogenesis. Interestingly, certain *Bifidobacterium* species such as *B. cebidarum*, *B. felsineum*, and *B. reuteri* were found to be elevated in IBS patients. This divergence in *Bifidobacterium* species abundance suggests a complex microbial shift in IBS, potentially driven by host factors or specific dietary influences.

Our findings show a relatively elevated abundance of Firmicutes in the gut microbiota of IBS patients, particularly *Faecalibacterium prausnitzii*, a species typically considered anti-inflammatory (Miquel et al., 2013). However, despite its known benefits, the increased abundance of *F. prausnitzii* in IBS patients could indicate a compensatory response to inflammation or a disrupted gut environment. Healthy controls, on the other hand, exhibited higher abundances of

other Firmicutes such as *Megasphaera elsdenii* and *Lactobacillus luminis*. *Lactobacillus* species are often associated with gut health, and their reduced presence in IBS patients might contribute to symptom severity (Ringel-Kulka et al., 2011). Bacteroidetes remained the most prevalent phylum in both groups, with *Segatella copri* (Formerly known as *Prevotella copri*) emerging as a dominant species. While *S. copri*'s role in gut health is debated, its high abundance in both groups suggests it might not be directly implicated in IBS but rather reflects a common gut microbiota feature (Tett et al., 2019). Conversely, the reduced abundance of *Bacteroides fragilis* in IBS patients aligns with its proposed protective role in gut health (Mazmanian et al., 2005).

4.4 Lower diversity in the gut microbiota of IBS patients

The alpha diversity analysis revealed that the Control group generally exhibited higher species richness and diversity across several metrics. Specifically, the Observed, Chao1, and ACE indices suggest that the Control group has a higher richness, particularly regarding abundant species as indicated by the ACE measure. The Shannon and Simpson indices further underscore the greater diversity and evenness within the Control group. Variability analysis indicates that the Control group displays wider interquartile ranges (IQRs) and more outliers across most measures, reflecting greater heterogeneity and a broader range of ecological conditions or sample characteristics. The presence of outliers in both groups highlights the significant differences in species richness and diversity among individual samples. Reduced evenness in IBS patients may reflect dominance by a few opportunistic taxa, consistent with dysbiosis (Pittayanon et al., 2019). In conclusion, the Control group exhibits higher overall species richness and diversity, along with greater variability, compared to the IBS group. This suggests that the microbial communities within the Control group are more diverse and exhibit a wider range of characteristics, possibly due to varying ecological conditions or differences in sample sources.

The beta diversity analysis depicted in the PCoA plot indicated significant differences in the gut microbiota composition between IBS patients and healthy controls. The clear separation of clusters indicated that IBS profoundly impacted the gut microbial community structure, leading to a distinct microbial profile in affected individuals. The clustering of control samples together suggested that healthy individuals tend to have a more consistent and uniform gut microbiome, whereas the IBS samples display greater variability and a distinct composition. This variability among IBS patients may reflect the heterogeneity of the disease and its multifactorial nature, which

includes genetic, environmental, and lifestyle influences. The inclusion of geographic location as a variable in the analysis shows some influence on microbial community composition, with samples from different locations exhibiting minor clustering patterns. However, the stronger separation observed between IBS and control groups emphasizes that the disease state exerts a more significant impact on gut microbiota than geographic factors. This suggests that while local environmental factors and diet may contribute to microbiota composition, the presence of IBS overrides these influences, leading to a distinct microbial profile.

The NMDS study showed that the microbial communities in the control and IBS groups were very distinct from each other and from each other in different parts of the world. The distinct clustering observed in the NMDS plot indicated that both health status and location influence the composition of microbial communities. The greater dispersion of IBS samples suggests that IBS is associated with a more variable and less stable microbial community compared to healthy controls. This variability in the gut microbiome could be contributing to the pathophysiology of IBS, potentially through mechanisms such as altered metabolic activity, immune modulation, or disruption of gut barrier function. The similarity between samples from Dhaka and Gazipur implies that living close to each other may lead to similar microbial populations. This could be because of common environmental factors like diet, lifestyle, and sanitation practices. The clear disparities seen in other places, on the other hand, show that regional variances have a big impact on the gut microbiome. These predictive findings would gain greater significance and validation if a larger number of samples from different geographical locations could be sequenced and analyzed using the same pipeline.

These findings have important implications for understanding the pathophysiology of IBS and developing targeted interventions. The unique microbial fingerprints linked to IBS may function as prospective biomarkers for the diagnosis and surveillance of the condition. Moreover, therapeutic strategies aimed at modulating the gut microbiome, such as probiotics, prebiotics, or dietary modifications, could be tailored to address the specific microbial alterations observed in IBS patients. The influence of geographic location also suggests that personalized approaches considering local environmental and dietary factors may enhance the effectiveness of such interventions.

The findings from functional genomics provide valuable insights into the complex microbiome landscape of IBS patients. The reduced abundance of specific functional genes in IBS patients may indicate disruptions in essential metabolic and regulatory pathways, which could contribute to disease pathogenesis. The presence of virulence factor genes underscores the potential for pathogenic interactions within the gut environment, possibly exacerbating the condition. The identification of antimicrobial resistance genes across various classes highlights the resilience of the gut microbiota to antibiotic treatment, which poses challenges for managing infections in IBS patients. Understanding these microbial dynamics and resistance patterns can inform future therapeutic strategies and guide personalized medicine approaches for better management of IBS.

4.5 Core Microbiome Analysis of Healthy Controls and IBS Samples

Segatella copri, *Bacteroides fragilis*, *Bacteroides ovatus*, and *Bifidobacterium* species were found in nearly every healthy person's core microbiome. These taxa are frequently linked to gut homeostasis, SCFA generation, and immunological regulation (Lopez-Siles et al., 2017; Callaghan & Sinderen, 2016). *Bacteroides* and *Bifidobacterium* numbers are lower in IBS patients, which is consistent with earlier research. This could be attributed to microbial imbalance and low anti-inflammatory capacity (Gun-Ha et al., 2023; Pittayanon et al., 2019).

The greater occurrence of "Not Assigned" taxa in IBS samples may indicate a more diversified or poorly defined microbial community, which could be influenced by environmental factors, nutrition, or host genetics (Sánchez-Pellicer et al., 2025). The dysbiosis hypothesis is supported further by the presence of *K. pneumoniae* and *Sutterella wadsworthensis*, both of which are related with inflammation and epithelial disruption (Shin et al., 2015; Skvortsova et al., 2024).

4.6 Higher Prevalence of Pathogenic Microbes in IBS

The current study demonstrated a much greater prevalence of harmful bacteria in IBS patients, such as *Klebsiella pneumoniae*, *Enterobacter cloaceae*, *Citrobacter freundii*, and *Streptococcus sanguinis*. These bacteria are known to cause intestinal inflammation, increased permeability, and immunological activation (Shin et al., 2015; Tu et al., 2025). The presence of *Acinetobacter baumannii*, *Photobacterium damselae*, and *Providencia rettgeri* typically opportunistic pathogens indicates a disturbed gut environment that may favor colonization by hazardous species.

The prevalence of *Vibrio* species (*V. vulnificus*, *V. parahaemolyticus*, and *V. alginolyticus*) in IBS patients is particularly problematic because they are linked to gastrointestinal infections and toxin production (Sun et al., 2022). Their presence in the gut microbiota could indicate temporary colonization or altered host defenses.

4.7 Relative Abundance of Bacteriophage

Bacteriophages play an important role in structuring microbial ecosystems through predation and horizontal gene transfer. The low abundance of *Erwinia*, *Salmonella*, and *Escherichia* phages in IBS patients shows a breakdown in phage-mediated microbial control (Shkoporov & Hill, 2019). In contrast, the increased prevalence of *Faecalibacterium* viruses (Toutatis, Lugh, Taranis, and Oengus) may indicate compensatory viral activity against dominant bacterial taxa in IBS (Gun-Ha et al., 2023). This change in phage composition could affect microbial dynamics and contribute to the instability seen in IBS microbiomes. Bacterial phages play an unknown role in IBS, but they may provide new treatment targets.

The reduced presence of bacteriophage sequences in IBS patients compared to healthy controls is noteworthy. Bacteriophages can influence bacterial populations and potentially modulate gut health (Manrique et al., 2017). The higher abundance of *Faecalibacterium* virus in IBS patients suggests a complex interplay between viruses and bacteria in the gut, possibly contributing to microbial instability.

4.8 Top-Most Abundant AMR Genes in HC vs. IBS

The compared heatmaps show a significant difference in resistome stability. Core AMR genes such as *acfB*, *acfC* were consistently found in healthy controls across several thresholds, demonstrating a robust and predictable resistome. This stability is most likely the result of a well-balanced microbial population with low turnover and negligible selective pressure (Zhernakova et al., 2016).

In contrast, IBS samples exhibited a broader and more erratic AMR gene profile. Genes like *toxR*, *ompU*, and *hlyA* often associated with virulence and environmental stress responses were detected at lower thresholds but diminished with increased stringency. This pattern suggests that these genes may be carried by transient or opportunistic bacteria, which are more prevalent in dysbiotic environments (Lloyd-Price et al., 2017). IBS is associated with increased microbial instability,

which can lead to the emergence of less dominant, potentially resistant strains (Simrén et al., 2013). Inflammation: Inflammatory conditions can alter microbial niches and promote the expansion of resistant organisms (Schirmer et al., 2016). Antibiotic history: Even distant antibiotic exposure can have enduring effects on the gut resistome, especially in those whose microbiota is not very strong (Palleja et al., 2018). These findings align with recent metagenomic studies showing that individuals with gastrointestinal disorders often harbor more diverse and dynamic resistomes, even when overall resistance gene abundance is comparable to healthy individuals (Y. Wang et al., 2019; Z. Zhang & Howe, 2022).

The presence of antimicrobial resistance (AMR) genes in both IBS patients and healthy controls highlights the significance of the gut microbiome as a reservoir for resistance determinants. The prevalence of tetracycline resistance genes, notably *tetQ*, *tetW*, and *tetM*, is consistent with prior research that identified these genes as among the most common in the human gut microbiota (Forslund et al., 2014; Yongfei et al., 2016). These genes are frequently associated with commensal anaerobes such as Bacteroides and Clostridium, which are common in the gut and can act as reservoirs for horizontal gene transfer (Sommer et al., 2009).

The presence of macrolide resistance genes (*ermF*, *ermX*, and *mphA*) as well as the beta-lactam resistance gene *cfxA6* suggests that even healthy people had been exposed to antibiotics in the past. These genes are typically found in mobile genetic elements, boosting their potential for spreading throughout microbial populations (Martínez et al., 2015).

Surprisingly, while the overall classes of AMR genes did not differ significantly between the IBS and control groups, their composition and distribution varied. This shows that, whereas both groups contain comparable resistance gene classes, their ecological contexts and microbial carriers may differ.

4.9 Network analysis of the Healthy Controls and IBS samples:

In our metabolic pathway network analysis, we observed distinct differences between IBS patients and healthy controls. Interestingly, the network of significant metabolic pathways in healthy controls appears to be less extensive compared to those in IBS patients. This suggests that the microbial communities in the gut of IBS patients are more abundant and exhibit a more complexly

correlated network. These findings suggest a potential role for this complex microbial network in the development and progression of IBS.

These findings have significant implications for our understanding of IBS pathology. The increased complexity and interconnection of the microbial metabolic network may not be purely a result of the disease. However, they may actively contribute to the development and progression of IBS symptoms. This complex microbial community, with its highly connected metabolic activity, may induce or perpetuate the gastrointestinal dysfunction associated with IBS. The findings suggest that therapeutic techniques aimed at this complex microbial network could provide viable possibilities for intervention in IBS management.

4.10 Worldwide Gut Microbial Diversity Assessment Differentiates IBS from Healthy Controls and Suggests Biomarker Signatures

The comparative analysis of global gut microbiome data between IBS patients and healthy controls (HC) highlighted several important ecological and clinical insights. By using rarefaction to make sequencing depth more consistent, we made sure that differences in microbial diversity were real biological differences and not just technical ones. This methodological rigor strengthened the reliability of our findings across diverse sequencing platforms and populations. Alpha diversity analyses consistently demonstrated that healthy individuals possess greater microbial richness and evenness compared to IBS patients. This observation aligned with the established notion that high microbial diversity is a hallmark of a resilient and stable gut ecosystem. The reduced diversity in IBS samples suggested a compromised microbial community structure, which may predispose individuals to dysbiosis and impaired gut functionality. LEfSe analysis provided further granularity by identifying potential microbial biomarkers that distinguish IBS from healthy states. Enrichment of *Bacteroides stercoris*, *Bacteroides uniformis*, and *Bifidobacterium catenulatum* in healthy controls supports their protective role in gut homeostasis, possibly through carbohydrate metabolism and immune modulation. Conversely, the increased abundance of *Klebsiella pneumoniae*, *Collinsella aerofaciens*, *Streptococcus salivarius*, and *Clostridium* species in IBS patients highlights taxa that may serve as diagnostic indicators or therapeutic targets. The presence of opportunistic pathogens such as *Klebsiella* further suggested that IBS-associated microbiota may predispose individuals to secondary infections or exacerbate gut inflammation.

4.11 Machine Learning based model development of IBS detection

The classification of IBS using microbiome-derived features presents a compelling challenge in computational biology, primarily due to the high dimensionality, compositional nature, and biological variability of microbiome data. In this study, multiple machine learning models were evaluated for their ability to discriminate between IBS patients and healthy individuals. Among them, Logistic Regression consistently demonstrated superior performance, both in cross-validation and independent testing, underscoring its utility in microbiome-based diagnostics.

4.11.1 Model Stability and Generalization

The application of 5-fold cross-validation revealed that Logistic Regression achieved the highest mean CV score with minimal variability, indicating strong generalization and stability. This finding is consistent with prior studies that emphasize the robustness of linear models in high-dimensional biological datasets (Karwowska et al., 2025; Zhou et al., 2022). In contrast, ensemble methods such as Gradient Boosting and XGBoost exhibited greater sensitivity to data partitioning, a phenomenon attributed to overfitting in sparse datasets (D. et al., 2020).

Microbiome datasets often suffer from the "curse of dimensionality," where the number of features (e.g., microbial taxa) far exceeds the number of samples. This imbalance can lead to model instability and poor generalization, particularly in complex models with large parameter spaces (Abegaz et al., 2025). Logistic Regression, with its linear decision boundary and reduced complexity, mitigates these risks, making it well-suited for microbiome classification tasks.

4.11.2 Performance Metrics and Interpretability

Independent testing further validated the efficacy of Logistic Regression, which achieved an accuracy of 0.80, F1-score of 0.80, and ROC-AUC of 0.88. These metrics reflect the model's ability to distinguish IBS from healthy samples across varying thresholds. The ROC-AUC value, in particular, indicates that the model correctly ranks IBS samples higher than healthy ones in 88% of randomly selected pairs, demonstrating high discriminative power (Palani et al., 2024).

Interpretability is a critical consideration in microbiome research, where understanding the biological relevance of features is as important as predictive accuracy. Logistic Regression offers a transparent mapping between features and outcomes, enabling researchers to identify key microbial taxa associated with IBS (Tarazona et al., 2023). This contrasts with ensemble and deep

learning models, which often function as “black boxes” and require complex post hoc interpretation techniques (Medina et al., 2025).

4.11.3 Biological Overlap and Visualization

The confusion matrix analysis revealed that while Logistic Regression performed well overall, a subset of samples near the classification boundary remained difficult to distinguish. This biological overlap is further visualized in the t-SNE plot, where IBS and healthy samples formed partially separable clusters. Such overlap is consistent with the hypothesis that IBS represents a spectrum of dysbiosis rather than a discrete microbial state (Fukui et al., 2020; Xu et al., 2020).

t-SNE, a nonlinear dimensionality reduction technique, preserves local structures in high-dimensional data and is particularly effective for visualizing microbiome profiles (Xu et al., 2020). The observed overlap in the t-SNE map suggests that certain individuals exhibit transitional microbial compositions, complicating binary classification. This finding aligns with studies indicating that environmental factors, diet, and host genetics contribute to microbiome heterogeneity (Hernández Medina et al., 2022; Zhou et al., 2022).

4.11.4 Ensemble Models and Overfitting

Despite their theoretical advantages, ensemble models such as Random Forest, Gradient Boosting, and XGBoost underperformed in this study. These models are designed to capture complex nonlinear relationships, but their effectiveness diminishes in datasets with limited sample sizes and high feature sparsity (Abegaz et al., 2025; D. et al., 2020). The tendency of ensemble methods to overfit specific data partitions was evident in their high variability across cross-validation folds.

Moreover, the interpretability of ensemble models remains a challenge. While techniques such as SHAP (SHapley Additive exPlanations) and feature importance rankings can provide insights, they often lack the direct transparency offered by linear models (Tarazona et al., 2023). In clinical applications, where understanding the biological basis of predictions is crucial, this lack of interpretability can hinder adoption and trust.

4.11.5 Data Transformation and Feature Selection

Effective preprocessing is essential for microbiome classification. Transformations such as centered log-ratio (CLR) and variance stabilization have been shown to improve model performance by addressing compositionality and zero inflation (Karwowska et al., 2025; Zhou et

al., 2022). In this study, feature selection and transformation played a pivotal role in enhancing the performance of Logistic Regression, allowing it to capture the most discriminative microbial signatures.

Feature selection also contributes to model interpretability by reducing dimensionality and highlighting biologically relevant taxa. Studies have demonstrated that targeted selection of microbial features can improve classification accuracy and facilitate biomarker discovery (D. et al., 2020; Tarazona et al., 2023).

4.11.6 Clinical Implications and Future Directions

The success of Logistic Regression in this study has important implications for microbiome-based diagnostics. Its simplicity, robustness, and interpretability make it an attractive choice for clinical applications, where transparency and reproducibility are paramount. Moreover, the ability to identify key microbial features associated with IBS can inform therapeutic and diagnostic strategies and personalized interventions (Fukui et al., 2020; Tarazona et al., 2023).

The primary findings of this study warrant further comprehensive research covering larger sample sizes, integrating multi-omics data, and exploring hybrid models that combine the interpretability of linear methods with the flexibility of nonlinear approaches. Additionally, longitudinal studies can provide insights into temporal dynamics of the microbiome and improve classification accuracy (Xia & Sun, 2023b). In summary, this study demonstrates that Logistic Regression is a powerful and interpretable tool for microbiome-based classification of IBS. Its superior performance across multiple evaluation metrics, combined with its robustness and transparency, highlights its potential for clinical application. While ensemble models offer theoretical advantages, their complexity and susceptibility to overfitting limit their utility in sparse microbiome datasets. The findings underscore the importance of model selection, data transformation, and feature optimization in achieving accurate and biologically meaningful classifications. Importantly with the progress of AI, ML modeling represents a promising acceptable, sustainable approach for microbiome associated disorders IBS and beyond. Validation with larger genomic data with regional customization, it can be a sustainable solution for disease detection, prediction and progress monitoring.

Chapter 05

Conclusion

5. **Conclusion**

This study demonstrated that a multi-layered metagenomic framework can provide a powerful and comprehensive understanding of the gut ecosystem in IBS. By integrating shotgun metagenomics, strain-resolved genome reconstruction, functional and resistome characterization, global dataset comparison, and data-driven machine-learning models, the work went beyond the conventional descriptive microbiome studies and offered a comprehensive systems-level view of IBS-associated microbial dysbiosis. Importantly, this approach highlighted that the dysbiosis observed in IBS is not limited to shifts in community composition but extends to deeper functional and genomic alterations that reflect a disrupted and unstable gut ecosystem. Reproducibility is crucial for developing universally applicable biomarkers, therapeutic targets, and diagnostic tools.

The successful application of machine-learning models further illustrated the translational potential of microbiome-based diagnostics. The ability of even relatively simple models to distinguish IBS from healthy individuals suggests that microbial profiles carry strong predictive signals. With continued refinement, such approaches may ultimately support clinical decision-making, risk stratification, or personalized intervention strategies.

Despite its strengths, the study also highlighted several considerations. The sample size of locally sequenced metagenomes was modest, but not significantly higher. Clinical metadata including dietary habits, symptom severity, drug history, and comorbidities could further improve interpretability and allow deeper exploration of host-microbe interactions. Moreover, the cross-sectional design limited the ability to determine causality, emphasizing the need for longitudinal studies to disentangle whether microbiome shifts are drivers or consequences of IBS.

Overall, this work contributed important insight into the microbiome architecture of IBS and established a robust methodological foundation for future studies. It emphasized the value of integrating metagenomics with computational and statistical frameworks to unravel complex gut disorders. The findings support the development of microbiome-guided therapeutic strategies and open pathways toward precision medicine approaches tailored to IBS and related gastrointestinal diseases. Future studies should incorporate larger, longitudinal cohorts to clarify causality between microbiome alterations and IBS development. Integrating host genomics, metabolomics, and immunological profiling with metagenomics will enable deeper mechanistic insights. Strain-level

functional validation may identify precise microbial targets for therapeutic and diagnostic interventions. Advancing machine-learning models with multi-omics data can support clinically reliable, non-invasive diagnostic tools. Ultimately, these findings may pave the way for personalized medicine approaches, including targeted microbial interventions, precision probiotics, and personalized dietary or microbial restoration therapies tailored to individual IBS phenotypes.

Chapter 06

References

6. References:

Abegaz, F., Abedini, D., Dong, L., & Smilde, A. K. (2025). *Analysis of microbiome experimental design data using generalized linear models and ANOVA simultaneous component analysis*. *October*, 1–16. https://doi.org/10.3389/frmbi.2025.1584516

Ahammad, I., Bhattacharjee, A., Chowdhury, Z. M., Rahman, A., Hossain, M. U., Dewan, G., Talukder, S., Das, K. C., Keya, C. A., & Salimullah, M. (2024). Gut microbiome composition reveals the distinctiveness between the Bengali people and the Indigenous ethnicities in Bangladesh. *Communications Biology*, *7*(1), 500. https://doi.org/10.1038/s42003-024-06191-9

Ahmadi, A., Kouhsari, E., Razavi, S., Mohamadzadeh, N., Besharat, S., Vakili, M. A., & Amiriani, T. (2025). Comparative analysis of dominant gut microbiota in Inflammatory Bowel Disease patients and healthy individuals: A case-control study. *New Microbes and New Infections*, *64*, 101567. https://doi.org/https://doi.org/10.1016/j.nmni.2025.101567

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery \& Data Mining*, 2623–2631. https://doi.org/10.1145/3292500.3330701

Bakhshi, H., Bakhshialiabad, M. H., & Hassanshahi, G. (2014). *Students ' perceptions of the educational environment in an Iranian Medical School , as measured by The Dundee*. 36–41.

Barbara, G., Cremon, C., Carini, G., Bellacosa, L., Zecchi, L., De Giorgio, R., Corinaldesi, R., & Stanghellini, V. (2011). The immune system in irritable bowel syndrome. *Journal of Neurogastroenterology and Motility*, *17*(4), 349–359. https://doi.org/10.5056/jnm.2011.17.4.349

Batuman, O., Britt-Ugartemendia, K., Kunwar, S., Yilmaz, S., Fessler, L., Redondo, A., Chumachenko, K., Chakravarty, S., & Wade, T. (2024). The Use and Impact of Antibiotics in Plant Agriculture: A Review. *Phytopathology®*, *114*(5), 885–909. https://doi.org/10.1094/PHYTO-10-23-0357-IA

Belete, D. M., & Huchaiah, M. D. (2022). Grid search in hyperparameter optimization of machine

learning models for prediction of HIV/AIDS test results. *International Journal of Computers and Applications*, *44*(9), 875–886. https://doi.org/10.1080/1206212X.2021.1974663

Belkaid, Y., & Hand, T. W. (2014). Role of the Microbiota in Immunity and Inflammation. *Cell*, *157*(1), 121–141. https://doi.org/10.1016/j.cell.2014.03.011

Bengio, Y. (2000). Gradient-Based Optimization of Hyperparameters. *Neural Computation*, *12*(8), 1889–1900. https://doi.org/10.1162/089976600300015187

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *J. Mach. Learn. Res.*, *13*(null), 281–305.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Brage, A., Pérez-Coleman, L. V., & Giacomodonato, R. (2013). Pseudomelanic Polyposis of the Colon. *Gastroenterology*, *144*(3), e3–e4. https://doi.org/10.1053/j.gastro.2012.09.045

Brewster, R., Tamburini, F. B., Asiimwe, E., Oduaran, O., Hazelhurst, S., & Bhatt, A. S. (2019). Surveying Gut Microbiome Research in Africans: Toward Improved Diversity and Representation. *Trends in Microbiology*, *27*(10), 824–835. https://doi.org/10.1016/j.tim.2019.05.006

Buja, A., & Kass, R. E. (1985). Comment. *Journal of the American Statistical Association*, *80*(391), 602–607. https://doi.org/10.1080/01621459.1985.10478159

Bull, M. J., & Plummer, N. T. (2014). Part 1: The Human Gut Microbiome in Health and Disease. *Integrative Medicine (Encinitas, Calif.)*, *13*(6), 17–22.

Callaghan, A. O., & Sinderen, D. Van. (2016). *Bifidobacteria and Their Role as Members of the Human Gut Microbiota*. *7*(June). https://doi.org/10.3389/fmicb.2016.00925

Canavan, C. (2014). *CLEP-40245-the-epidemiology-of-irritable-bowel-syndrome*. 71–80.

Carco, C., Young, W., Gearry, R. B., Talley, N. J., McNabb, W. C., & Roy, N. C. (2020). Increasing Evidence That Irritable Bowel Syndrome and Functional Gastrointestinal Disorders Have a Microbial Pathogenesis. *Frontiers in Cellular and Infection Microbiology*,

10(September), 1–24. https://doi.org/10.3389/fcimb.2020.00468

Caruana, E. J., Roman, M., Hernández-Sánchez, J., & Solli, P. (2015). Longitudinal studies. *Journal of Thoracic Disease*, *7*(11), E537-40. https://doi.org/10.3978/j.issn.2072-1439.2015.10.63

Chan, A. T., & Williams, C. S. (2020). Covering the Cover. *Gastroenterology*, *159*(4), 1193–1194. https://doi.org/10.1053/j.gastro.2020.09.021

Chao, A. (1987). Estimating the Population Size for Capture-Recapture Data with Unequal Catchability. *Biometrics*, *43*(4), 783–791. https://doi.org/10.2307/2531532

Chey, W. D., Kurlander, J., & Eswaran, S. (2015). Irritable bowel syndrome: a clinical review. *JAMA*, *313*(9), 949–958. https://doi.org/10.1001/jama.2015.0954

Chicco, D., Tötsch, N., & Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining*, *14*(1), 13. https://doi.org/10.1186/s13040-021-00244-z

Cho, H., Kim, Y., Lee, E., Choi, D., Lee, Y., & Rhee, W. (2020). Basic Enhancement Strategies When Using Bayesian Optimization for Hyperparameter Tuning of Deep Neural Networks. *IEEE Access*, *8*, 52588–52608. https://doi.org/10.1109/ACCESS.2020.2981072

Colella, M., Charitos, I. A., Ballini, A., Cafiero, C., Topi, S., Palmirotta, R., & Santacroce, L. (2023). Microbiota revolution: How gut microbes regulate our lives. *World Journal of Gastroenterology*, *29*(28), 4368–4383. https://doi.org/10.3748/wjg.v29.i28.4368

Compaoré, M., Meda, R. N.-T., Bakasso, S., Vlase, L., & Kiendrebeogo, M. (2016). Antioxidative, anti-inflammatory potentials and phytochemical profile of Commiphora africana (A. Rich.) Engl. (Burseraceae) and Loeseneriella africana (Willd.) (Celastraceae) stem leaves extracts. *Asian Pacific Journal of Tropical Biomedicine*, *6*(8), 665–670. https://doi.org/https://doi.org/10.1016/j.apjtb.2016.06.001

Costea, P. I., Zeller, G., Sunagawa, S., Pelletier, E., Alberti, A., Levenez, F., Tramontano, M., Driessen, M., Hercog, R., Jung, F.-E., Kultima, J. R., Hayward, M. R., Coelho, L. P., Allen-Vercoe, E., Bertrand, L., Blaut, M., Brown, J. R. M., Carton, T., Cools-Portier, S., … Bork,

P. (2017). Towards standards for human fecal sample processing in metagenomic studies. *Nature Biotechnology*, *35*(11), 1069–1076. https://doi.org/10.1038/nbt.3960

Cryan, J. F., & Dinan, T. G. (2012). Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. *Nature Reviews Neuroscience*, *13*(10), 701–712. https://doi.org/10.1038/nrn3346

D., T. B., A., L. N., T., R. M., Jenna, W., & D., S. P. (2020). A Framework for Effective Application of Machine Learning to Microbiome-Based Classification Problems. *MBio*, *11*(3), 10.1128/mbio.00434-20. https://doi.org/10.1128/mbio.00434-20

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., Ling, A. V, Devlin, A. S., Varma, Y., Fischbach, M. A., Biddinger, S. B., Dutton, R. J., & Turnbaugh, P. J. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, *505*(7484), 559–563. https://doi.org/10.1038/nature12820

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J. B., Massart, S., Collini, S., Pieraccini, G., & Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proceedings of the National Academy of Sciences*, *107*(33), 14691–14696. https://doi.org/10.1073/pnas.1005963107

Dethlefsen, L., & Relman, D. A. (2011). Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proceedings of the National Academy of Sciences*, *108*(supplement_1), 4554–4561. https://doi.org/10.1073/pnas.1000087107

Dhariwal, A., Chong, J., Habib, S., King, I. L., Agellon, L. B., & Xia, J. (2017). MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Research*, *45*(W1), W180–W188. https://doi.org/10.1093/nar/gkx295

Didari, T., Mozaffari, S., Nikfar, S., & Abdollahi, M. (2015). Effectiveness of probiotics in irritable bowel syndrome: Updated systematic review with meta-analysis . *World Journal of Gastroenterology*, *21*(10), 3072–3084. https://doi.org/10.3748/wjg.v21.i10.3072

Dong, T. S., & Gupta, A. (2019). Influence of Early Life, Diet, and the Environment on the

Microbiome. *Clinical Gastroenterology and Hepatology*, *17*(2), 231–242. https://doi.org/https://doi.org/10.1016/j.cgh.2018.08.067

Drossman, D. A. (2016). Functional Gastrointestinal Disorders: History, Pathophysiology, Clinical Features, and Rome IV. *Gastroenterology*, *150*(6), 1262-1279.e2. https://doi.org/10.1053/j.gastro.2016.02.032

El-Salhy, M., Hatlebakk, J. G., Gilja, O. H., Bråthen Kristoffersen, A., & Hausken, T. (2020). Efficacy of faecal microbiota transplantation for patients with irritable bowel syndrome in a randomised, double-blind, placebo-controlled study. *Gut*, *69*(5), 859 LP – 867. https://doi.org/10.1136/gutjnl-2019-319630

Fahim, S. M., Hossain, M. S., Sen, S., Das, S., Hosssain, M., Ahmed, T., Rahman, S. M. M., Rahman, M. K., & Alam, S. (2021). Nutrition and Food Security in Bangladesh: Achievements, Challenges, and Impact of the COVID-19 Pandemic. *The Journal of Infectious Diseases*, *224*(Supplement_7), S901–S909. https://doi.org/10.1093/infdis/jiab473

Filardo, S., Di Pietro, M., & Sessa, R. (2024). Current progresses and challenges for microbiome research in human health: a perspective. *Frontiers in Cellular and Infection Microbiology*, *Volume 14*. https://doi.org/10.3389/fcimb.2024.1377012

Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P., & Forano, E. (2012). Microbial degradation of complex carbohydrates in the gut. *Gut Microbes*, *3*(4), 289–306. https://doi.org/10.4161/gmic.19897

Forry, S. P., Servetas, S. L., Kralj, J. G., Soh, K., Hadjithomas, M., Cano, R., Carlin, M., Amorim, M. G. de, Auch, B., Bakker, M. G., Bartelli, T. F., Bustamante, J. P., Cassol, I., Chalita, M., Dias-Neto, E., Duca, A. Del, Gohl, D. M., Kazantseva, J., Haruna, M. T., … Jackson, S. A. (2024). Variability and bias in microbiome metagenomic sequencing: an interlaboratory study comparing experimental protocols. *Scientific Reports*, *14*(1), 9785. https://doi.org/10.1038/s41598-024-57981-4

Forslund, K., Sunagawa, S., Coelho, L. P., & Bork, P. (2014). *shape it*. 316–329. https://doi.org/10.1002/bies.201300143

Franzosa, E. A., McIver, L. J., Rahnavard, G., Thompson, L. R., Schirmer, M., Weingart, G.,

Lipson, K. S., Knight, R., Caporaso, J. G., Segata, N., & Huttenhower, C. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. *Nature Methods*, *15*(11), 962–968. https://doi.org/10.1038/s41592-018-0176-y

Fukui, H., Nishida, A., Matsuda, S., Kira, F., Watanabe, S., Kuriyama, M., Kawakami, K., Aikawa, Y., Oda, N., Arai, K., Matsunaga, A., Nonaka, M., Nakai, K., Shinmura, W., Matsumoto, M., Morishita, S., Takeda, A. K., & Miwa, H. (2020). Usefulness of Machine Learning-Based Gut Microbiome Analysis for Identifying Patients with Irritable Bowels Syndrome. In *Journal of Clinical Medicine* (Vol. 9, Issue 8). https://doi.org/10.3390/jcm9082403

Fusco, W., Lorenzo, M. B., Cintoni, M., Porcari, S., Rinninella, E., Kaitsas, F., Lener, E., Mele, M. C., Gasbarrini, A., Collado, M. C., Cammarota, G., & Ianiro, G. (2023). Short-Chain Fatty-Acid-Producing Bacteria: Key Components of the Human Gut Microbiota. *Nutrients*, *15*(9). https://doi.org/10.3390/nu15092211

Ge, S. X., Jung, D., & Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics*, *36*(8), 2628–2629. https://doi.org/10.1093/bioinformatics/btz931

Ghoshal, U. C., Sachdeva, S., Pratap, N., Karyampudi, A., Mustafa, U., Abraham, P., Bhatt, C. B., Chakravartty, K., Chaudhuri, S., Goyal, O., Makharia, G. K., Panigrahi, M. K., Parida, P. K., Patwari, S., Sainani, R., Sadasivan, S., Srinivas, M., Upadhyay, R., & Venkataraman, J. (2023). Indian consensus statements on irritable bowel syndrome in adults: A guideline by the Indian Neurogastroenterology and Motility Association and jointly supported by the Indian Society of Gastroenterology. *Indian Journal of Gastroenterology : Official Journal of the Indian Society of Gastroenterology*, *42*(2), 249–273. https://doi.org/10.1007/s12664-022-01333-5

Gönen, M. (2006). *NESUG 2006 Data Manipulation and Analysis Receiver Operating Characteristic (ROC) Curves Mithat Gönen , Memorial Sloan-Kettering Cancer Center NESUG 2006 Data Manipulation and Analysis FN + FP*. *2001*, 1–18.

Gralnek, I. M., Hays, R. D., Kilbourne, A., Naliboff, B., & Mayer, E. A. (2000). The impact of irritable bowel syndrome on health-related quality of life. *Gastroenterology*, *119*(3), 654–660. https://doi.org/10.1053/gast.2000.16484

Gun-Ha, K., Kihyun, L., & Ok, S. J. (2023). Gut Bacterial Dysbiosis in Irritable Bowel Syndrome: a Case-Control Study and a Cross-Cohort Analysis Using Publicly Available Data Sets. *Microbiology Spectrum*, *11*(1), e02125-22. https://doi.org/10.1128/spectrum.02125-22

Halmos, E. P., Christophersen, C. T., Bird, A. R., Shepherd, S. J., Gibson, P. R., & Muir, J. G. (2015). Diets that differ in their FODMAP content alter the colonic luminal microenvironment. *Gut*, *64*(1), 93 LP – 100. https://doi.org/10.1136/gutjnl-2014-307264

Heeney, D. D., Gareau, M. G., & Marco, M. L. (2018). Intestinal Lactobacillus in health and disease, a driver or just along for the ride? *Current Opinion in Biotechnology*, *49*, 140–147. https://doi.org/https://doi.org/10.1016/j.copbio.2017.08.004

Hernández Medina, R., Kutuzova, S., Nielsen, K. N., Johansen, J., Hansen, L. H., Nielsen, M., & Rasmussen, S. (2022). Machine learning and deep learning applications in microbiome research. *ISME Communications*, *2*(1), 98. https://doi.org/10.1038/s43705-022-00182-9

Holowatyj, A. N., Gigic, B., Herpel, E., Scalbert, A., Schneider, M., Ulrich, C. M., Achaintre, D., Brezina, S., van Duijnhoven, F. J. B., Gsur, A., Keski-Rahkonen, P., Weijenberg, M. P., Abbenhardt-Martin, C., Boehm, J., Boucher, K., Habermann, N., Haffa, M., Hardikar, S., Himbert, C., … Tavtigian, S. V. (2020). Distinct Molecular Phenotype of Sporadic Colorectal Cancers Among Young Patients Based on Multiomics Analysis. *Gastroenterology*, *158*(4), 1155-1158.e2. https://doi.org/https://doi.org/10.1053/j.gastro.2019.11.012

Ianiro, G., Porcari, S., Settanni, C. R., Bibbò, S., Ponziani, F. R., Zileri dal Verme, L., Franceschi, F., Cammarota, G., Gasbarrini, A., & Group, G. against C.-19 W. (2020). Letter: prevalence and patterns of gastrointestinal symptoms in a large Western cohort of patients with COVID-19. *Alimentary Pharmacology & Therapeutics*, *52*(5), 902–903. https://doi.org/https://doi.org/10.1111/apt.15946

Jeffery, I. B., O'Toole, P. W., Öhman, L., Claesson, M. J., Deane, J., Quigley, E. M. M., & Simrén, M. (2012). An irritable bowel syndrome subtype defined by species-specific alterations in faecal microbiota. *Gut*, *61*(7), 997 LP – 1006. https://doi.org/10.1136/gutjnl-2011-301501

Jin, S. S., Amnon, A., L., M. J., R., A. K., Zech, X. Z., Greg, H., & Rob, K. (2016). Preservation

Methods Differ in Fecal Microbiome Stability, Affecting Suitability for Field Studies. *MSystems*, *1*(3), 10.1128/msystems.00021-16. https://doi.org/10.1128/msystems.00021-16

Johnsen, P. H., Hilpüsch, F., Cavanagh, J. P., Leikanger, I. S., Kolstad, C., Valle, P. C., & Goll, R. (2018). Faecal microbiota transplantation versus placebo for moderate-to-severe irritable bowel syndrome: a double-blind, randomised, placebo-controlled, parallel-group, single-centre trial. *The Lancet Gastroenterology & Hepatology*, *3*(1), 17–24. https://doi.org/10.1016/S2468-1253(17)30338-2

Jung, K. W., & Myung, S.-J. (2023). An Asian perspective on irritable bowel syndrome. *Intestinal Research*, *21*(2), 189–195. https://doi.org/10.5217/ir.2021.00136

Jyoti, & Dey, P. (2025). Mechanisms and implications of the gut microbial modulation of intestinal metabolic processes. *Npj Metabolic Health and Disease*, *3*(1), 24. https://doi.org/10.1038/s44324-025-00066-1

Kamboj, A. K., Stroh, G. R., Smoot, R. L., & Prichard, D. O. (2018). A Rare Case of Recurrent Small Bowel Obstructions. *Gastroenterology*, *155*(5), e5–e7. https://doi.org/https://doi.org/10.1053/j.gastro.2018.05.048

Karwowska, Z., Aasmets, O., Metspalu, M., Metspalu, A., Milani, L., Esko, T., Kosciolek, T., Org, E., & team, E. B. research. (2025). Effects of data transformation and model selection on feature importance in microbiome classification data. *Microbiome*, *13*(1), 2. https://doi.org/10.1186/s40168-024-01996-6

Kassinen, A., Krogius-Kurikka, L., Mäkivuokko, H., Rinttilä, T., Paulin, L., Corander, J., Malinen, E., Apajalahti, J., & Palva, A. (2007). The Fecal Microbiota of Irritable Bowel Syndrome Patients Differs Significantly From That of Healthy Subjects. *Gastroenterology*, *133*(1), 24–33. https://doi.org/10.1053/j.gastro.2007.04.005

Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks*, *4*, 1942–1948 vol.4. https://doi.org/10.1109/ICNN.1995.488968

Khan, I., Ullah, N., Zha, L., Bai, Y., Khan, A., Zhao, T., Che, T., & Zhang, C. (2019). Alteration of Gut Microbiota in Inflammatory Bowel Disease (IBD): Cause or Consequence? IBD

Treatment Targeting the Gut Microbiome. *Pathogens (Basel, Switzerland)*, *8*(3). https://doi.org/10.3390/pathogens8030126

Knight, R., Vrbanac, A., Taylor, B. C., Aksenov, A., Callewaert, C., Debelius, J., Gonzalez, A., Kosciolek, T., McCall, L.-I., McDonald, D., Melnik, A. V, Morton, J. T., Navas, J., Quinn, R. A., Sanders, J. G., Swafford, A. D., Thompson, L. R., Tripathi, A., Xu, Z. Z., … Dorrestein, P. C. (2018). Best practices for analysing microbiomes. *Nature Reviews Microbiology*, *16*(7), 410–422. https://doi.org/10.1038/s41579-018-0029-9

Konopiński, M. K. (2020). Shannon diversity index: a call to replace the original Shannon's formula with unbiased estimator in the population genetics studies. *PeerJ*, *8*, e9391. https://doi.org/10.7717/peerj.9391

Lacy, B. E., & Patel, N. K. (2017). Rome Criteria and a Diagnostic Approach to Irritable Bowel Syndrome. In *Journal of Clinical Medicine* (Vol. 6, Issue 11). https://doi.org/10.3390/jcm6110099

Lamb, C. A., Kennedy, N. A., Raine, T., Hendy, P. A., Smith, P. J., Limdi, J. K., Hayee, B., Lomer, M. C. E., Parkes, G. C., Selinger, C., Barrett, K. J., Davies, R. J., Bennett, C., Gittens, S., Dunlop, M. G., Faiz, O., Fraser, A., Garrick, V., Johnston, P. D., … Hawthorne, A. B. (2019). British Society of Gastroenterology consensus guidelines on the management of inflammatory bowel disease in adults. *Gut*, *68*(Suppl 3), s1 LP-s106. https://doi.org/10.1136/gutjnl-2019-318484

Langelier, C., Lu, D., Kalantar, K., Chu, V., Glascock, A., Guerrero, E., Bernick, N., Butcher, X., Ewing, K., Fahsbender, E., Holmes, O., Hoops, E., Jones, A., Lim, R., McCanny, S., Reynoso, L., Rosario, K., Tang, J., Valenzuela, O., … McArthur, A. (2024). Simultaneous detection of pathogens and antimicrobial resistance genes with the open source, cloud-based, CZ ID pipeline. *Research Square*. https://doi.org/10.21203/rs.3.rs-4271356/v1

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods 2012 9:4*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lee, B. W., Ha, T. K. Q., Cho, H. M., An, J.-P., Kim, S. K., Kim, C.-S., Kim, E., & Oh, W. K. (2020). Antiviral activity of furanocoumarins isolated from Angelica dahurica against

influenza a viruses H1N1 and H9N2. *Journal of Ethnopharmacology*, *259*, 112945. https://doi.org/https://doi.org/10.1016/j.jep.2020.112945

Leventi, A., Clifford, T. F. J., Arnold, A., Knowles, C. H., & Martin, J. E. (2018). A case of sigmoid volvulus. *Gut*, *67*(12), 2084 LP – 2123. https://doi.org/10.1136/gutjnl-2017-315465

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/BIOINFORMATICS/BTP352

Li, J., Ghosh, T. S., Arendt, E., Shanahan, F., & O'Toole, P. W. (2024). Cross-Cohort Gut Microbiome Signatures of Irritable Bowel Syndrome Presentation and Treatment. *Advanced Science*, *11*(41), 2308313. https://doi.org/https://doi.org/10.1002/advs.202308313

Lin, A., Bik, E. M., Costello, E. K., Dethlefsen, L., Haque, R., Relman, D. A., & Singh, U. (2013). Distinct Distal Gut Microbiome Diversity and Composition in Healthy Children from Bangladesh and the United States. *PLOS ONE*, *8*(1), e53838. https://doi.org/10.1371/journal.pone.0053838

Liu, B., Zheng, D., Jin, Q., Chen, L., & Yang, J. (2019). VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Research*, *47*(D1), D687–D692. https://doi.org/10.1093/NAR/GKY1080

Lloyd-Price, J., Mahurkar, A., Rahnavard, G., Crabtree, J., Orvis, J., Hall, A. B., Brady, A., Creasy, H. H., McCracken, C., Giglio, M. G., McDonald, D., Franzosa, E. A., Knight, R., White, O., & Huttenhower, C. (2017). Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature*, *550*(7674), 61–66. https://doi.org/10.1038/nature23889

Lo, C.-C., & Chain, P. S. G. (2014). Rapid evaluation and quality control of next generation sequencing data with FaQCs. *BMC Bioinformatics*, *15*(1), 366. https://doi.org/10.1186/s12859-014-0366-2

Lopez-Siles, M., Duncan, S. H., Garcia-Gil, L. J., & Martinez-Medina, M. (2017). Faecalibacterium prausnitzii: from microbiology to diagnostics and prognostics. *The ISME Journal*, *11*(4), 841–852. https://doi.org/10.1038/ismej.2016.176

Lovell, R. M., & Ford, A. C. (2012). Global Prevalence of and Risk Factors for Irritable Bowel

Syndrome: A Meta-analysis. *Clinical Gastroenterology and Hepatology*, *10*(7), 712-721.e4. https://doi.org/10.1016/j.cgh.2012.02.029

Maaten, L. Van Der, & Hinton, G. (2008). *Visualizing Data using t-SNE*. *9*, 2579–2605.

Macura, B., Kiecka, A., & Szczepanik, M. (2024). Intestinal permeability disturbances: causes, diseases and therapy. *Clinical and Experimental Medicine*, *24*(1), 232. https://doi.org/10.1007/s10238-024-01496-9

Malfertheiner, P., & Delchier, J. C. (2014). Letter: bismuth quadruple therapy with Pylera for Helicobacter pylori infection – authors' reply. *Alimentary Pharmacology & Therapeutics*, *40*(6), 736–737. https://doi.org/https://doi.org/10.1111/apt.12909

Mandal, R. S., Saha, S., & Das, S. (2015). Metagenomic Surveys of Gut Microbiota. *Genomics, Proteomics & Bioinformatics*, *13*(3), 148–158. https://doi.org/10.1016/j.gpb.2015.02.005

Mannan, A., Chakma, K., Dewan, G., Saha, A., Chy, N. U. H. A., Mehedi, H. M. H., Hossain, A., Wnaiza, J., Ahsan, M. T., Rana, M. M., & Alam, N. (2024). Prevalence and determinants of antibiotics self-medication among indigenous people of Bangladesh: a cross-sectional study. *BMJ Open*, *14*(3), e071504. https://doi.org/10.1136/bmjopen-2022-071504

Manrique, P., Dills, M., & Young, M. J. (2017). The Human Gut Phage Community and Its Implications for Health and Disease. *Viruses*, *9*(6). https://doi.org/10.3390/v9060141

Martínez, J. L., Coque, T. M., & Baquero, F. (2015). What is a resistance gene? Ranking risk in resistomes. *Nature Reviews Microbiology*, *13*(2), 116–123. https://doi.org/10.1038/nrmicro3399

Mayer, E. A., Tillisch, K., Gupta, A., Mayer, E. A., Tillisch, K., & Gupta, A. (2015a). *Gut / brain axis and the microbiota*. *125*(3), 926–938. https://doi.org/10.1172/JCI76304.Several

Mayer, E. A., Tillisch, K., Gupta, A., Mayer, E. A., Tillisch, K., & Gupta, A. (2015b). Gut / brain axis and the microbiota Find the latest version : Gut / brain axis and the microbiota. *Journal of Clinical Investiogation*, *125*(3), 926–938. https://doi.org/10.1172/JCI76304.Several

Mazmanian, S. K., Liu, C. H., Tzianabos, A. O., & Kasper, D. L. (2005). An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell*, *122*(1),

107–118. https://doi.org/10.1016/j.cell.2005.05.007

McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE*, *8*(4), e61217. https://doi.org/10.1371/JOURNAL.PONE.0061217

Medina, S., Romero, E., Cruz-roa, A., & González, F. A. (2025). *Interpretable weakly-supervised learning through kernel density matrices : A digital pathology use case*. 1–20. https://doi.org/10.1371/journal.pone.0335826

Michael, J. J., & L., A. S. (2007). 16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls. *Journal of Clinical Microbiology*, *45*(9), 2761–2764. https://doi.org/10.1128/jcm.01228-07

Miquel, S., Martín, R., Rossi, O., Bermúdez-Humarán, L. G., Chatel, J. M., Sokol, H., Thomas, M., Wells, J. M., & Langella, P. (2013). Faecalibacterium prausnitzii and human intestinal health. *Current Opinion in Microbiology*, *16*(3), 255–261. https://doi.org/10.1016/j.mib.2013.06.003

Mirsepasi, H., Persson, S., Struve, C., Andersen, L. O. B., Petersen, A. M., & Krogfelt, K. A. (2014). Microbial diversity in fecal samples depends on DNA extraction method: easyMag DNA extraction compared to QIAamp DNA stool mini kit extraction. *BMC Research Notes*, *7*(1), 50. https://doi.org/10.1186/1756-0500-7-50

Moran, N. A. (2017). Old and new symbiotic partners in lachnine aphids. *Environmental Microbiology*, *19*(1), 7. https://doi.org/https://doi.org/10.1111/1462-2920.13516

Mostafavi Abdolmaleky, H., & Zhou, J.-R. (2024). Gut Microbiota Dysbiosis, Oxidative Stress, Inflammation, and Epigenetic Alterations in Metabolic Diseases. *Antioxidants (Basel, Switzerland)*, *13*(8). https://doi.org/10.3390/antiox13080985

Nayfach, S., Rodriguez-Mueller, B., Garud, N., & Pollard, K. S. (2016). An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. *Genome Research*, *26*(11), 1612–1625. https://doi.org/10.1101/gr.201863.115

Nicholson, J. K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., & Pettersson, S. (2012).

Host-Gut Microbiota Metabolic Interactions. *Science*, *336*(6086), 1262–1267. https://doi.org/10.1126/science.1223813

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Research*, *27*(5), 824–834. https://doi.org/10.1101/gr.213959.116

O'Halloran, S. A., Lacy, K. E., Grimes, C. A., Woods, J., Campbell, K. J., & Nowson, C. A. (2017). A novel processed food classification system applied to Australian food composition databases. *Journal of Human Nutrition and Dietetics*, *30*(4), 534–541. https://doi.org/https://doi.org/10.1111/jhn.12445

O'Keefe, S. J. D., Li, J. V, Lahti, L., Ou, J., Carbonero, F., Mohammed, K., Posma, J. M., Kinross, J., Wahl, E., Ruder, E., Vipperla, K., Naidoo, V., Mtshali, L., Tims, S., Puylaert, P. G. B., DeLany, J., Krasinskas, A., Benefiel, A. C., Kaseb, H. O., … Zoetendal, E. G. (2015). Fat, fibre and cancer risk in African Americans and rural Africans. *Nature Communications*, *6*(1), 6342. https://doi.org/10.1038/ncomms7342

Oakland, K., Chadwick, G., East, J. E., Guy, R., Humphries, A., Jairath, V., McPherson, S., Metzner, M., Morris, A. J., Murphy, M. F., Tham, T., Uberoi, R., Veitch, A. M., Wheeler, J., Regan, C., & Hoare, J. (2019). Diagnosis and management of acute lower gastrointestinal bleeding: guidelines from the British Society of Gastroenterology. *Gut*, *68*(5), 776 LP – 789. https://doi.org/10.1136/gutjnl-2018-317807

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Sólymos, P., Stevens, M. H. H., & Wagner, H. (2012). *vegan: Community Ecology Package*.

Ondov, B. D., Bergman, N. H., & Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*, *12*(1), 1–10. https://doi.org/10.1186/1471-2105-12-385/FIGURES/4

Palani, D., Bapatdhar, N., & Manchiraju, R. (2024). *IntegrIBS : Evaluating the Potential and Challenges of Building a Robust IBS Classifier with Integrated Microbiome Data*. 1–21.

Palleja, A., Mikkelsen, K. H., Forslund, S. K., Kashani, A., Allin, K. H., Nielsen, T., Hansen, T.

H., Liang, S., Feng, Q., Zhang, C., Pyl, P. T., Coelho, L. P., Yang, H., Wang, J., Typas, A., Nielsen, M. F., Nielsen, H. B., Bork, P., Wang, J., … Pedersen, O. (2018). Recovery of gut microbiota of healthy adults following antibiotic exposure. *Nature Microbiology*, *3*(11), 1255–1265. https://doi.org/10.1038/s41564-018-0257-9

Pathan, M., Keerthikumar, S., Ang, C. S., Gangoda, L., Quek, C. Y. J., Williamson, N. A., Mouradov, D., Sieber, O. M., Simpson, R. J., Salim, A., Bacic, A., Hill, A. F., Stroud, D. A., Ryan, M. T., Agbinya, J. I., Mariadason, J. M., Burgess, A. W., & Mathivanan, S. (2015). FunRich: An open access standalone functional enrichment and interaction network analysis tool. *PROTEOMICS*, *15*(15), 2597–2601. https://doi.org/10.1002/PMIC.201400515

Pethakamsetty, L., Pola, S., & Giduthuri, J. G. (2023). *The Microbiome Antibiotic Resistome: Significant Strategies Toward Microbiome-Targeted Therapeutic Interventions in Medicine BT - Human Microbiome in Health, Disease, and Therapy* (P. Veera Bramhachari (ed.); pp. 241–264). Springer Nature Singapore. https://doi.org/10.1007/978-981-99-5114-7_13

Pham, V. T., Dold, S., Rehman, A., Bird, J. K., & Steinert, R. E. (2021). Vitamins, the gut microbiome and gastrointestinal health in humans. *Nutrition Research*, *95*, 35–53. https://doi.org/https://doi.org/10.1016/j.nutres.2021.09.001

Pittayanon, R., Lau, J. T., Yuan, Y., Leontiadis, G. I., Tse, F., Surette, M., & Moayyedi, P. (2019). Gut Microbiota in Patients With Irritable Bowel Syndrome—A Systematic Review. *Gastroenterology*, *157*(1), 97–108. https://doi.org/https://doi.org/10.1053/j.gastro.2019.03.049

Pozuelo, M., Panda, S., Santiago, A., Mendez, S., Accarino, A., Santos, J., Guarner, F., Azpiroz, F., & Manichanh, C. (2015). Reduction of butyrate- and methane-producing microorganisms in patients with Irritable Bowel Syndrome. *Scientific Reports*, *5*(1), 12693. https://doi.org/10.1038/srep12693

Probst, P. (2019). *Tunability : Importance of Hyperparameters of Machine Learning Algorithms*. *20*, 1–32.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang,

H., Zheng, H., … Consortium, M. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, *464*(7285), 59–65. https://doi.org/10.1038/nature08821

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, *35*(9), 833–844. https://doi.org/10.1038/nbt.3935

Rafiqul Islam, S. M., Foysal, M. J., Hoque, M. N., Mehedi, H. M. H., Rob, M. A., Salauddin, A., Tanzina, A. Y., Biswas, S., Noyon, S. H., Siddiki, A. M. A. M. Z., Tay, A., & Mannan, A. (2022). Dysbiosis of Oral and Gut Microbiomes in SARS-CoV-2 Infected Patients in Bangladesh: Elucidating the Role of Opportunistic Gut Microbes. *Frontiers in Medicine*, *Volume 9*-. https://doi.org/10.3389/fmed.2022.821777

Rajilić–Stojanović, M., Biagi, E., Heilig, H. G. H. J., Kajander, K., Kekkonen, R. A., Tims, S., & de Vos, W. M. (2011). Global and Deep Molecular Analysis of Microbiota Signatures in Fecal Samples From Patients With Irritable Bowel Syndrome. *Gastroenterology*, *141*(5), 1792–1801. https://doi.org/10.1053/j.gastro.2011.07.043

Ringel-Kulka, T., Palsson, O. S., Maier, D., Carroll, I., Galanko, J. A., Leyer, G., & Ringel, Y. (2011). Probiotic bacteria Lactobacillus acidophilus NCFM and Bifidobacterium lactis Bi-07 versus placebo for the symptoms of bloating in patients with functional bowel disorders: a double-blind study. *Journal of Clinical Gastroenterology*, *45*(6), 518–525. https://doi.org/10.1097/MCG.0b013e31820ca4d6

Sánchez-Pellicer, P., Álamo-Marzo, J. M., Martínez-Villaescusa, M., Núñez-Delegido, E., Such-Ronda, J. F., Huertas-López, F., Serrano-López, E. M., Martínez-Moreno, D., & Navarro-López, V. (2025). Comparative Analysis of Gut Microbiota in Patients with Irritable Bowel Syndrome and Healthy Controls. *Journal of Clinical Medicine*, *14*(4). https://doi.org/10.3390/jcm14041198

Schirmer, M., Smeekens, S. P., Vlamakis, H., Jaeger, M., Oosting, M., Franzosa, E. A., ter Horst, R., Jansen, T., Jacobs, L., Bonder, M. J., Kurilshikov, A., Fu, J., Joosten, L. A. B., Zhernakova, A., Huttenhower, C., Wijmenga, C., Netea, M. G., & Xavier, R. J. (2016). Linking the Human Gut Microbiome to Inflammatory Cytokine Production Capacity. *Cell*, *167*(4), 1125-1136.e8. https://doi.org/10.1016/j.cell.2016.10.020

Scutari, R., Alteri, C., Vicenti, I., Di Carlo, D., Zuccaro, V., Incardona, F., Borghi, V., Bezenchek, A., Andreoni, M., Antinori, A., Perno, C. F., Cascio, A., De Luca, A., Zazzi, M., & Santoro, M. M. (2020). Evaluation of HIV-1 integrase resistance emergence and evolution in patients treated with integrase inhibitors. *Journal of Global Antimicrobial Resistance*, *20*, 163–169. https://doi.org/https://doi.org/10.1016/j.jgar.2019.07.015

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., & Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. *Genome Biology*, *12*(6), R60. https://doi.org/10.1186/gb-2011-12-6-r60

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., & Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods*, *9*(8), 811–814. https://doi.org/10.1038/nmeth.2066

Sekirov, I., Russell, S. L., Antunes, L. C. M., & Finlay, B. B. (2010). Gut Microbiota in Health and Disease. *Physiological Reviews*, *90*(3), 859–904. https://doi.org/10.1152/physrev.00045.2009

Sharpton, T. J. (2014). An introduction to the analysis of shotgun metagenomic data. *Frontiers in Plant Science*, *Volume 5*-. https://doi.org/10.3389/fpls.2014.00209

Shin, N.-R., Whon, T. W., & Bae, J.-W. (2015). Proteobacteria: microbial signature of dysbiosis in gut microbiota. *Trends in Biotechnology*, *33*(9), 496–503. https://doi.org/10.1016/j.tibtech.2015.06.011

Shireen, K., Weidong, C., William, G., Katie, B., & J., M. A. (2017). Spread from the Sink to the Patient: In Situ Study Using Green Fluorescent Protein (GFP)-Expressing Escherichia coli To Model Bacterial Dispersion from Hand-Washing Sink-Trap Reservoirs. *Applied and Environmental Microbiology*, *83*(8), e03327-16. https://doi.org/10.1128/AEM.03327-16

Shkoporov, A. N., & Hill, C. (2019). Bacteriophages of the Human Gut: The “Known Unknown” of the Microbiome. *Cell Host & Microbe*, *25*(2), 195–209. https://doi.org/10.1016/j.chom.2019.01.017

Shreiner, A. B., Kao, J. Y., & Young, V. B. (2015). The gut microbiome in health and in disease. *Current Opinion in Gastroenterology*, *31*(1). https://journals.lww.com/co-

gastroenterology/fulltext/2015/01000/the_gut_microbiome_in_health_and_in_disease.12.aspx

Simrén, M., Barbara, G., Flint, H. J., Spiegel, B. M. R., Spiller, R. C., Vanner, S., Verdu, E. F., Whorwell, P. J., & Zoetendal, E. G. (2013). Intestinal microbiota in functional bowel disorders: a Rome foundation report. *Gut*, *62*(1), 159–176. https://doi.org/10.1136/gutjnl-2012-302167

Siraj, M. J., Ahmad, T., & Ijtihadie, R. M. (2022). *Analyzing ANOVA F -test and Sequential Feature Selection for Intrusion Detection Systems*. *14*(2). https://doi.org/10.15849/IJASCA.220720.13

Skvortsova, O., Kaiumov, K., Ulivanova, V., Alekseev, D., Lyamin, A., Migacheva, N., & Ismatullin, D. (2024). Prevalence of Sutterella wadsworthensis in the fecal microbiota of obese children. *Russian Journal of Infection and Immunity*, *14*, 187–190. https://doi.org/10.15789/2220-7619-POS-16600

Somerfield, P. J., Clarke, K. R., & Warwick, R. M. (2008). Simpson index. In S. V Jorgensen, B. B. T.-E. of E. Fath, S. V Jorgensen, & B. Fath (Eds.), *Encyclopedia of Ecology* (pp. 3252–3255). Elsevier. https://doi.org/10.1016/B978-008045405-4.00084-7

Sommer, M. O. A., Dantas, G., & Church, G. M. (2009). Functional Characterization of the Antibiotic Resistance Reservoir in the Human Microflora. *Science*, *325*(5944), 1128–1131. https://doi.org/10.1126/science.1176950

Spiller, R., & Major, G. (2016). IBS and IBD — separate entities or on a spectrum? *Nature Reviews Gastroenterology & Hepatology*, *13*(10), 613–621. https://doi.org/10.1038/nrgastro.2016.141

Sun, H., Zhu, C., Fu, X., Khattak, S., Wang, J., Liu, Z., Kong, Q., Mou, H., & Secundo, F. (2022). Effects of intestinal microbiota on physiological metabolism and pathogenicity of Vibrio. *Frontiers in Microbiology*, *13*, 947767. https://doi.org/10.3389/fmicb.2022.947767

Tap, J., Derrien, M., Törnblom, H., Brazeilles, R., Cools-Portier, S., Doré, J., Störsrud, S., Le Nevé, B., Öhman, L., & Simrén, M. (2017). Identification of an Intestinal Microbiota Signature Associated With Severity of Irritable Bowel Syndrome. *Gastroenterology*, *152*(1),

111-123.e8. https://doi.org/10.1053/j.gastro.2016.09.049

Tarazona, S., Lopes, M. B., Klammsteiner, T., Ibrahimi, E., Eckenberger, J., Tonda, A., Simeon, A., Tangaro, S., Lahti, L., Temko, A., & Tan, Z. (2023). *Machine learning approaches in microbiome research : challenges and best practices*. *September*, 1–21. https://doi.org/10.3389/fmicb.2023.1261889

Tett, A., Huang, K. D., Asnicar, F., Fehlner-Peach, H., Pasolli, E., Karcher, N., Armanini, F., Manghi, P., Bonham, K., Zolfo, M., De Filippis, F., Magnabosco, C., Bonneau, R., Lusingu, J., Amuasi, J., Reinhard, K., Rattei, T., Boulund, F., Engstrand, L., … Segata, N. (2019). The Prevotella copri Complex Comprises Four Distinct Clades Underrepresented in Westernized Populations. *Cell Host and Microbe*, *26*(5), 666-679.e7. https://doi.org/10.1016/J.CHOM.2019.08.018/ATTACHMENT/EA844E49-40CE-4DF3-B25D-3B09A250022A/MMC6.XLSX

Tu, H., Xu, J., Wei, L., Kong, Z., & Cai, L. (2025). *Comparison with healthy controls : meta-analysis of changes in intestinal fl ora in Chinese patients with irritable bowel syndrome*. *October*, 1–14. https://doi.org/10.3389/fcimb.2025.1657501

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., Sogin, M. L., Jones, W. J., Roe, B. A., Affourtit, J. P., Egholm, M., Henrissat, B., Heath, A. C., Knight, R., & Gordon, J. I. (2009). A core gut microbiome in obese and lean twins. *Nature*, *457*(7228), 480–484. https://doi.org/10.1038/nature07540

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The Human Microbiome Project. *Nature*, *449*(7164), 804–810. https://doi.org/10.1038/nature06244

Veitch, A. M., Uedo, N., Yao, K., & East, J. E. (2015). Optimizing early upper gastrointestinal cancer detection at endoscopy. *Nature Reviews Gastroenterology & Hepatology*, *12*(11), 660–667. https://doi.org/10.1038/nrgastro.2015.128

Wang, T., Liu, J., Shen, L., Tonti-Filippini, J., Zhu, Y., Jia, H., Lister, R., Whitaker, J. W., Ecker, J. R., Millar, A. H., Ren, B., & Wang, W. (2013). STAR: an integrated solution to management and visualization of sequencing data. *Bioinformatics*, *29*(24), 3204–3210.

https://doi.org/10.1093/bioinformatics/btt558

Wang, Y., Lu, J., Mao, L., Li, J., Yuan, Z., Bond, P. L., & Guo, J. (2019). Antiepileptic drug carbamazepine promotes horizontal transfer of plasmid-borne multi-antibiotic resistance genes within and across bacterial genera. *The ISME Journal*, *13*(2), 509–522. https://doi.org/10.1038/s41396-018-0275-x

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, *20*(1), 257. https://doi.org/10.1186/s13059-019-1891-0

Wu, J., Chen, X.-Y., Zhang, H., Xiong, L.-D., Lei, H., & Deng, S.-H. (2019). Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimizationb. *Journal of Electronic Science and Technology*, *17*(1), 26–40. https://doi.org/https://doi.org/10.11989/JEST.1674-862X.80904120

Wu, T. D., Reeder, J., Lawrence, M., Becker, G., & Brauer, M. J. (2016). *GMAP and GSNAP for Genomic Sequence Alignment: Enhancements to Speed, Accuracy, and Functionality BT - Statistical Genomics: Methods and Protocols* (E. Mathé & S. Davis (eds.); pp. 283–334). Springer New York. https://doi.org/10.1007/978-1-4939-3578-9_15

Xia, Y., & Sun, J. (2023a). *Alpha Diversity BT - Bioinformatic and Statistical Analysis of Microbiome Data: From Raw Sequences to Advanced Modeling with QIIME 2 and R* (Y. Xia & J. Sun (eds.); pp. 289–333). Springer International Publishing. https://doi.org/10.1007/978-3-031-21391-5_9

Xia, Y., & Sun, J. (2023b). *Linear Mixed-Effects Models for Longitudinal Microbiome Data BT - Bioinformatic and Statistical Analysis of Microbiome Data: From Raw Sequences to Advanced Modeling with QIIME 2 and R* (Y. Xia & J. Sun (eds.); pp. 557–586). Springer International Publishing. https://doi.org/10.1007/978-3-031-21391-5_15

Xu, X., Ocansey, D. K. W., Hang, S., Wang, B., Amoah, S., Yi, C., Zhang, X., Liu, L., & Mao, F. (2022). The gut metagenomics and metabolomics signature in patients with inflammatory bowel disease. *Gut Pathogens*, *14*(1), 26. https://doi.org/10.1186/s13099-022-00499-9

Xu, X., Xie, Z., Yang, Z., Li, D., & Xu, X. (2020). *A t-SNE Based Classification Approach to Compositional Microbiome Data*. *11*(December), 1–10.

https://doi.org/10.3389/fgene.2020.620143

Yongfei, H., Xi, Y., Jing, L., Na, L., Fei, L., Jun, W., C., L. I. Y., Na, W., C., W. B., F., G. G., Yulan, L., & Baoli, Z. (2016). The Bacterial Mobile Resistome Transfer Network Connecting the Animal and Human Microbiomes. *Applied and Environmental Microbiology*, *82*(22), 6672–6681. https://doi.org/10.1128/AEM.01802-16

Zhang, S., Wang, B., Wan, L., & Li, L. M. (2017). Estimating Phred scores of Illumina base calls by logistic regression and sparse modeling. *BMC Bioinformatics*, *18*(1), 335. https://doi.org/10.1186/s12859-017-1743-4

Zhang, Z., & Howe, S. (2022). *Weaning Time Affects the Archaeal Community Structure and Functional Potential in Pigs*. *13*(March). https://doi.org/10.3389/fmicb.2022.845621

Zhao, Y., Tang, H., & Ye, Y. (2012). RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. *Bioinformatics*, *28*(1), 125–126. https://doi.org/10.1093/bioinformatics/btr595

Zhernakova, A., Kurilshikov, A., Bonder, M. J., Tigchelaar, E. F., Schirmer, M., Vatanen, T., Mujagic, Z., Vila, A. V., Falony, G., Vieira-Silva, S., Wang, J., Imhann, F., Brandsma, E., Jankipersadsing, S. A., Joossens, M., Cenit, M. C., Deelen, P., Swertz, M. A., study, L. cohort, … Fu, J. (2016). Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science*, *352*(6285), 565–569. https://doi.org/10.1126/science.aad3369

Zhou, H., He, K., Chen, J., & Zhang, X. (2022). LinDA: linear models for differential abundance analysis of microbiome compositional data. *Genome Biology*, *23*(1), 95. https://doi.org/10.1186/s13059-022-02655-5

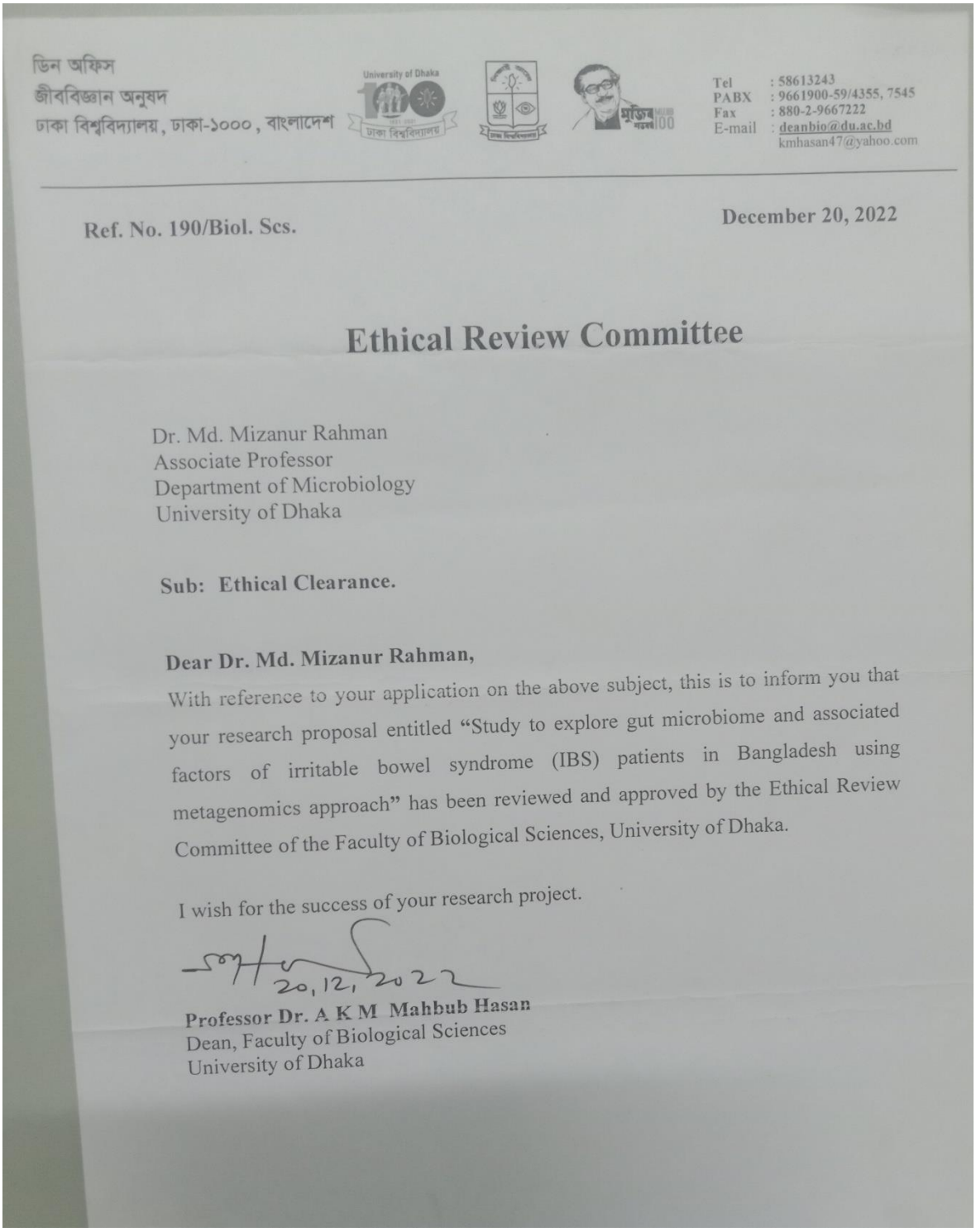
Zhu, B., Wang, X., & Li, L. (2010). Human gut microbiome: the second genome of human body. *Protein & Cell*, *1*(8), 718–725. https://doi.org/10.1007/s13238-010-0093-z

Zoetendal, E. G., Rajilić-Stojanović, M., & de Vos, W. M. (2008). High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota. *Gut*, *57*(11), 1605 LP – 1615. https://doi.org/10.1136/gut.2007.133603

Chapter 07

Appendices

7. Appendices



ডিন অফিস
জীববিজ্ঞান অনুষদ
ঢাকা বিশ্ববিদ্যালয়, ঢাকা-১০০০, বাংলাদেশ

Tel : 58613243
PABX : 9661900-59/4355, 7545
Fax : 880-2-9667222
E-mail : deanbio@du.ac.bd
kmhasan47@yahoo.com

Ref. No. 190/Biol. Scs.

December 20, 2022

Ethical Review Committee

Dr. Md. Mizanur Rahman
Associate Professor
Department of Microbiology
University of Dhaka

Sub: Ethical Clearance.

Dear Dr. Md. Mizanur Rahman,

With reference to your application on the above subject, this is to inform you that your research proposal entitled "Study to explore gut microbiome and associated factors of irritable bowel syndrome (IBS) patients in Bangladesh using metagenomics approach" has been reviewed and approved by the Ethical Review Committee of the Faculty of Biological Sciences, University of Dhaka.

I wish for the success of your research project.

20,12,2022

Professor Dr. A K M Mahbub Hasan
Dean, Faculty of Biological Sciences
University of Dhaka

Ethical clearance letter from IRB, University of Dhaka

Participant Consent Form

Title: Identification and characterization of microbial signatures and potential biomarkers for Irritable bowel syndrome (IBS) in Bangladesh

Declaration by Participant

I have read the Participant Information Sheet or someone has read it to me in a language that I understand.

I understand the purposes, procedures and risks of the research described in the project.

I have had an opportunity to ask questions and I am satisfied with the answers I have received.

I freely agree to participate in this research project as described and understand that I am free to withdraw at any time during the project without affecting

I understand that I will be given a signed copy of this document to keep.

Name of Participant_____

Signature_____Date _____

Declaration by Person performing the informed consent discussion

I have given a verbal explanation of the research project, its procedures and risks and I believe that the participant has understood that explanation.

Name _____

Signature_____Date _____

Patient's information:

Name:

Address:

Phone Number:

Age:

Gender:

History of taking antibiotics past 6-12 months:

Stool types: i. Constipation (C)

ii. Diarrhea (D)

iii. Mixed (M)

iv. Alternating (A)

Lactose Intolerance: Yes No

Education:

Marital status: Married Single

Physical condition:

i. Diabetes Yes No

ii. Hypertension Yes No

iii. Others:

History of IBS in patient's family members:

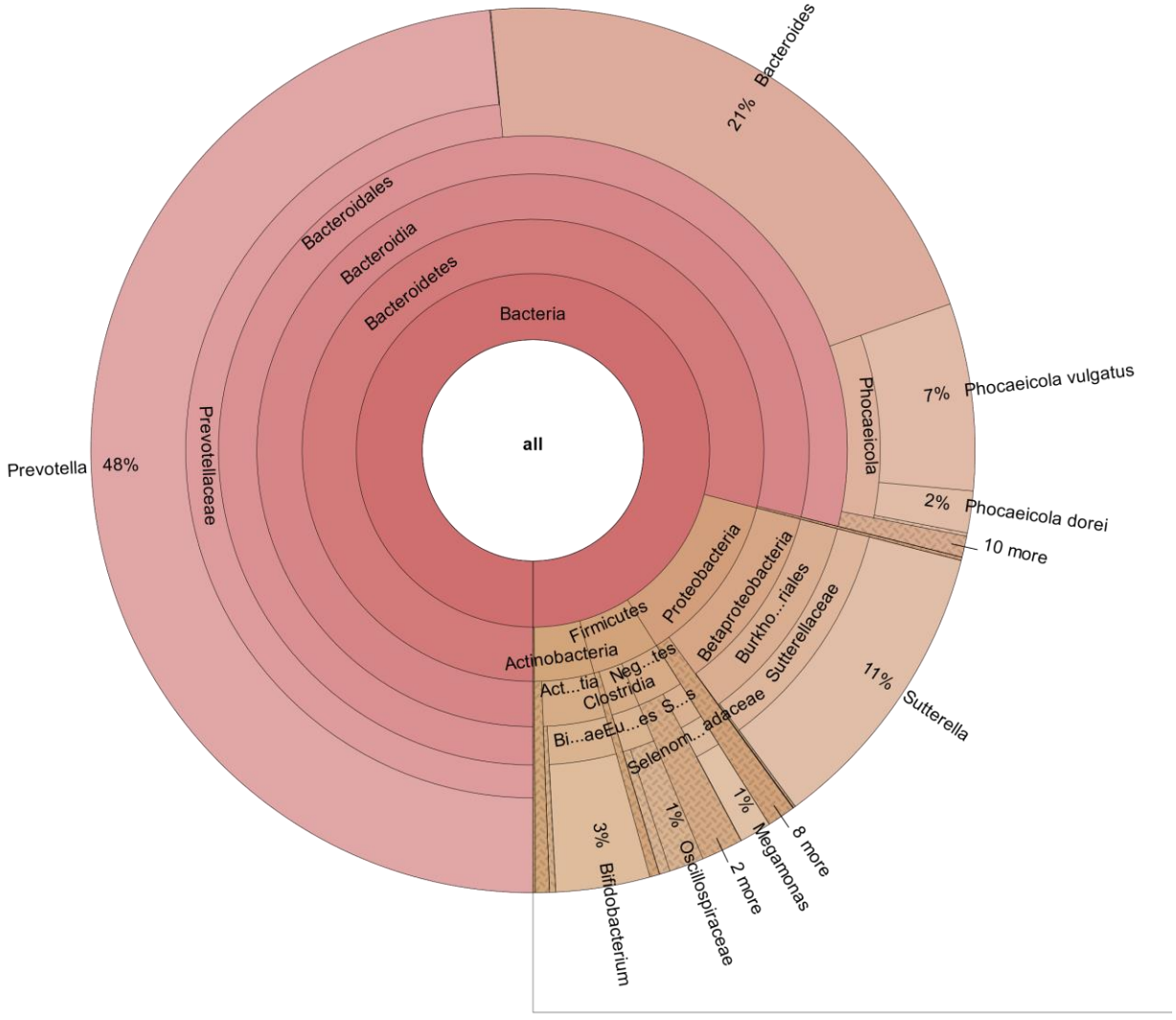


Figure S1: Krona plot showing the distribution of bacterial species in the gut of HC1

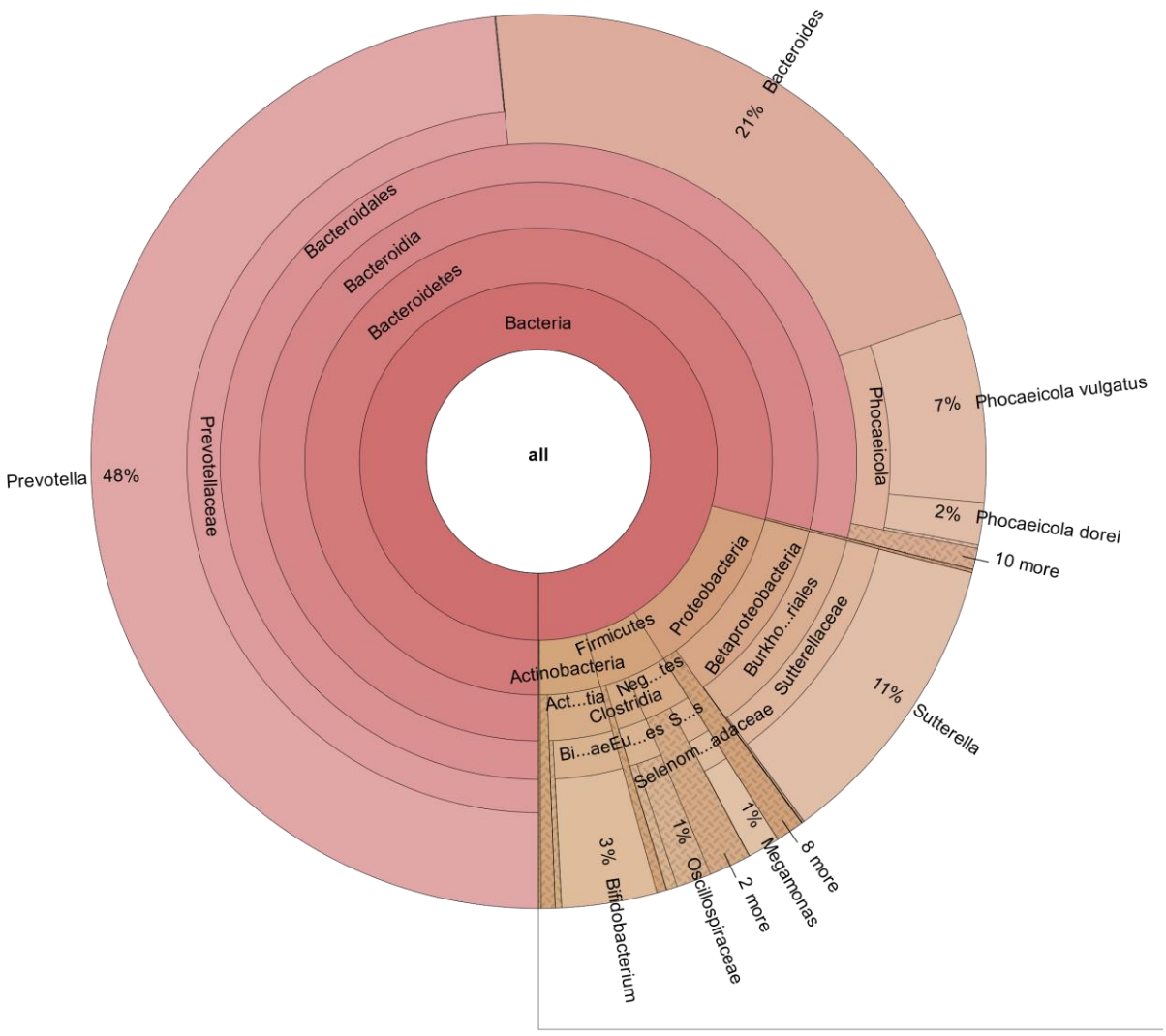


Figure S2: Krona Plot showing the distribution of bacterial species in the gut of HC2

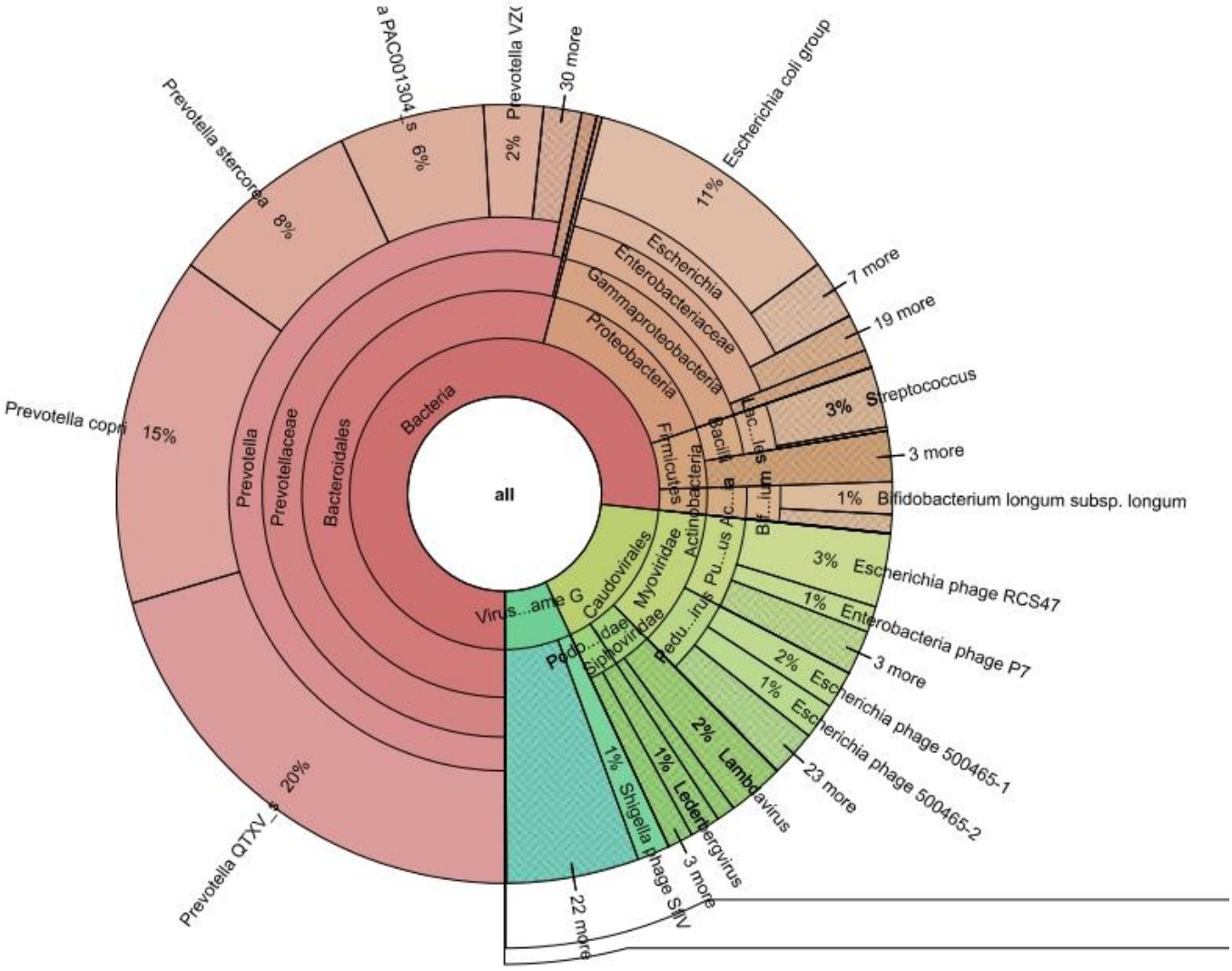


Figure S3: Krona Plot showing the distribution of bacterial species in the gut of IBS9

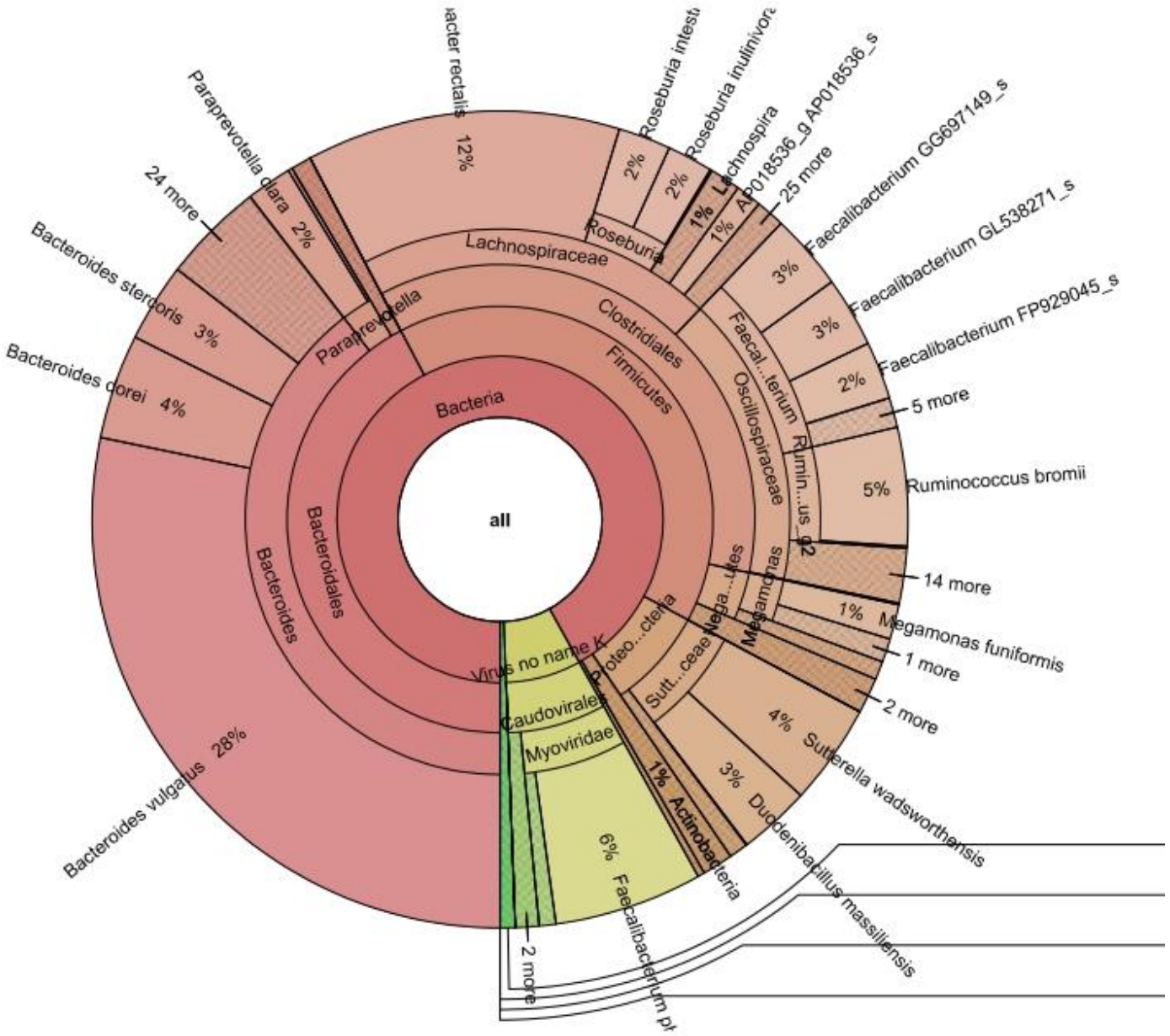


Figure S4: Krona Plot showing the distribution of bacterial species in the gut of IBS5