Original Article

Integrative Metagenomic Analysis Reveals Human Gut Microbiome-Derived Candidate Non-invasive Biomarkers for Type II Diabetes Mellitus

Wajeeha Mujahid, MSc; Sahar Safdar, MSc; Ayesha Aftab, MSc; Sania Tabassum, MPhil; Syeda Marriam Bakhtiar, PhD

Department of Bioinformatics and Biosciences, Capital University of Science and Technology, Islamabad, Pakistan

Correspondence:
Sahar Safdar, MSc;
Capital University of Science and Technology, Postal code: 44000, Islamabad, Pakistan
Tel: +92 321 9801645
Email: Saharsafdarss26@gmail.com
Received: 09 August 2025
Revised: 30 January 2026
Accepted: 27 February 2026

What's Known

- Type II Diabetes Mellitus (T2DM) is strongly associated with alterations in the gut microbiome, which influence host metabolic processes, immune responses, and insulin sensitivity.
- Dysbiosis of gut microbial communities has been implicated in disease progression, highlighting the microbiome as a critical factor in T2DM pathophysiology.

What's New

- This study identified disease-specific gut microbial biomarkers associated with T2DM using an integrated subtractive metagenomics framework combining assembly-based and read-based profiling approaches.
- Key microbial taxa, including *Bacteroides dorei*, *Faecalibacterium prausnitzii,* and *Bacteroides vulgatus,* demonstrated distinct abundance patterns in T2DM and may serve as potential non-invasive diagnostic biomarkers.

Abstract

Background: Type II Diabetes Mellitus (T2DM) is increasingly associated with alterations in the gut microbiome, which influences host metabolism, inflammation, and insulin sensitivity. Metagenomic profiling has emerged as a promising non-invasive strategy for identifying disease-associated microbial signatures. However, distinguishing disease-specific biomarkers from general dysbiosis remains a major challenge. This study aimed to develop an integrative subtractive metagenomic framework to identify candidate disease-specific gut microbial biomarkers.

Methods: This *in silico* case-control study used publicly available metagenomics datasets from healthy controls and individuals with T2DM. Assembly-based and read-based taxonomic profiling approaches were integrated. Differential abundance analysis using the Wilcoxon rank-sum test identified key microbial taxa significantly associated with T2DM.

Results: Potential microbial biomarkers were identified as *Bacteroides dorei, Bacteroides gracilis, Bacteroides stercoris, Bacteroides ovatus, Bacteroides thetaiotaomicron, Bacteroides uniformis, Bacteroides vulgatus, Bacteroides xylanisolvens, Eggerthella lenta, Escherichia coli, Faecalibacterium prausnitzii, Parabacteroides distasonis, Ruminococcus torques,* and *Subdoligranulum.* These taxa are involved in gut metabolic homeostasis and may serve as candidate non-invasive biomarkers for T2DM.

Conclusion: The results of this study advance understanding of microbiome-disease crosstalk and form the basis for further *in vitro* and *in vivo* validation and microbiome-targeted therapeutic approaches. Overall, this integrative metagenomic study supports alteration of microbial ecology in T2DM, validating the use of gut microbiome profiling as a diagnostic and therapeutic tool in metabolic disease research.

Please cite this article as: Mujahid W, Safdar S, Aftab A, Tabassum S, Bakhtiar SM. Integrative Metagenomic Analysis Reveals Human Gut Microbiome-Derived Candidate Non-invasive Biomarkers for Type II Diabetes Mellitus. Iran J Med Sci. doi: 10.30476/ijms.2026.108150.4295.

Keywords ● Metagenomics ● Type II diabetes mellitus ● Non-invasive biomarkers ● Microbiome ● Therapeutics

Introduction

Diabetes mellitus is one of the oldest recognized metabolic disorders and was first documented in Egyptian literature approximately 3,000 years ago. It is a chronic condition

characterized by persistent hyperglycemia resulting from impaired insulin secretion, defective insulin action, or both.[1,2] Approximately 90% of instances of diabetes are type 2 diabetes, which is caused by a combination of decreased insulin production and insulin resistance.[3] According to reports published in 2019, diabetes is the fifth leading cause of mortality worldwide, with a high prevalence.[2] The rapidly rising incidence of type 2 diabetes is attributed to several factors, including the aging of the population, the rising incidence of obesity, and the fall in physical activity levels that comes with industrialization. Risk factors for Type II Diabetes Mellitus (T2DM) include being older, obese, having prior gestational diabetes, having a family history of the condition, having low glucose tolerance, being inactive, and being a certain race or ethnicity.[4]

Nowadays, oral hypoglycemic medications are the primary treatment for diabetes. Twelve different types of medications, including sulfonylureas, meglitinides, biguanides, α-glucosidase inhibitors, and dipeptidyl peptidase (DPP-4) inhibitors, have been approved by the Food and Drug Administration (FDA) to help patients control their blood glucose levels. However, most people with type 2 diabetes cannot have their blood glucose levels adequately controlled for an extended period of time with the current medications, and they will experience some adverse effects.[5] As a result, new medications and novel therapeutic targets are being extensively investigated. The makeup of intestinal microbiota in humans and animals has recently attracted considerable attention in research on the mechanisms behind type 2 diabetes and its treatment.[6] Numerous metabolic disorders, including diabetes and inflammatory bowel disease, have been linked to gut microbiota.[7]

It is now acknowledged that the gut microbiota is a separate organ that controls numerous physiological processes and influences a wide range of host activities.[8] These microbial communities can be perturbed by external factors such as antibiotic use, age, diet, pregnancy, and stress, among others, and undergo a change in either their composition or abundance, which is known as dysbiosis.[9] The state of dysbiosis can increase susceptibility to various diseases and presentation of different phenotypes, including neurological, immunological, physiological, psychic, and metabolic disorders. One such metabolic disorder caused by dysbiosis of the human gut microbiome is T2DM.[10] Previous studies have identified biomarkers that are not specific to T2DM. With the advent of Next Generation Sequencing (NGS) technology, there has been a rapid surge in metagenomic studies, substantially advancing our knowledge of previously unculturable microorganisms and their association with prevalent diseases and ecological changes.[11] The conventional approach adopted to discover disease-specific biomarkers from metagenomic data often involves a differential method.[10] However, in many cases, biomarkers identified using this differential approach reflect a mix of disease-specific characteristics and the features of general dysbiosis.

One of the most effective and widely applicable methods for integrating genomic and molecular data into clinical practice is biomarker discovery. Human microbial communities can function as biomarkers of host characteristics such as lifestyle and illness, according to recent metagenomic analyses. These results have also attracted attention to the importance of tasks such as class prediction, which determines the subtype of a new sample, and class discovery, which finds new subtypes of a disease. These findings are supported by comparisons between healthy and diseased samples.[12]

A critical evaluation of such studies reflects an immediate need to improve current approaches adopted to evaluate the predicted diagnostic biomarkers so that general dysbiosis features can be separated from disease-specific biomarkers. By acknowledging the existing limitations of conventional study designs that often lead to the false discovery of disease-specific biomarkers and re-discovering the diagnostic biomarkers for critical diseases such as diabetes, we hypothesize that integrating assembly-based and read-based profiling within a subtractive metagenomic framework can improve the identification of robust, disease-associated microbial candidates across heterogeneous datasets. Therefore, this study aimed to identify candidate non-invasive gut microbial biomarkers associated with T2DM.

Materials and Methods

This *in silico* case-control study was conducted in 2024 at Capital University of Science and Technology, Islamabad, Pakistan. The integrated subtractive metagenomics pipeline combining read-based and assembly-based approaches for the identification of disease-specific gut microbial biomarkers in T2DM is adopted.

Figure 1 provides a comprehensive overview of the methodology adopted for the metagenomic analysis of the human gut microbiome as a tool towards noninvasive biomarkers for diagnosing T2DM.

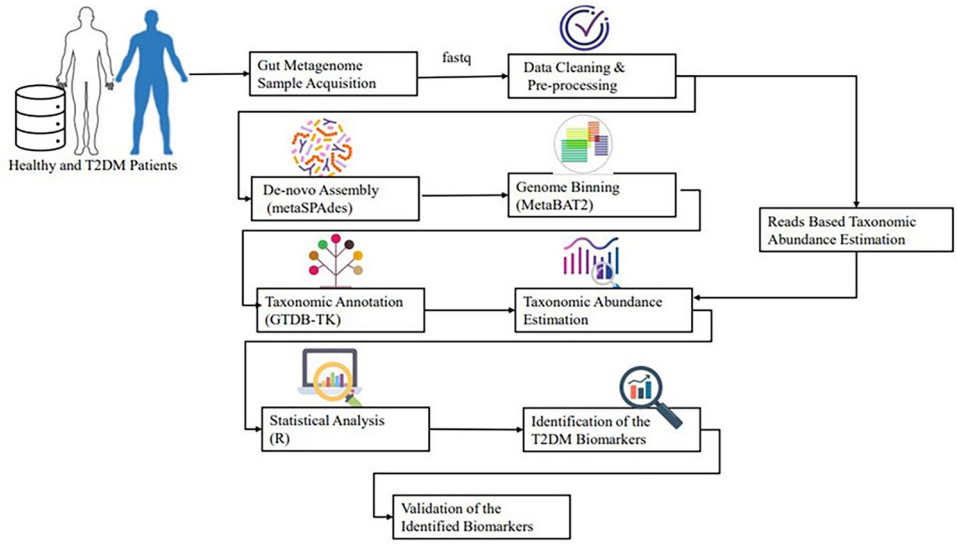

Figure 1: This figure presents the integrated workflow used for human gut metagenomic analysis to identify candidate non-invasive biomarkers for type II diabetes mellitus.

Human Gut Metagenome Sample Retrieval

This study employs publicly available data obtained from T2DM patients and healthy controls. Compressed Fastq files of paired-end raw reads were downloaded from the National Center of Biotechnology Information (NCBI) Sequence Read Archive (National Center for Biotechnology Information, USA). The dataset comprised 50 healthy and 50 T2DM patients' metagenomes from different cohorts available under BioProject accession numbers PRJNA422434 and PRJEB11419, respectively. Sample size was determined based on the availability of high-quality publicly deposited metagenomic datasets with comparable sequencing depth. An equal number of control and disease samples were selected to minimize group imbalance in exploratory analyses.

As datasets originated from independent BioProjects, potential batch effects related to sample collection and sequencing protocols cannot be fully excluded. However, the study design aimed to identify microbial signatures consistently observed across heterogeneous datasets.

FastQC Analysis and Data Pre-processing

Following the retrieval of raw FASTQ files of gut metagenomes, selected datasets were subjected to quality assessment and preprocessing using a combination of FastQC v0.12.1 (Babraham Bioinformatics, UK) and Fastp tools v0.23.2 (OpenGene, China).[13] Quality evaluation and data pre-processing were carried out in a Linux environment (Linus Torvalds/Open-Source Community, USA) using command-line programs. Subsequently, Trimmomatic (version 0.39; Usadel Lab, Max Planck Institute for Developmental Biology, Tübingen, Germany) was used to trim raw reads for quality, read length, and adaptor content. To investigate read sequences for anomalies, the over-representation of k-mers (a sign of contamination or genomic repeats that can lead to complex metagenomic assemblies), and to facilitate the removal of sequencing adapter sequences, these tools make use of both the quality information provided by sequencing instruments and databases that contain collections of adapters and primer sequences. They also evaluate the quantity or percentage of unclear bases, guanine-to-cytosine (GC) content, read length, and quality scores. Post-trimming reads with a length below a specified threshold are eliminated.[14]

De novo Metagenome Assembly and Genome Binning

Based on the number of reads and the complexity of the microbial species in the sample, MAGs are assembled from short reads either as a complete or a draft genome. *De novo* assembly is based on either the overlap-layout-consensus (OLC) or De Bruijn graph (DBG) approaches, where the DBG is the most prevalent method.[15] Hence, this study has also used the DBG-based assembler, including metaSPAdes v3.15.3 (St. Petersburg State University, Russia), which uses k-mers for assembling the selected metagenomes.[16] The resulting contigs were binned using MetaBAT2 (version 1.7; Lawrence Berkeley National Laboratory, USA) at default settings and a minimum contig length of 2500 bp. The combination of metaSPAdes and metaBAT2 is quite successful in recovering low-abundance species.[17]

Taxonomic Annotation and Abundance Estimation

An assembly-based method was used to assess the microbial communities and estimate their relative abundances. High-quality reads were subjected to taxonomic profiling using Kaiju (version 1.9.0; Max Planck Institute for Informatics, Germany), and Taxonomic abundances were estimated at the genus and species level in Greedy mode.[18] The reference database used was RefSeq Genomes, and the low-abundance filter was set to ≥0.05. Taxonomic abundance profiles were generated at both the genus and species levels. Genome Taxonomy Database–ToolKit (version 2.3.2; GTDB-Tk; University of Queensland, Australia) was used to annotate the bacterial genomes according to their taxonomy. A recovery threshold value of 0.05 was used to include the low-abundance species along with the most prevalent and highly abundant taxa.[19] Later, the results from the two approaches were combined to get the final taxonomic profiles of healthy and T2DM patients.

Biomarker Identification

The Wilcoxon rank-sum test was performed in R (R Foundation for Statistical Computing, Austria) to assess significant differences between the control and diseased gut metagenomic profiles at the genus and species levels. Differentially abundant taxa between the control and T2DM metagenomes were selected as potential microbial biomarkers. To compare microbial divergences between individuals with T2DM and healthy controls, we performed a comparative approach based on species-level relative abundance data. The datasets of the T2DM and control groups were then merged using taxonomic identifiers to facilitate comparison. Missing values were imputed with a minimal abundance constant (1e-6) to allow for abundance logarithmic transformations and to prevent division by zero. Alpha diversity was analyzed using three ecological metrics of the system, which are the number of observed genera (species richness), the Shannon diversity index, and the Simpson index. These indices can provide information on species count as well as species evenness of distribution within each group. The Shannon index was taken as the negative of the sum of the proportion of each genus multiplied by its natural logarithm. In contrast, the Simpson index was computed as the sum of squared proportions, thereby characterizing the community's dominance.[20] To investigate the beta diversity and recapitulate compositional separation between the groups, we ran Principal Component Analysis (PCA).[21] Synthetic sample replicates (n=5 per group) were generated from the observed genus-level abundance profiles by adding Gaussian noise to simulate inter-individual biological variation. The derived prevalence data were then log-transformed, normalized as z scores, and run on PCA. The group clustering was plotted with the first two principal components (PC1 and PC2). Each sample was linked to its corresponding group centroid, and intra-group dispersion (standard deviation ellipses) was added. For differential abundance analysis, for each genus, log2 fold change values were computed using the ratio of the T2DM and control abundances. Genera were enriched (log□FC>1), depleted (log□FC<-1), or unchanged (log□FC between -1 and 1) in T2DM. The cumulative relative abundance of each category was then worked out for each group and presented in a stacked bar chart. This revealed patterns of enrichment and depletion for a disease status-associated change in the T2DM microbiome (broad taxonomic shifts).

Results

High-quality metagenomic sequencing data were obtained following preprocessing with an average read length of 107 bp and over 96% of bases exceeding a Q-score of 20. Differential abundance analysis revealed significant alterations in gut microbial composition between patients and healthy controls, highlighting distinct taxonomic shifts associated with disease status. The basic statistics obtained from FastQC analysis demonstrated that total sequences were 19998693, with total bases of 2.9 Gbp encoded through Sanger/Illumina 1.9. No poor-quality sequence was flagged. The detected sequence length ranged from 35 to 151. Quality reports of raw sequencing data were obtained using FastQC. The generated reports illustrated that some of the reads of the input samples contained adapter contamination, abnormal G+C content, and per-base sequence content. Moreover, some samples showed reads with phred scores less than 20 that needed to be removed. This cleaning was conducted in a Linux environment and produced reads with a mean length of 107 bp and 96% of bases having Q-scores >20. During filtering, reads that passed filters were 16.034314 M (95.647623%), reads with low quality were 695.022000 K (4.145934%), reads with too many Ns were 14.740000 K (0.087927%), and reads that are too short were 19.868000 K (0.118516%).

De novo Assembly and Recovery of MAGs

De novo assembly of human gut metagenomic samples performed using metaSPAdes had a mean assembly length of 78477.50±29765 Kbp for healthy controls and 79533±2500 Kbp for T2DM. The larger assembly size observed in the T2DM group may reflect increased microbial diversity of the gut microbiome in diabetics. However, this difference is not significant as determined by the Wilcoxon rank-sum test ($P>0.05$). Subjecting these assemblies to genome binning yielded 120 and 170 MAGs for healthy and T2DM patients. These MAGs were subjected to quality assessment using CheckM and classification into high-quality (HQ), medium-quality (MQ), and low-quality (LQ) MAGs using the MIMAG criteria, as shown in table 1.

The taxonomic profiles of the control group at the genus level show lower diversity in microbial communities than the T2DM group. Analysis of the resulting taxonomic profiles revealed the highest proportions of *Bacteroides* and *Phocaeicola* (27.2 % and 17.2%, respectively) in the healthy-control metagenomes. Among 142 identified genera, 10 were present at high abundance >1%, and the remaining 132 were present at a relative abundance of <1%. *Acinetobacter*, *Desulfitobacterium*, *Caproiciproducens*, and *Niabella* are among the low-abundance genera, present at 0.001%. At the species level, *Prevotella copri*, *Faecalibacterium prausnitzii*, *Phocaeicola vulgatus*, *Phocaeicola dorei*, *Bacteroides uniformis*, *Bacteroides thetaiotaomicron*, *Bacteroides stercoris*, *Lachnospira eligens*, and *Bacteroides fragilis* were found at relatively high abundance, i.e., 6.23%-1.43%. Whereas *Caproiciproducens* sp. NJN-50, *Streptococcus australis*, and *Streptococcus gallolyticus* were found at low abundance, i.e., 0.001% among others.

In the gut metagenome profiles of T2DM patients, 135 genera and 274 species were identified. Compared with the control group, the gut microbiome of diabetic patients was less diverse at the genus level and more diverse at the species level. This discrepancy may be due to the loss of beneficial genera and acquisition of pathogens or opportunistic pathogens involved in the onset of T2DM.

At the genus level, a relatively higher proportion (26%-0.89%) of *Bacteroides*, *Phocaeicola*, *Prevotella*, *Faecalibacterium*, *Bifidobacterium,* *Lachnospira*, *Alistipes*, *Blautia*, *Phascolarctobacterium*, *Roseburia*, and *Mediterraneibacter* was observed. Whereas *Vibrio*, *Lactobacillus*, *Pseudoalteromonas*, *Crassaminicella*, *Sphaerochaeta*, *Aeromonas*, *Ndongobacter*, *Alkaliphilus*, and *Mageeibacillus* were found to be present at low abundance, i.e., 0.001%. A slightly lower proportion of *Bacteroides* may indicate diabetes, as a higher abundance of this genus is positively correlated with improved insulin resistance in a mouse model. But this genus shows a contradictory association with T2DM. A lower proportion of *Ruminococcus*, *Fusobacterium*, and *Blautia* found in T2DM gut metagenomic profiles than in the healthy controls is positively associated with T2DM.[22] The *Faecalibacterium* genus has also been shown to be reduced in diabetic patients in response to the various antidiabetic medications.[15] Additionally, a higher abundance of *Bifidobacterium* than the control may suggest a protective effect in response to immune system activation.[23] The taxonomic profiles of diabetic patients at the species level contained more species than the control group. The microbial communities constituting the major proportion of the taxonomic profiles were *F. prausnitzii, P. copri, Phocaeicola vulgatus, B. stercoris, Lachnospira eligens,* and *Phocaeicola dorei.* These species' abundance ranged from 4.1% to 1.5% in the metagenomic samples.

Compared with healthy controls, increased relative abundances (approximately twofold) of *Ruminococcus* gnavus, *Ruminococcus bicirculans*, *Ruminococcus bovis*, and *Ruminococcus albus* were observed. In contrast, lower abundances of *F. prausnitzii*, *Subdoligranulum*, *Roseburia intestinalis*, *Roseburia inulinivorans*, *Eubacterium rectale*, and *Akkermansia muciniphila* were identified in T2DM metagenomes. These findings indicate species-specific variation within the genus *Ruminococcus*, despite differences observed at the overall genus level.

Table 1: The quality of MAGs recovered from the human gut metagenome of healthy and type II diabetes mellitus patients

Condition	Healthy	Type II diabetes mellitus
MAG Count	1591	1721
HQ%	35.4	35.7
MQ%	45.6	55.0
LQ%	19.0	9.3

MAG: Metagenome-assembled genome; HQ: High quality; MQ: Medium quality; LQ: Low quality

Microbial Biomarkers of T2DM

The gut microbiota modulates inflammation, interacts with dietary constituents, influences the gut barrier, glucose and lipid metabolism, insulin sensitivity, and maintains homeostasis in the human body. To identify genera with different abundance between T2DM and healthy subjects, a Wilcoxon rank-sum test was conducted. The binary heatmap shows genera distinctly enriched in either group ($P<0.05$). Enriched genera in the T2DM group are marked red, and enriched in controls are marked in green. The distribution presents a clear microbial signature, as some genera are consistently more abundant in one group versus the other, indicating potential disease-associated microbial changes, as shown in figure 2.

An integrated view of differential microbial abundance in T2DM patients and healthy controls is shown in figure 3. The upper panel shows a bar plot of log2 fold changes at the genus level; positive values are assigned to genera enriched in T2DM, whereas negative values are assigned to depleted (compared to diabetic) ones (figure 3A). This visualization emphasizes significant changes in microbial taxa; e.g., *P. copri* and *Bacteroides ovatus* are strongly depleted in T2DM, whereas other genera are strikingly enriched. The central heatmap demonstrates differential abundance (absolute) of groups with a distinct trend for increased or decreased representation of key genera across conditions (figure 3B). The rightmost clustered heatmap, according to hierarchical clustering, visually confirms group-specific microbial signatures as well as taxonomic relationships (figure 3C). Together, these analyses consistently demonstrate a systematic alteration of the gut microbiota in T2DM subjects, suggesting its potential role in disease onset and progression.

Alpha diversity analysis showed some decrease in microbial richness and diversity in patients with T2DM compared to healthy controls, as shown in figure 4 (A). A lower observed number of genera was found in the T2DM group, implying a reduction in taxonomic richness. Likewise, the Shannon and Simpson diversity indices were reduced in the T2DM group, suggesting not only decreased richness but also variation in community evenness. These results support the hypothesis that links a disrupted gut microbiome profile in T2DM to reduced diversity. When using the difference abundance analysis that relied on log 2-fold changes, there were different clusters of T2DM-enriched and T2DM-depleted genera, as shown in figure 4 (B). Genes with a fold change >2 (log□FC>1) were considered enriched in T2DM, and those with <0.5 (log□FC<-1) were considered depleted. The total relative abundance of T2DM-enriched taxa was significantly higher in the diabetic group than that of the T2DM-depleted taxa within the control group's microbiome. The "Other" category, where taxa show no significant change values, was fairly stable among groups. These results underscore the occurrence of particular microbial signatures that may be associated with disease pathogenesis or host metabolic modulation in T2DM. A separation between the groups of microbial profiles in the T2DM and control groups was evident from the PCA, as shown in figure 4 (C). There was a substantial proportion of the variance across samples, which was explained by PC1 and PC2 jointly. Single samples simulated to represent inter-individual variation displayed their own distinct clusters by group, with T2DM samples separated from controls. Centroid connections and dispersion ellipses accentuated the difference in microbial composition across groups.

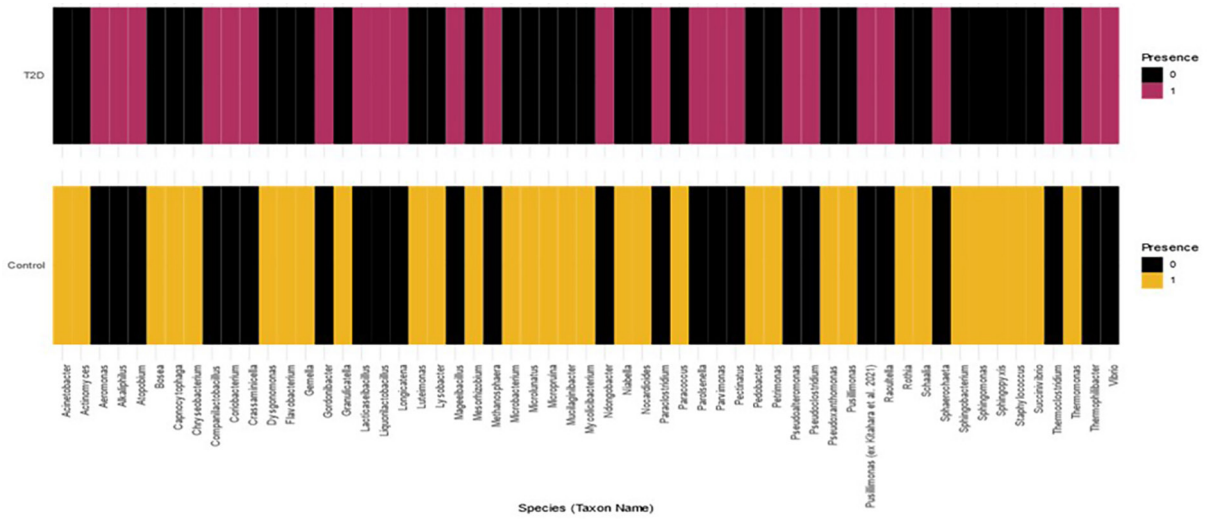

Figure 2: This figure presents candidate biomarkers found exclusively in control and T2DM gut metagenomes.

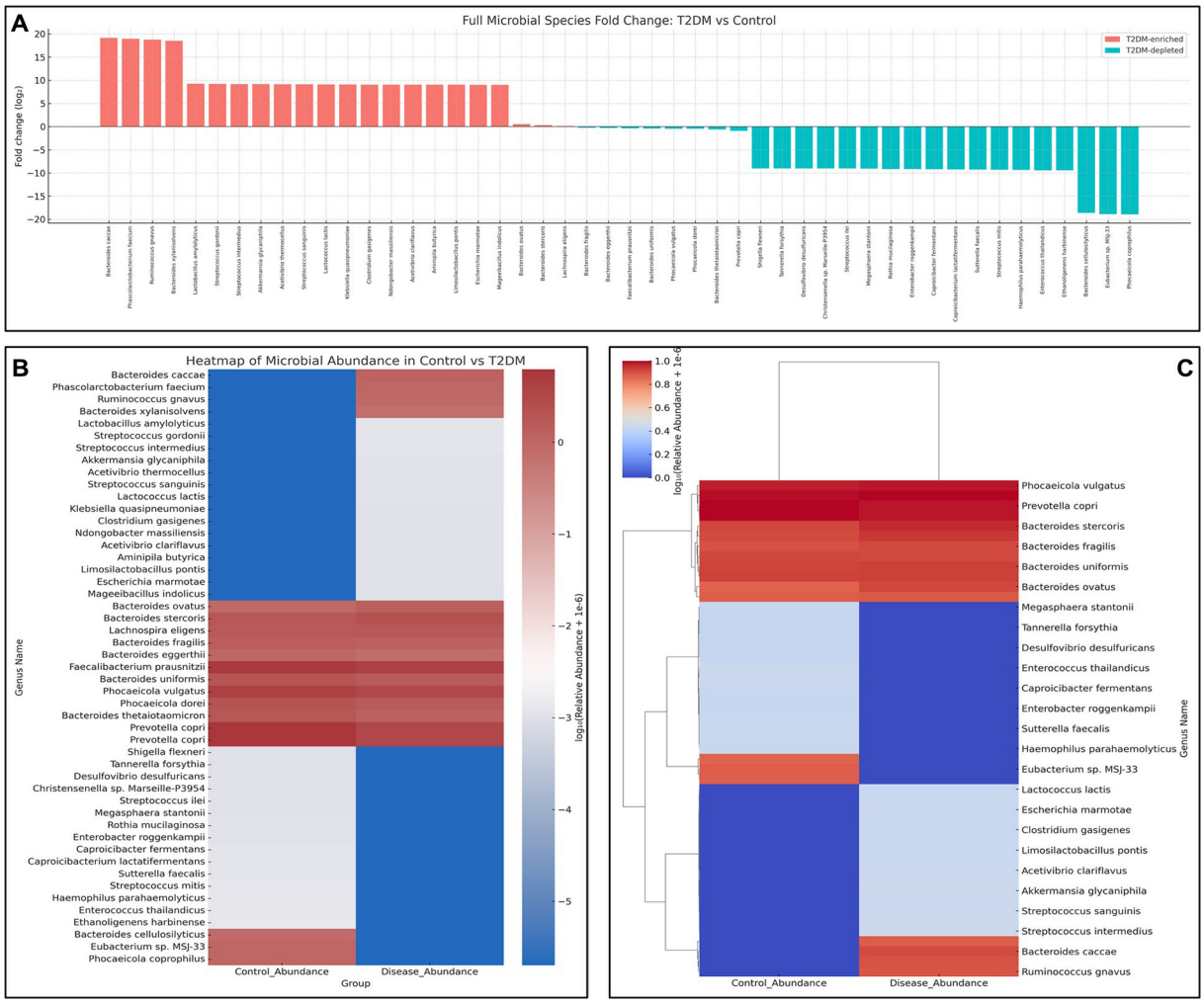

Figure 3: This figure illustrates differential microbial profiling in type 2 diabetes mellitus patients and the control group. (A) The top bar plot represents the log□ fold changes in abundances at the genus level, defining type 2 diabetes mellitus patient-enriched species (positive values, red bars) and type 2 diabetes mellitus patient-depleted species (negative values, blue bars). In genera where shifts in abundance are most pronounced, specific dysbiosis patterns in type 2 diabetes mellitus patients are revealed. (B) The left heatmap shows the abundance of selected genera within groups in a two-column format, consistent depletion or enrichment profiles in type 2 diabetes mellitus patients versus controls. (C) The right clustered heatmap proceeds to classify the genera according to an order of similarity, showing group-specific clustering and microbial composition shifts. These visualizations together highlight a distinct separation in gut microbial ecology related to the pathophysiology of type 2 diabetes mellitus.

This dissociation further supports the notion that T2DM is associated with persistent alterations in gut microbial composition, consistent with the dysbiotic changes characteristic of the disease.

Discussion

This study applied an integrative subtractive metagenomic approach to identify candidate gut microbial taxa associated with T2DM across heterogeneous cohorts. The resulting comprehensive pool of biomarkers was further associated with healthy and T2DM states, suggesting their use as non-invasive biomarkers. Overall, the results demonstrate clear compositional and functional differences in healthy and patient groups. Comparative abundance analysis revealed a distinct microbial signature in T2DM, characterized by enrichment of opportunistic and metabolically disruptive taxa and decreased beneficial microbial species. Furthermore, the PCA identified a significant variation in microbial community, showing a distinct difference between the patient and control groups. Diversity metrics showed subtle but consistent differences. This suggests disease-associated dysbiosis is driven more by compositional shifts rather than a complete loss of microbial diversity. The increased proportion of enriched species in patients compared with controls is clearly visualized in taxonomic characterization, indicating the presence of disease-specific microbial patterns. This hybrid approach is key to accurately identifying

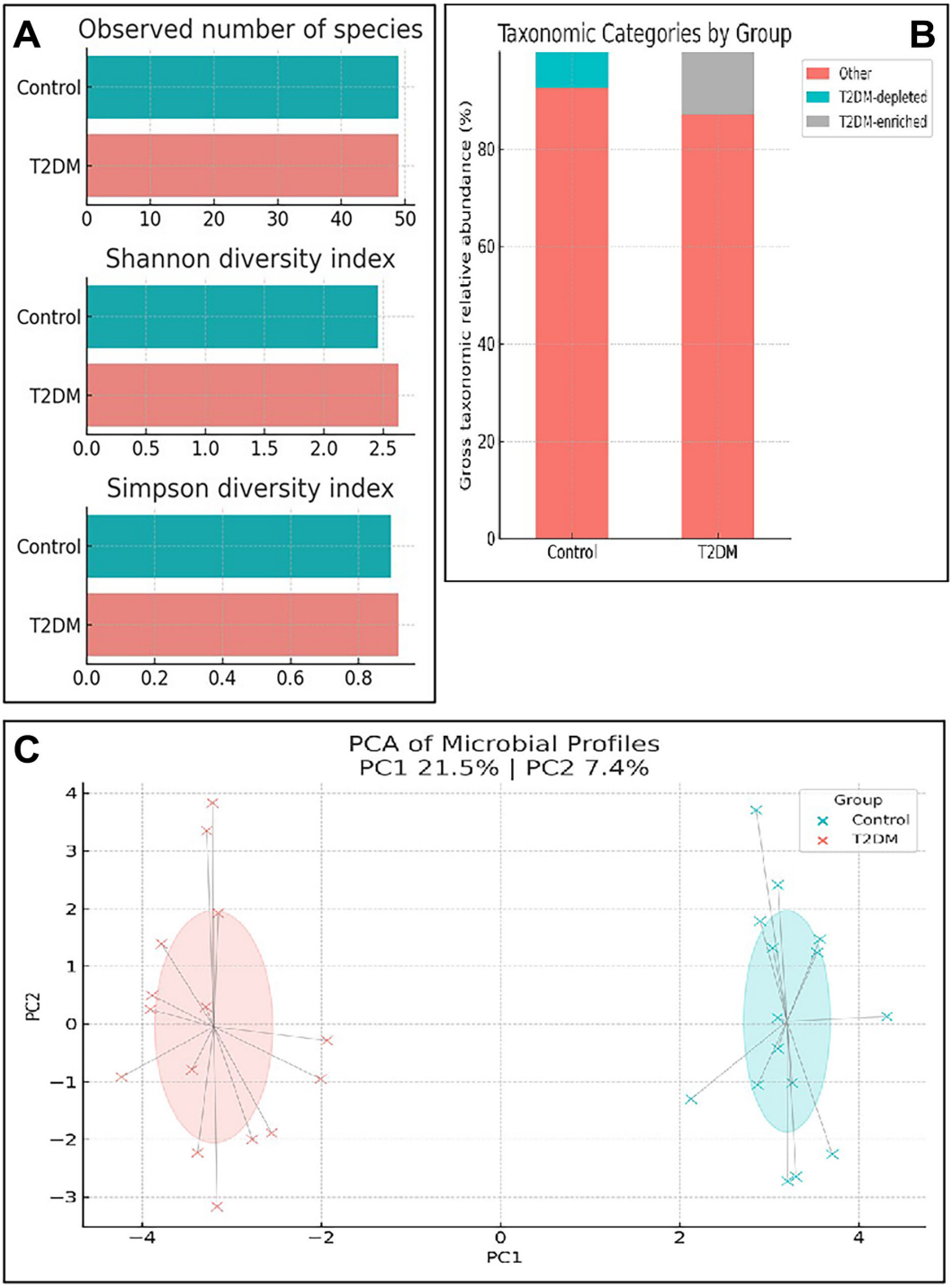

Figure 4: This figure presents differential microbial profiling in type 2 diabetes mellitus patients and the control group. (A) The top bar plot represents the log fold changes in abundances at the genus level, defining type 2 diabetes mellitus patients enriched species (positive values, red bars) and type 2 diabetes mellitus-depleted species (negative values, blue bars). In the genera with the most pronounced shifts in abundance, specific dysbiotic patterns in type 2 diabetes mellitus are revealed. (B) The left heatmap shows the abundance of selected genera within groups with a two-column format, consistent depletion or enrichment profiles, and type 2 diabetes mellitus patients compared with controls. (C) The right clustered heatmap proceeds to classify the genera according to an order of similarity, which shows group-specific clustering and microbial composition shifts. All these visualizations combined highlight a distinct spatial distinction in gut microbial ecology related to the pathophysiology of type 2 diabetes mellitus patients.

microbes and candidate biomarkers that not only correlate with T2DM but also shed light on metabolic processes, inflammation, and insulin resistance. This could transform T2DM diagnosis and management.[24]

Integrating non-invasive biomarkers into T2DM management provides significant advantages over traditional invasive methods. One major advantage is reduced patient discomfort, since using fecal samples eliminates

the need for repeated skin punctures. These biomarkers enable early identification and constant follow-up of the disease, which in turn helps prevent late diagnoses and the complications that usually follow through the provision of real-time information on blood glucose levels. This non-invasive diagnostic approach is less burdensome for patients and may improve participation among individuals who are usually reluctant to engage in regular health monitoring.[25] Less invasive practices reduce the risks of such complications as infection and scarring, hence making it appropriate for patients who have slow wound recovery rates. The convenience of these tests promotes patient engagement, which is essential for effective long-term disease management.[26] Moreover, non-invasive biomarkers allow personalized treatment approaches because they reflect how patients react to the modification of therapeutic interventions.[27] This will also help scientists to investigate the gut microbiota's contribution to T2DM and determine interesting new targets for therapy.

The findings from the integrated approach underscore that different populations of gut bacteria are altered in individuals with T2DM and that these changes significantly impact glucose metabolism and overall metabolic health.[27, 28] Higher abundances of *B. dorei* and *B. gracilis* have been associated with improved glycemic control, suggesting that increasing the abundance of these species may contribute to better blood glucose regulation in individuals with T2DM. However, *Bacteroides vulgatus* is less abundant in individuals with T2DM, supporting its potential utility as a candidate diagnostic biomarker, although its abundance may also be influenced by factors such as dietary habits and antibiotic exposure. *E. coli* also functions as a diagnostic biomarker for T2DM and is associated with high-fat intake and increased gut permeability, both risk factors for inflammation and metabolic disturbances.[29, 30] Species such as *F. prausnitzii* are not considered direct risk factors for T2DM; however, their abundance is often reduced in affected individuals, potentially reflecting the influence of factors such as unhealthy dietary patterns and antibiotic exposure on the gut microbiota.[31] It is believed that eating too much fat and too little fiber, being sedentary, and being under constant stress can, by tipping the balance, favor the harmful organisms in the gut. Increased gut permeability allows toxic microorganisms to enter the bloodstream and trigger inflammation, thereby aggravating microbiome disorders, making it harder to manage diabetes and promote insulin resistance. It has been noted that certain species of the gut microbiome are more abundant in healthy individuals than in those with T2DM.[32] Major species detected are *F. prausnitzii* (anti-inflammatory effects), *Prevotella copri* (glucose management promoted with fiber-rich foods), *B. stercoris* (fiber fermenter), and *L. eligens* (supports carbohydrate metabolism).[33]

Additionally, recent research has shown a strong correlation between diabetes and the composition of the gut microbiota. When compared to the nondiabetic controls, the diabetic group's proportions of *Firmicutes* and *Clostridium* were found to be considerably lower. In addition to metabolic microbiota-host crosstalk, increased intestinal permeability leading to raised systemic levels of lipopolysaccharides that promote insulin resistance and hyperinsulinemia are other possible causes for the function of gut microbiota in diabetes.[34] Moreover, in a mouse model, levels of *Bifidobacterium* were substantially and favorably associated with enhanced glucose tolerance and reduced inflammation. A study by Brown suggested *Prevotella*'s relative absence indicates a deficiency of mucin on the intestinal epithelial layer and could be a sign of present or future gut permeability. Additionally, patients have a significantly higher population of bacteria such as *Bacteroides* than controls. Propionate, acetate, and succinate are the products of these bacteria that ferment glucose and lactate. These short-chain fatty acids do not cause the formation of mucin, in contrast to butyrate.

This connection between T2DM and gut microbiome has fueled a rapid wave of scientific research. In an effort to find microbial signatures that could serve as non-invasive T2DM complication biomarkers, the current study was undertaken. By performing a differential analysis method of metagenome profiles from healthy and control patients, it was possible to find potential biomarkers of T2DM. These candidate biomarkers can be experimentally and practically examined. The research by Zhang and others provided perspectives on the function of gut bacteria in the onset and progression of diabetes as well as innovative precision-medicine strategies to combat type 2 diabetes. Additionally, this study demonstrated that examining metabolic pathways and functions is more significant for understanding the role of gut microbiota in disease etiology than merely comparing microbial populations.[35] Wu and others reported a substantial nonlinear negative association between DI-GM and diabetes risk. The risk of diabetes may be considerably decreased by keeping the DI-GM

score above 6.191. Increasing the production of short-chain fatty acids (SCFAs) and preserving the gut microbiota diversity may help lower the risk of diabetes. To elucidate the mechanisms behind the nonlinear association between DI-GM and diabetes risk, more research is required.[36] There is growing evidence that the homeostatic regulation of glucose metabolism is significantly influenced by the gut microbiota.[37] Low-grade endotoxemia due to increased gut permeability, imbalanced production of SCFAs and branched-chain amino acids, and disrupted bile acid metabolism are some of the ways that dysbiosis of the gut microbiota contributes to the development of glucose intolerance and insulin resistance in T2DM.[38]

The alteration in gut microbiome composition in patients with T2DM demonstrating consistent pro-inflammatory taxa are enriched in individuals with diabetes, while a significant reduction is observed in SCFA-producing bacteria. This is due to gut permeability, chronic inflammation, and impaired glucose metabolism.[30] There was a positive and statistically significant association of the *Bacteroidetes*-to-*Firmicutes* ratio in individuals with T2DM with plasma glucose levels. However, the relationship did not change with body weight, meaning that the connection between this ratio and decreased tolerance to glucose was not disrupted. The identified candidate species biomarkers include *B. dorei, B. gracilis, B. stercoris, B. ovatus, B. thetaiotaomicron, B. uniformis, B. vulgatus, Bacteroides xylanisolvens, Escherichia lenta, E. coli, F. prausnitzii, Lachnospiraceae, Parabacteroides distasonis, Ruminococcus torques,* and *Subdoligranulum.* These species play a vital role in maintaining a balanced gut microbiome, which is crucial for metabolic health and the prevention of diseases such as T2DM.[23] This research will improve our understanding of how the microbiome influences diabetes and guide targeted treatments.

The limitations of this study arise from the complexity of host-microbiome interactions. Although this was an integrative study, the study should include the functional impact of host genetic variability shaping the microbiome composition in the disease state. Furthermore, the analysis focused on microbial abundance rather than microbial activity, and no predictive validation was performed. Therefore, identified taxa should be considered candidate biomarkers requiring further experimental and clinical validation.

Conclusion

This study identified disease-specific gut microbial candidate biomarkers associated with T2DM using an integrated subtractive metagenomics approach. Notably, *B. dorei* and *F. prausnitzii* were identified as potential non-invasive diagnostic markers. The findings demonstrate that combining assembly-based and read-based analyses improves the robustness of biomarker identification. Our integrated approach provides more reliable and comparable results, in turn reducing the influence of external factors and enhancing the value of our work for a wider population base. Our findings reveal microbial patterns associated with early disease stage manifestations that pave the way for effective, non-invasive practices for the detection and treatment of T2DM.

Authors' Contribution

W.M. contributed to the conception and design of the study, methodology development, data analysis, and interpretation of results. S.S. contributed to data analysis, interpretation of results, drafting, proofreading, and critical revisions. A.A. and S.T. contributed to manuscript drafting, proofreading, and critical revision. S.M.B supervised the study, contributed to study design and interpretation, and critically revised the manuscript. All authors approved the final revision of the manuscript and agree to be accountable for all aspects of the work, including the accuracy and integrity of the study.

Declaration of AI

The authors declare that no AI tools were used in the preparation of this manuscript.

Conflict of Interest: None declared.

References

1 Reed J, Bain S, Kanamarlapudi V. A Review of Current Trends with Type 2 Diabetes Epidemiology, Aetiology, Pathogenesis, Treatments and Future Perspectives. Diabetes Metab Syndr Obes. 2021;14:3567-602. doi: 10.2147/DMSO.S319895. PubMed PMID: 34413662; PubMed Central PMCID: PMC8369920.

2 Sadagopan A, Mahmoud A, Begg M, Tarhuni M, Fotso M, Gonzalez NA, et al. Understanding the Role of the Gut Microbiome in Diabetes and Therapeutics Targeting Leaky Gut: A Systematic Review. Cureus. 2023;15:e41559. doi: 10.7759/cureus.41559. PubMed PMID: 37554593; PubMed Central PMCID: PMC10405753.

3 Berbudi A, Rahmadika N, Tjahjadi AI, Ruslami R. Type 2 Diabetes and its Impact on the Immune System. Curr Diabetes Rev. 2020;16:442-9. doi: 10.2174/1573399815666191024085838. PubMed PMID: 31657690; PubMed Central PMCID: PMC7475801.
4 Alam S, Hasan MK, Neaz S, Hussain N, Hossain MF, Rahman T. Diabetes mellitus: insights from epidemiology, biochemistry, risk factors, diagnosis, complications and comprehensive management. Diabetology. 2021;2:36-50. doi: 0.3390/diabetology2020004.
5 Anitha T, Vallazhath A, Maniyal V, Menon D. Advances in therapeutic strategies for atherosclerosis: from pharmacologics to stents and stent coatings. J Mater Chem B. 2026;14:2407-37. doi: 10.1039/d5tb02293b. PubMed PMID: 41668608.
6 Wu H, Esteve E, Tremaroli V, Khan MT, Caesar R, Manneras-Holm L, et al. Metformin alters the gut microbiome of individuals with treatment-naive type 2 diabetes, contributing to the therapeutic effects of the drug. Nat Med. 2017;23:850-8. doi: 10.1038/nm.4345. PubMed PMID: 28530702.
7 Zhang L, Chu J, Hao W, Zhang J, Li H, Yang C, et al. Gut Microbiota and Type 2 Diabetes Mellitus: Association, Mechanism, and Translational Applications. Mediators Inflamm. 2021;2021:5110276. doi: 10.1155/2021/5110276. PubMed PMID: 34447287; PubMed Central PMCID: PMC8384524.
8 Bodke H, Jogdand S. Role of Probiotics in Human Health. Cureus. 2022;14:e31313. doi: 10.7759/cureus.31313. PubMed PMID: 36514580; PubMed Central PMCID: PMC9733784.
9 Bakir-Gungor B, Bulut O, Jabeer A, Nalbantoglu OU, Yousef M. Discovering Potential Taxonomic Biomarkers of Type 2 Diabetes From Human Gut Microbiota via Different Feature Selection Methods. Front Microbiol. 2021;12:628426. doi: 10.3389/fmicb.2021.628426. PubMed PMID: 34512559; PubMed Central PMCID: PMC8424122.
10 Brockley LJ, Souza VGP, Forder A, Pewarchuk ME, Erkan M, Telkar N, et al. Sequence-Based Platforms for Discovering Biomarkers in Liquid Biopsy of Non-Small-Cell Lung Cancer. Cancers (Basel). 2023;15. doi: 10.3390/cancers15082275. PubMed PMID: 37190212; PubMed Central PMCID: PMC10136462.
11 Wirbel J, Pyl PT, Kartal E, Zych K, Kashani A, Milanese A, et al. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. Nat Med. 2019;25:679-89. doi: 10.1038/s41591-019-0406-6. PubMed PMID: 30936547; PubMed Central PMCID: PMC7984229.
12 Yadav A, Vishwakarma S, Krishna N, Katara P. Integrative omics: current status and future directions. In: Recent Trends in Computational Omics: Concepts and Methodology. New York: Nova Science Publishers; 2020. p. 1-46.
13 Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884-i90. doi: 10.1093/bioinformatics/bty560. PubMed PMID: 30423086; PubMed Central PMCID: PMC6129281.
14 Liu S, Rodriguez JS, Munteanu V, Ronkowski C, Sharma NK, Alser M, et al. Analysis of metagenomic data. Nat Rev Methods Primers. 2025;5. doi: 10.1038/s43586-024-00376-6. PubMed PMID: 40688383; PubMed Central PMCID: PMC12276902.
15 Hauptfeld E, Pappas N, van Iwaarden S, Snoek BL, Aldas-Vargas A, Dutilh BE, et al. Integrating taxonomic signals from MAGs and contigs improves read annotation and taxonomic profiling of metagenomes. Nat Commun. 2024;15:3373. doi: 10.1038/s41467-024-47155-1. PubMed PMID: 38643272; PubMed Central PMCID: PMC11032395.
16 Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 2017;27:824-34. doi: 10.1101/gr.213959.116. PubMed PMID: 28298430; PubMed Central PMCID: PMC5411777.
17 Qayyum H, Talib MS, Ali A, Kayani MUR. Evaluating the potential of assembler-binner combinations in recovering low-abundance and strain-resolved genomes from human metagenomes. Heliyon. 2025;11:e41938. doi: 10.1016/j.heliyon.2025.e41938. PubMed PMID: 39897886; PubMed Central PMCID: PMC11786835.
18 Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun. 2016;7:11257. doi: 10.1038/ncomms11257. PubMed PMID: 27071849; PubMed Central PMCID: PMC4833860.
19 Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics. 2022;38:5315-6. doi: 10.1093/bioinformatics/btac672. PubMed PMID: 36218463; PubMed Central

PMCID: PMC9710552.
20 Kunakh O, Volkova A, Tutova G, Zhukov O. Diversity of diversity indices: Which diversity measure is better? Biosyst Divers. 2023;31:131-46. doi: 10.15421/012314.
21 Calle ML. Statistical Analysis of Metagenomics Data. Genomics Inform. 2019;17:e6. doi: 10.5808/GI.2019.17.1.e6. PubMed PMID: 30929407; PubMed Central PMCID: PMC6459172.
22 Zhou Z, Sun B, Yu D, Zhu C. Gut Microbiota: An Important Player in Type 2 Diabetes Mellitus. Front Cell Infect Microbiol. 2022;12:834485. doi: 10.3389/fcimb.2022.834485. PubMed PMID: 35242721; PubMed Central PMCID: PMC8886906.
23 Que Y, Cao M, He J, Zhang Q, Chen Q, Yan C, et al. Gut Bacterial Characteristics of Patients With Type 2 Diabetes Mellitus and the Application Potential. Front Immunol. 2021;12:722206. doi: 10.3389/fimmu.2021.722206. PubMed PMID: 34484230; PubMed Central PMCID: PMC8415158.
24 Blanco-Miguez A, Beghini F, Cumbo F, McIver LJ, Thompson KN, Zolfo M, et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. Nat Biotechnol. 2023;41:1633-44. doi: 10.1038/s41587-023-01688-w. PubMed PMID: 36823356; PubMed Central PMCID: PMC10635831
25 Zafar H, Saier MH, Jr. Gut Bacteroides species in health and disease. Gut Microbes. 2021;13:1-20. doi: 10.1080/19490976.2020.1848158. PubMed PMID: 33535896; PubMed Central PMCID: PMC7872030.
26 Padilla-Martinez F, Wojciechowska G, Szczerbinski L, Kretowski A. Circulating Nucleic Acid-Based Biomarkers of Type 2 Diabetes. Int J Mol Sci. 2021;23. doi: 10.3390/ijms23010295. PubMed PMID: 35008723; PubMed Central PMCID: PMC8745431.
27 Hills RD, Jr., Pontefract BA, Mishcon HR, Black CA, Sutton SC, Theberge CR. Gut Microbiome: Profound Implications for Diet and Disease. Nutrients. 2019;11. doi: 10.3390/nu11071613. PubMed PMID: 31315227; PubMed Central PMCID: PMC6682904.
28 Cunningham AL, Stephens JW, Harris DA. Gut microbiota influence in type 2 diabetes mellitus (T2DM). Gut Pathog. 2021;13:50. doi: 10.1186/s13099-021-00446-0. PubMed PMID: 34362432; PubMed Central PMCID: PMC8343927.
29 Kim HB, Cho YJ, Choi SS. Metformin increases gut multidrug resistance genes in type 2 diabetes, potentially linked to Escherichia coli. Sci Rep. 2024;14:21480. doi: 10.1038/s41598-024-72467-z. PubMed PMID: 39277620; PubMed Central PMCID: PMC11401871.
30 Rios-Covian D, Salazar N, Gueimonde M, de Los Reyes-Gavilan CG. Shaping the Metabolism of Intestinal Bacteroides Population through Diet to Improve Human Health. Front Microbiol. 2017;8:376. doi: 10.3389/fmicb.2017.00376. PubMed PMID: 28326076; PubMed Central PMCID: PMC5339271.
31 Parsaei M, Sarafraz N, Moaddab SY, Ebrahimzadeh Leylabadlo H. The importance of Faecalibacterium prausnitzii in human health and diseases. New Microbes New Infect. 2021;43:100928. doi: 10.1016/j.nmni.2021.100928. PubMed PMID: 34430035; PubMed Central PMCID: PMC8365382.
32 Sharma S, Tripathi P. Gut microbiome and type 2 diabetes: where we are and where to go? J Nutr Biochem. 2019;63:101-8. doi: 10.1016/j.jnutbio.2018.10.003. PubMed PMID: 30366260.
33 Woldeamlak B, Yirdaw K, Biadgo B. Role of Gut Microbiota in Type 2 Diabetes Mellitus and Its Complications: Novel Insights and Potential Intervention Strategies. Korean J Gastroenterol. 2019;74:314-20. doi: 10.4166/kjg.2019.74.6.314. PubMed PMID: 31870137; PubMed Central PMCID: PMC12285806.
34 Sasidharan Pillai S, Gagnon CA, Foster C, Ashraf AP. Exploring the Gut Microbiota: Key Insights Into Its Role in Obesity, Metabolic Syndrome, and Type 2 Diabetes. J Clin Endocrinol Metab. 2024;109:2709-19. doi: 10.1210/clinem/dgae499. PubMed PMID: 39040013; PubMed Central PMCID: PMC11479700.
35 Zhang L, Luo J, Li X, Guo S, Shi D. 16S rRNA Sequencing and Metagenomics Study of Gut Microbiota: Implications of BDB on Type 2 Diabetes Mellitus. Mar Drugs. 2020;18. doi: 10.3390/md18090469. PubMed PMID: 32957565; PubMed Central PMCID: PMC7551199.
36 Wu Z, Gong C, Wang B. The relationship between dietary index for gut microbiota and diabetes. Sci Rep. 2025;15:6234. doi: 10.1038/s41598-025-90854-y. PubMed PMID: 39979448; PubMed Central PMCID: PMC11842723.
37 Brial F, Chilloux J, Nielsen T, Vieira-Silva S, Falony G, Andrikopoulos P, et al. Human

and preclinical studies of the host-gut microbiome co-metabolite hippurate as a marker and mediator of metabolic health. Gut. 2021;70:2105-14. doi: 10.1136/gutjnl-2020-323314. PubMed PMID: 33975870; PubMed Central PMCID: PMC8515120.

38 Canfora EE, Meex RCR, Venema K, Blaak EE. Gut microbial metabolites in obesity, NAFLD and T2DM. Nat Rev Endocrinol. 2019;15:261-73. doi: 10.1038/s41574-019-0156-z. PubMed PMID: 30670819.