Analysis of Gut Microbiome Dynamics Following Prebiotic Supplementation

Vanessa Rodriguez[1], Clarisse Nobre[1] and Andreia Salvador[1]

[1] Centre of Biological Engineering, University of Minho
PG49131@ALUNOS.UMINHO.PT

Abstract. The human gut microbiota is a complex and dynamic ecosystem, whose composition and functionality are modulated by components such as prebiotics. This project aims to investigate how prebiotic supplementation alters microbial composition and to uncover the metabolomic basis behind the enrichment of specific bacterial taxa, focusing on their enzymatic capacity to metabolize prebiotics. We propose a modular bioinformatics pipeline that integrates taxonomic profiling via 16S rRNA sequencing and functional prediction using tools such as GUMPP, PICRUSt2, KEGG, CAZy, and UPIMAPI. The methodology includes statistical analysis of microbiome data (e.g., Yacon polysaccharides and microbial cellulose case studies), followed by enzyme mapping and metabolic pathway reconstruction. The expected outcome is a refined tool that bridges taxonomic shifts with functional capacity, providing mechanistic insights into microbiota-prebiotic interactions and enabling the prediction of structure-function relationships in the gut microbiome.

Keywords: prebiotics, gut microbiota, functional profiling.

1 State of the Art

Human gut microbiota is a complex ecosystem of trillions of microorganisms that play essential roles in digestion, immunity, and overall health [1]. Prebiotics, defined as substrates thar are selectively utilized by host microorganisms conferring a health benefit, are known to influence the growth of beneficial bacteria [2,3]. Recent advances have shown that the specific response of gut microbiota to prebiotics is dependent on their genomic capability to metabolize these compounds [4].

Understanding how gut microbiota metabolize prebiotics requires not only biological insight but also the application of bioinformatics tools capable of interpreting large-scale sequencing data. Advancements in high-throughput sequencing, especially 16S rRNA gene sequencing, have allowed researchers to profile microbial communities at unprecedented depth. The 16S rRNA gene, highly conserved across bacteria but with hypervariable regions, is commonly used for bacterial identifications and phylogenetic studies. This gene contains nine hypervariable regions (V1-V9) interspersed with conserved sequences and amplifying and sequencing specific variable regions allows researchers to classify bacteria to the genus or species level [5]. When using Illumina sequencing, a method of high-throughput sequencing technology, microbial DNA is

first extracted from samples such as fecal material, to then amplify a target region, commonly V3-V4, of the 16S rRNA gene using specific primers [6]. Then, the amplicons are barcoded, prepared for sequencing and the Illumina MiSeq generates millions of short reads, 150-300 bp paired-end reads, which originates a FASTQ file which is a text file that contains the raw DNA sequence reads and corresponding quality scores (Phred scores) used for downstream analysis, which indicate the confidence in each base call [7,8]. These files form the starting point of any microbiome bioinformatic workflow.

The FASTQ files are processed to quality filter, trim reads to remove low-quality bases, merge paired-end reads and de-noise to distinguish Amplicon Sequence Variants (ASVs) or cluster Operational Taxonomic Units (OTUs). For this, two widely used command-line bioinformatics platforms are used: QIIME2 and mothur. QIIME2 offers modular plugins for each step: importing data, denoising, taxonomic assignment, diversity analysis, and visualization [9,10]. Meanwhile, mothur is another powerful pipeline, particularly strong for OUT-based workflows, also performed via the command line with a sequence of commands for each processing step [10,11]. The processed sequences are then matched to known bacterial sequences in reference databases such as greengenes which is a commonly used database for 16S rRNA gene sequences and SILVA and RDP which are alternatives with broader or more updated coverage [12-14]. Moreover, to analyze and interpret the large datasets generated in microbiome studies, MicrobiomeAnalyst enables statistical analysis, visualization, and interpretation of microbiome data, making it accessible for researchers with varying levels of bioinformatics expertise [15].

However, 16S rRNA sequencing only provides taxonomic insight and lacks direct functional information, and to overcome this bioinformatics tools have been developed. For example, PICRUSt2 was developed to predict the functional potential of microbial communities based on taxonomy which help infer the metabolic pathway that are likely present in a microbiome and thus allow the identification of bacteria capable of performing key functions, such as those involved in prebiotics utilization [16]. Studies have demonstrated that taxa like Bifidobacterium and certain Bacteroides species often increase in abundance following prebiotic intake, such as inulin or fructooligosaccharides (FOS) [17,18]. The ability of these bacteria to thrive is linked to their possession of specific glycoside hydrolases and other carbohydrate-active enzymes (CAZymes), which are encoded in their genomes. Mining these genetic features provides a mechanistic understanding of microbiota shifts in response to prebiotic interventions. Therefore, while 16S rRNA sequencing provides a cost-effective and scalable method for profiling microbial communities, metagenomics and metatranscriptomics offer higher resolution insights into gene content and expression, respectively. Nonetheless, predictive tools like PICRUSt2 serve as a bridge between taxonomic and functional data, especially when metagenomic sequencing is not feasible.

While tools like PICRUST2 and databases such as CAZy allow for functional predictions based on 16S rRNA sequencing data, there is currently no streamlined tool that correlates taxonomic growth patterns under specific prebiotic conditions with the presence of enzymes responsible for metabolizing those prebiotics. More recently, GUMPP (General Unified Microbiome Profiling Pipeline) has emerged as a comprehensive

framework for reproducible, large-scale 16S rRNA analysis. GUMPP unifies taxonomic and functional profiling using QIIME2 and PICRUSt2, with built-in support for KEGG, MetaCyc, and enzyme commission (EC) number annotations [19]. Existing tools focus either on broad functional prediction or enzyme annotation from known genomes, lacking direct linkage to experimental outcomes. Developing a custom tool or script that automates the identification of metabolic enzymes in bacteria that increased in relative abundance would provide a novel approach to understanding microbiota-prebiotic interactions.

2 Aims

In this project, we aim to analyze how prebiotic supplementation affects microbiota composition and explore why certain bacterial taxa thrive by investigating their genomic potential, particularly their ability to produce enzymes that metabolize specific prebiotics. Thus, the main aims of this project are:

- Conduct a bioinformatic survey for functional analysis based on the taxonomic composition of gut microbiota communities.
- Define a bioinformatics workflow for:
    - Extracting relevant functional information on gut microorganisms selected based on statistical reports.
    - Investigating the functional potential of these microorganisms by identifying enzymes and metabolic pathways associated with prebiotic metabolism.
- Perform statistical and functional analysis of gut microbiome metataxonomic data exposed to different prebiotics.
- Identify the potential for degrading specific prebiotics and their associated enzymes using a custom bioinformatics tool.

Therefore, the custom bioinformatic tool will help to predict structure-prebiotic relationship of dietary compounds in the microbiome.

3 Preliminary Problem Analysis

Despite advancements in sequencing and bioinformatics, a gap remains between identifying which bacteria grow in response to prebiotics and understanding why these taxa are favored. Knowing the relative abundance of bacteria after prebiotic treatment does not provide insight into their functional capabilities. To bridge this gap, it is necessary to investigate the genomic potential of bacteria, particularly focusing on genes that encode enzymes involved in prebiotic metabolism.

Furthermore, 16S rRNA sequencing offers limited resolution, as it identifies bacteria only at the genus or species level and does not directly inform functional genes. Functional inference using tools like PICRUSt2 introduces predictive capabilities, but these are limited by the availability and quality of reference genomes. Accurate functional predictions require well-annotated reference databases and sophisticated statistical

analysis to validate findings. Meanwhile GUMPP unifies the taxonomic and functional prediction process in a reproducible and modular fashion, reducing variability and manual overhear in the analysis. It provides a scalable platform for generating consistent outputs across large sample sets and integrates enzyme-level predictions that can be directly linked to prebiotic metabolism.

However, even with GUMPP, most current pipelines focus either on broad taxonomic shifts or general functional predictions but lack specificity in correlating bacterial taxa with the exact enzymes or pathways involved in prebiotic degradation and/or other enzymes involved in other metabolic processes of interest. Developing a custom tool that builds on GUMPP's output to highlight taxon-specific enzyme repertoires would provide new insights into the metabolic basis of microbiota-prebiotic interactions, specifically for the novel prebiotic substrates of Yacon polysaccharides and microbial cellulose, allowing a more precise and biologically grounded explanation for observed microbiota shifts.

4 Methodology

This project proposes a modular bioinformatics pipeline designed to integrate taxonomic and functional analysis of gut microbiota communities in response to prebiotic supplementation.

The bioinformatics pipeline will follow a modular structure with two phases: (1) taxonomic profiling and (2) functional inference and enzymatic mapping, the latter representing the core innovative contribution of this project.

The modular structure allows researchers to either initiate the pipeline from raw sequencing data or from preprocessed taxonomic abundance tables, the focuses on the selection of microorganisms of interest from the complex microbial community (likely the most active) based on statistical analysis of taxonomic shifts, while the second phase involves functional characterization using computational tools, i.e., the identification of the genes in the genomes of the microorganisms selected in phase 1, that are involved in relevant metabolic processes.

4.1 Phase 1: Selection of Microorganisms of Interest

The analysis begins with the results of the taxonomic classification using 16S rRNA sequencing data. The input consists of an Excel file containing taxonomic and abundance data, formatted for compatibility with MicrobiomeAnalyst, a tool used to perform statistical analysis, visualization, and diversity assessments.

The selection of microorganisms is based on statistical reports identifying taxa that significantly increase in abundance following prebiotic administration, compared to baseline and control conditions, as well as those that remain abundant over time (Figure 1).

This phase is designed to identify microorganisms that significantly increase in relative abundance following prebiotic intervention. To accommodate varying levels of data processing, this phase supports two entry points:

- Option A: start from raw 16S rRNA sequencing data in FASTQ format, which are processed using GUMPP, integrating several bioinformatic tools including QIIME2 for sequence quality control, taxonomic classification, and diversity analysis.
- Option B: accepts pre-analyzed taxonomic abundance data, such as Excel files containing genus- or species-level classification with corresponding relative abundances.

In both input scenarios, the next step involves statistical analysis using MicrobiomeAnalyst, to identify microbial taxa that significantly change in abundance under prebiotic treatment compared to baseline and control conditions, as well as those that remain abundant over time (Figure 1).

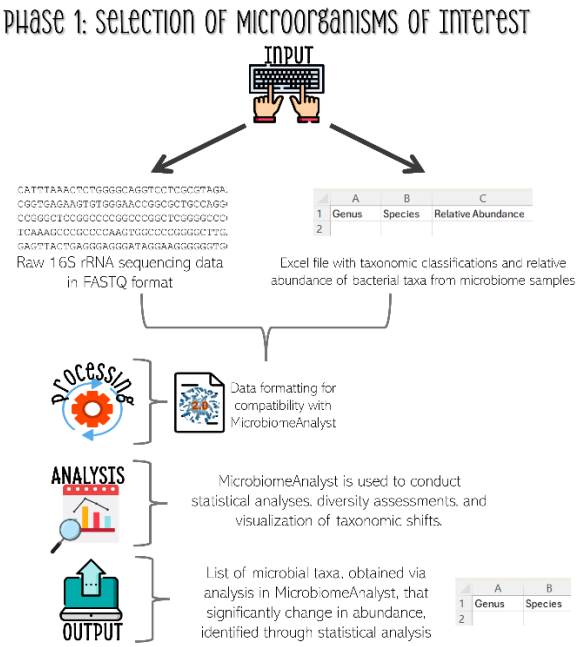

Fig. 1. Representation of the workflow for selecting microbial taxa that significantly changes in abundance following prebiotic intervention (Phase 1). The input could consist of either a FASTQ file of raw 16S rRNA sequencing data or a preprocessed Excel file with taxonomic classification and relative abundances data. The data is formatted for compatibility with MicrobiomeAnalyst, which is then used to perform statistical analysis, diversity assessments, and visualization of taxonomic shifts. The output is a list of microbial taxa that significantly increase in abundance compared to baseline and control conditions or remain abundant over time.

The selection process of bacteria will be automated but is expected to allow for manual refinement by the user. The outcome of Phase 1 is a refined list of microbial taxa that are enriched or consistently abundant in response to specific prebiotics, such as polysaccharides from Yacon or microbial cellulose. This list forms the foundation for functional exploration in Phase 2.

4.2 Phase 2: Functional Analysis

Phase 2 constitutes the main innovation of the project, focusing on the functional characterization of the microbial taxa identified in Phase 1. The aim is to understand the metabolic potential of these bacteria, particularly their ability to degrade or utilize specific prebiotic compounds and other metabolic processes of interest, such as those linked to short-chain fatty acid production SCFAs (e.g., butyrate) and key metabolic intermediates such as acetyl-CoA.

In this phase, GUMPP will be used to provide predicted gene family profiles, through its PICRUSt2 integration, and pathway reconstruction with the aid of KEGG Orthology, MetaCyc pathways and EC numbers, facilitating the correlation between microorganisms and their metabolic capabilities.

The core component of Phase 2 is a custom-developed module that automates the mapping of microbial taxa to enzymes and metabolic pathways relevant to prebiotic metabolism. This tool queries databases such as CAZy and KEGG to identify enzyme families known to act on the chemical structures of the studied prebiotics. By cross-referencing these enzymes with the taxa enriched in Phase 1, the module can generate hypotheses about which organisms are metabolically equipped to degrade specific substrates. Additionally, the tool integrates UPIMAPI to enrich functional annotations by retrieving UniProt-based information, including EC numbers, enzyme names, Gene Ontology (GO) terms, and links to relevant pathways (Figure 2).

The final output of this phase is a structured report linking microbial taxa to their functional roles in the degradation of prebiotics, providing a mechanistic interpretation of observed taxonomic shifts. This includes not only the identification of prebiotic-specific enzymatic capabilities but also broader metabolic functions such as SCFAs production and carbohydrate metabolism. By bridging the gap between taxonomic and functional data, this custom pipeline offers a novel and biologically grounded approach to understanding gut microbiota responses to dietary interventions.

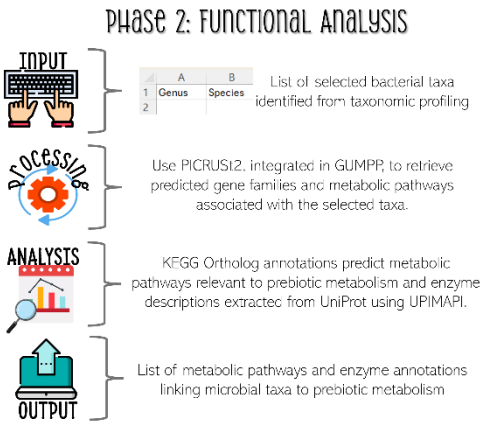

Fig. 2. Illustration of the functional analysis workflow to determine the metabolic potential of the selected microorganisms (Phase 2). The input consists of a list of microbial taxa identified in Phase 1. GUMPP will then be used to provide gene family profiles and reconstruct metabolic

pathways of the aforementioned bacteria through PICRUSt2, KEGG, MetaCyc and EC integration. The analysis includes KEGG Ortholog annotations to predict metabolic pathways involved in prebiotic metabolism, as well as enzyme descriptions obtained from UniProt using UPIMAPI. The final output is a comprehensive functional characterization linking microbial taxa to relevant metabolic pathways.

4.3 Work Plan

The methodology will be systematically implemented through the following steps:

1. Literature review
2. Modular pipeline design
3. Statistical analysis of two study cases: polysaccharides from Yacon and Microbial Cellulose, with MicrobiomeAnalyst
4. Functional prediction
5. Link taxa to prebiotic metabolism using CAZy and/or KEGG databases
6. Functional annotation using UPIMAPI
7. Develop and test a custom enzyme detection script to automatize workflow
8. Statistical analysis, visualization, report drafting and validation applied to the two study cases

By integrating taxonomic shifts with functional predictions, this workflow will provide a more detailed understanding of how gut microbiota respond to prebiotic interventions and identifies key microorganisms involved in these processes.

References

1. Flint HJ, Scott KP, Louis P, Duncan SH. 2012. The role of the gut microbiota in nutrition and health. Nature Reviews Gastroenterology & Hepatology. 9(10):577–89
2. Bedu-Ferrari C, Biscarrat P, Langella P, Cherbuy C. 2022. Prebiotics and the Human Gut Microbiota: From Breakdown Mechanisms to the Impact on Metabolic Health. Nutrients. 14(10):2096
3. Yoo S, Jung S-C, Kwak K, Kim J-S. 2024. The Role of Prebiotics in Modulating Gut Microbiota: Implications for Human Health. International journal of molecular sciences. 25(9):4834–34
4. Sonnenburg Erica D, Sonnenburg Justin L. 2014. Starving our Microbial Self: The Deleterious Consequences of a Diet Deficient in Microbiota-Accessible Carbohydrates. Cell Metabolism. 20(5):779–86
5. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 1991. 16S ribosomal DNA amplification for phylogenetic study. Journal of bacteriology. 173(2):697–703
6. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, et al. 2012. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Research. 41(1):e1–1
7. Cock PJA, Fields CJ, Goto N, Heuer ML, Rice PM. 2010. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Research. 38(6):1767–71
8. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. 2013. Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence

Data on the MiSeq Illumina Sequencing Platform. Applied and Environmental Microbiology. 79(17):5112–20

9. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, et al. 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome. 6(1):
10. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, et al. 2009. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Applied and Environmental Microbiology. 75(23):7537–41
11. Nearing JT, Douglas GM, Comeau AM, Langille MGI. 2018. Denoising the Denoisers: an independent evaluation of microbiome sequence error-correction approaches. PeerJ. 6:e5364
12. Christian Q, Elmar P, Pelin Y, Jan G, Timmy S, et al. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research. 41(1):590–96
13. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, et al. 2013. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Research. 42(D1):D633–42
14. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, et al. 2006. Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. Applied and Environmental Microbiology. 72(7):5069–72
15. Lu Y, Zhou G, Ewald J, Pang Z, Shiri T, Xia J. 2023. MicrobiomeAnalyst 2.0: comprehensive statistical, functional and integrative analysis of microbiome data. Nucleic Acids Research. 51(1):310–18
16. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, et al. 2020. PICRUSt2 for prediction of metagenome functions. Nature Biotechnology. 38(6):685–88
17. Rossi M, Corradini C, Amaretti A, Nicolini M, Pompei A, et al. 2005. Fermentation of Fructooligosaccharides and Inulin by Bifidobacteria: a Comparative Study of Pure and Fecal Cultures. Applied and Environmental Microbiology. 71(10):6150–58
18. Yin P, Yi S, Du T, Zhang C, Yu L, et al. 2024. Dynamic response of different types of gut microbiota to fructooligosaccharides and inulin. Food & Function. 15(3):1402–16
19. Murovec B, Deutsch L, Stres B. 2021. General Unified Microbiome Profiling Pipeline (GUMPP) for Large Scale, Streamlined and Reproducible Analysis of Bacterial 16S rRNA Data to Predicted Microbial Metagenomes, Enzymatic Reactions and Metabolic Pathways. Metabolites. 11(6):336–36