

MICROBIOME SEQUENCING AND METHODOLOGICAL CHALLENGES: IMPLICATIONS FOR PERSONALIZED NUTRITION AND MEDICINE

AUTHORS:

María del Rosario Taussig ¹, Andrea Liliana Millán ², Rodrigo Peralta ^{1,3}, Josefina Arcagni ¹, María Constanza Pautasso ², Kyle McClary ⁵, Yap Chew ⁵, Jeffrey Koble ⁶, Gustavo Daniel Frechtel ^{2,4}, Juan Pablo Bustamante ^{1,3}

1. Laboratorio de Investigación, Desarrollo y Transferencia de la Universidad Austral (LIDTUA-CIC). Facultad de Ingeniería de la Universidad Austral. Pilar, Argentina.

2. Laboratorio de Diabetes y Metabolismo. Instituto de Inmunología, Genética y Metabolismo (INIGEM), UBA-CONICET, Buenos Aires, Argentina.

3. Facultad de Ingeniería, Universidad Nacional de Entre Ríos (FI-UNER), Oro Verde, Entre Ríos, Argentina.

4. Hospital de Clínicas, Cátedra de Nutrición, Facultad de medicina. Universidad de Buenos Aires, Argentina.

5. Zymo Research Corporation, Irvine, CA, USA.

6. Illumina, San Diego, CA, USA.

CORRESPONDING AUTHOR: jbustamante@austral.edu.ar

doi.org/10.55634/3.3.5

ABSTRACT:

This study investigates the impact of sequencing methodologies on microbial composition and functional dynamics in the context of dietary patterns. By analyzing the same stool samples using both 16S rRNA and shotgun metagenomics, we uncovered significant discrepancies in taxonomic resolution, with shotgun metagenomics identifying a broader range of microorganisms, including 22 phyla and 1489 species, compared to the 15 phyla and 509 species identified by 16S rRNA. Our results highlight the dependency of microbial abundances on sequencing methodology, raising concerns about the generalizability of associations reported in the literature. Notably, despite herein selected dietary patterns, plant-based and western, key genera and phyla displayed markedly different relative abundances, illustrating the complexity and variability inherent in microbiome research. Furthermore, we employed metatranscriptomics to differentiate between potential and active metabolic pathways, revealing that not all detected pathways at the DNA level are actively expressed, emphasizing the need for multi-omics approaches in future studies. The limited sample size in our analysis restricts comprehensive comparisons between dietary groups, underscoring the necessity for larger cohort studies. This work calls for standardized protocols in microbiome research and highlights the importance of addressing methodological sequencing biases, which can critically affect interpretations and applications looking for the development of actionable strategies for personalized nutrition and future health interventions.

KEYWORDS: *gut microbiota, gut microbiome, diet, 16S rRNA, shotgun metagenomics, metatranscriptomics, sequencing methodologies.*

INTRODUCTION

The study of the human gut microbiota has gained increasing attention in recent years due to its crucial role in maintaining health and its potential to modulate disease states¹⁻³. Publications in this matter are increasingly growing, which denotes a markable interest in research and application specifically on human health and disease balance about the microorganisms that inhabit the human body, particularly, at the gut⁴. As it is well known, the largest amount of

human microbiota resides in the colon⁵. The gut microbiota reveals a high degree of functional redundancy^{6,7}. Energy uptake of host nutrient metabolism taken from food components⁸, digestion of insoluble fibers from vegetable origin⁹, maintenance of structural integrity of the gut mucosal barrier^{9,10}, regulate host fat storage and modify bone mineral density¹¹, collaborate in the balance of the hormones ghrelin and leptin⁸, synthesize vitamin K and B12¹¹, help protect against damage or pathogen entry and stimulate intestinal angiogenesis¹¹, mature the immune systems,

regulating response and inflammation for example through the production of derived aromatic amino acids¹⁰, produce short chain fatty acids (SCFA)^{9,10}, regulate bile acids through the synthesis of metabolites derived from cholesterol¹⁰, and the level of phospholipid synthesis with the anaerobic metabolism of choline¹⁰ are some key functions of our gut microbiota.

Among all the factors that modulate the microbiota, diet plays a predominant daily role¹². It is, therefore, important for the advancement of science and personalized nutrition to identify health promoting or detrimental foods on dietary patterns, understanding their impact on gut microbiota, minimizing methodological bias. Particularly, two of the most worldwide referenced diet approaches¹²⁻¹⁴, western and plant-based diets are here considered as examples of how diverse the results can be independently of chosen diets. Evidence linking the gut microbiota with food intake and diverse health states is steadily increasing; however, knowledge about its composition and variation remains incomplete and lacks consensus. The use of diverse methodologies to analyze the microbiota has hindered global agreement on both its composition and its associations with different diseases.

As sequencing technologies advance, researchers have adopted a range of methodologies to explore the composition and functional capabilities of the gut microbiota. Among these, 16S rRNA gene sequencing and shotgun metagenomics are two of the most widely used approaches. However, both methods have distinct advantages and limitations, which can lead to differing results when interpreting microbiome data^{15,16}. 16S rRNA sequencing is often employed for its cost-effectiveness and ability to provide taxonomic resolution of bacterial communities at the genus level¹⁷. In contrast, shotgun metagenomics offers a more comprehensive view of the microbiota, enabling species-level identification and functional insights by sequencing the entire microbial genome¹⁸. Despite the richness of data that shotgun metagenomics can provide, it is also more resource-intensive and computationally demanding.

This study seeks to evaluate how these two methodologies influence the analysis of the gut microbiota in individuals following omnivorous and plant-based diets. By comparing taxonomic and functional outcomes derived from both methods, we aim to highlight the methodological biases that can shape our understanding of diet-microbiota interactions.

METHODS:

INDIVIDUALS RECRUITMENT AND SAMPLE COLLECTION

Data from forty eight participants of the clinical trial MicrobiAr: "Characterization and Follow-up of Microbiome and Health Indicators in Obese, Pre-Diabetes and Type 2 Diabetes Cohorts Undergoing a Plant-based Diet and Lifestyle Intervention" were used for this study. MicrobiAr was approved in 2021 by the Ethics Committee of Hospital de Clínicas and the Faculty of Pharmacy and Biochemistry at the University of Buenos Aires. It is also registered with the National Institutes of Health (NIH) in the United States under ID NCT05372445. The forty eight participants have signed the corresponding informed consent forms, 24 of whom following a western and 24 a plant-based dietary pattern. Fecal samples were collected by participants according to the uses detailed in the Bunny Wipe kit from Zymo Research Corporation. This kit has inside a collector tube where people should collect their own sample and drop it in a tube containing the reagent DNA/RNA Shield™. This reagent is a stabilizing solution for nucleic acids in any biological sample. It preserves the genetic integrity and expression profiles of the samples at room temperature and completely inactivates other biological agents, maintaining the sample as it was extracted. The unique DNA and RNA stabilization solution also prevents degradation from freeze-thaw cycles. Once the patient extracts and delivers the sample he/she has to fill out a form with related information.

DNA/RNA EXTRACTION

Starting from aliquots of the stabilized tube, the extraction of the nucleic acids takes place. For DNA extraction, the ZymoBIOMICS™ DNA Miniprep Kit was used, whereas for RNA, the RNA extraction RNeasy PowerMicrobiome® Kit of Qiagen. Concentration of nucleic acids extracted was determined with Qubit® 4.0. For both DNA and RNA, High Sensitivity detection kits from Invitrogen were used; one specific for DNA and RNA respectively. DNA and RNA integrity was checked in an electrophoretic gel with molecular weight marker lambda/HindIII. Extracted DNA and RNA were stored at -80°C.

16S RRNA TARGETED LIBRARY PREPARATION AND SEQUENCING

The DNA samples were prepared for targeted sequencing with the Quick-16S™ NGS Library Prep Kit (Zymo Research, Irvine, CA) using primer set V1-V2. These primers were custom-designed by Zymo Research to provide the best coverage of the 16S gene while maintaining high sensitivity. The sequencing library was prepared using an innovative library preparation process in which PCR reactions were performed in real-time PCR machines to control cycles and therefore limit PCR chimera formation. The final PCR products were quantified with qPCR fluorescence readings and pooled together based on equal molarity. The final pooled library was cleaned up with the Select-a-Size DNA Clean & Concentrator™ (Zymo Research, Irvine, CA), then quantified with TapeStation® (Agilent Technologies, Santa Clara, CA) and Qubit® (Thermo Fisher Scientific, Waltham, WA). The final library was sequenced on Illumina® NextSeq 2000™ with a P1 reagent kit (600 cycles) (Illumina, San Diego, CA). The sequencing was performed with 25% PhiX spike-in.

CONTROL SAMPLES

The ZymoBIOMICS® Microbial Community Standard (Zymo Research, Irvine, CA) was used as a positive control for each DNA extraction, if performed. The ZymoBIOMICS® Microbial Community DNA Standard (Zymo Research, Irvine, CA) was used as a positive control for each targeted library preparation. Negative controls (i.e. blank extraction control, blank library preparation control) were included to assess the level of bioburden carried by the wet-lab process.

SHOTGUN METAGENOMICS PROCEDURES

Once the DNA has been eluted and checked, the metagenomics libraries were prepared following the Illumina DNA Prep protocol according to Illumina Inc. All procedures were performed under sterile conditions to avoid contamination. As a reference control for fecal microbiome diversity, the Gut Microbiome Standard Reference from Zymo Research was used (data not shown). Once the libraries were done, they were quantified individually using Qubit® 4.0 and High Sensitivity kits as it was previously described at DNA/RNA extraction section. The pooling was carried out and measured in a Bioanalyzer 2100 with the DNA High Sensitivity kit. DNA sequencing was performed on Illumina's NovaSeq 6000.

METATRANSCRIPTOMICS PROCEDURES

Once the RNA has been eluted and checked, the libraries

were prepared according to the Illumina Stranded Total RNA Prep protocol from Illumina Inc. All procedures were performed under sterile conditions to avoid contamination. Once the libraries were finalized, they were quantified individually as already explained for shotgun metagenomics, and the pooling was measured as the same for shotgun metagenomics too. Sequencing was performed on Illumina's NovaSeq 6000.

BIOINFORMATIC ANALYSIS

After obtaining the sequencing data from amplicon sequences, raw paired-end reads (average sequencing depth 234,120 reads \pm 39,883 reads) were grouped based on their barcode sequences (demultiplexing) using the demux option in cutadapt¹⁹, which also removed the barcodes and primers. Sequence quality and changes following preprocessing were visualized using fastp²⁰. Deblur²¹ was used for selecting representative sequences, serving as proxies for species, based on denoising into ASVs. A feature table (ASV table) was then generated by quantifying the frequency of each feature sequence in every sample. Simultaneously, the feature sequences obtained in this step were used for taxonomic assignment at the phylum, genus, and species levels, offering a dimensionality reduction perspective on the microbiota. This was done by comparing the feature sequences with reference amplicon sequences stored in the SILVA²²—database (v138.2), using the VSEARCH²³ taxonomy classifier.

Quality control for the shotgun metagenomics (average sequencing depth 33,169,156 reads \pm 10,378,603 reads) and metatranscriptomic (average sequencing depth 30,816,243 reads \pm 20,782,514 reads) sequences were performed using Trimmomatic²⁴. Human DNA sequences were removed with Kneaddata (<https://bitbucket.org/biobakery/kneaddata/wiki/Home>) using the GRCh38 human genome (<https://www.ncbi.nlm.nih.gov/genome/guide/human/>) as a reference. Similar to the amplicon analysis, sequence quality was visualized using fastp.

The taxonomic profile was generated using MetaPhlan4²⁵, which aligns metagenomic reads to a predefined marker-gene database (mpa_vJan21_CHOCOPHlanSGB_202103) for precise taxonomic classification. Humann3²⁶ was used to obtain the functional profile, enabling the exploration of within- and between-sample contributive diversity, revealing the contributions of individual species to specific functions.

The evaluation of statistically significant differences

between the 16S rRNA and shotgun metagenomics techniques was performed using the Wilcoxon test (p value < 0.05) implemented in the Scipy²⁷python module. The data visualization was performed with Python modules: pandas²⁸, matplotlib²⁹and seaborn³

RESULTS:

TAXONOMIC RESOLUTION DIFFERENCES BETWEEN 16S RRNA AND SHOTGUN METAGENOMICS

The same samples were processed via both methods, 16S rRNA and shotgun metagenomics, to explore how these methodological differences in sequencing approaches may contribute to observed variations. Our analysis revealed substantial differences in the taxonomic profiles generated by 16S rRNA and shotgun metagenomics sequencing (see Figure 1 and Figure 2). The 16S rRNA approach provided a comprehensive view at the genus level but struggled to resolve differences at the species level. In contrast, shotgun metagenomics allowed for more detailed taxonomic resolution, identifying species that were not detected by 16S sequencing.

As it is expected, shotgun metagenomics sequencing identifies more taxas in all taxonomic levels looking at bacteria. A total of 15 phyla, 250 genera and 509 species were identified according to 16S rRNA while 22 phyla, 783 genera and 1489 species through shotgun metagenomics. At the genus level, of 897 genera identified across all samples, only 136 genera (15%) were identified with both amplicon and shotgun metagenomic sequencing. Amplicon sequencing of 16S rRNA gene identified 114 genera not found by metagenomic profiling, while only 647 genera were uniquely found using shotgun metagenomic sequencing. At the phylum level, of 26 phylum identified across all samples, only 11 phylum (42%) were identified with both amplicon and shotgun metagenomic sequencing. Amplicon sequencing of 16S rRNA gene identified 4 phyla not found by metagenomic profiling, while only 11 phyla were uniquely found using shotgun metagenomic sequencing.

Relative abundances of phyla (Figure 1) and genera (Figure 2) for most abundant microbes processed by shotgun metagenomics and 16S rRNA are shown, in blue and orange, respectively, for people from two groups, one of them (group A) following a western (Figure 1.A, Figure 2.A) and the other (group B) following a plant-based dietary pattern (Figure 1.B, Figure 2.B). For both groups A and B, the most abundant phyla are Firmicutes,

Bacteroidetes, Proteobacteria, Actinobacteria, Fusobacteria and Verrucomicrobia. Differences between both sequencing methods were evaluated with a Wilcoxon test. For both groups A and B, significant differences related to the chosen sequencing method were observed, with a wide range of variation in the abundances of Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria. These results support the fact that, especially for most reported phyla in the literature about the human gut (ratio

Firmicutes/Bacteroidetes, Actinobacteria and Proteobacteria), relative abundances strongly depend on the chosen methodology, regardless of the people's dietary pattern.

Regarding genera level (Figure 2), the most abundant shown in the graphic are Bacteroides, Bifidobacterium, Blautia, Faecalibacterium, Lactobacillus and Prevotella. Differences evaluated with a Wilcoxon test show significantly different results for group A for Bacteroides; Bifidobacterium and Lactobacillus, whereas for group B, only Lactobacillus show significant differences.

As observed, certain genera reported here, such as Faecalibacterium (Figure 2), did not show significant differences based on the sequencing method, despite belonging to a phylum that did exhibit notable variation, like Firmicutes (Figure 1). This scenario highlights that within the Firmicutes phylum, there are microbes, at genera level, with marked differences in measured relative abundances, depending on the sequencing methodology employed.

At the species level (Figure 3), the top 10 most abundant microbes identified by each sequencing method are shown. Notably, none of these species overlap between the two methods, with no single species shared in the top 10, regardless of the diet followed by the individuals.

ANALYSIS OF POTENTIAL AND ACTIVE FUNCTIONAL GROUPS

The same samples were also analyzed using metatranscriptomics. In Group A (Western diet), 552 potential metabolic pathways were identified through shotgun metagenomics, while 432 active pathways were detected via metatranscriptomics. Similarly, in Group B (plant-based diet), 515 putative pathways were found via shotgun metagenomics, compared to 389 detected through metatranscriptomics. These findings align with the well-established observation that not all metabolic pathways identified at the DNA level are actively transcribed into RNA, underscoring the difference between potential functions and those that are truly

expressed. Figure 4 highlights 9 selected metabolic pathways to illustrate the discrepancies between genes identified as potential functions (via shotgun metagenomics, MG) and those actively expressed, as measured by metatranscriptomics (MT), for individuals following a western (A) or plant-based diet (B). Results are presented as heat maps, showing significant differences in the reported values using a graded color scale, according to the chosen methodological approaches, MG and MT.

Overall, the majority of metabolic pathways are active in both groups, regardless of the observed relative abundance of corresponding genes indicated by the MG approach, or even independent of dietary patterns. There are pathways identified as potential based on MG data that are inactive when analyzed through MT, as illustrated in Figure 4. The transitions from cyan to blue or even to black indicate the shift from MG to MT, respectively. Interestingly, there are also instances where very low relative abundance is observed in MG, yet high expression is detected in MT, as shown by transitions from black to red.

DISCUSSION:

The findings presented in this study highlight significant discrepancies between microbial abundances observed in stool samples, which could represent a high impact when looking for associations according to dietary patterns when processed through different sequencing methodologies (as herein for 16S rRNA and shotgun metagenomics). Our results underscore the complexity and variability inherent in microbiome research, raising important considerations for future studies that seek to link dietary intake with gut microbial composition and function, ultimately aiming to inform actionable strategies for personalized nutrition and medicine.

METHODOLOGICAL BIASES AND THEIR IMPLICATIONS

It is important to note when selecting the sequencing methodology that there is a threefold difference in associated costs between choosing 16S rRNA and shotgun metagenomics. This cost difference may lead researchers to opt for the cheaper option, but it's crucial to understand the scope and limitations, especially when aiming to find actionable insights for advancing personalized medicine.

One of the most striking observations in our analysis is the profound influence of the chosen sequencing

methodology on the resulting microbial profiles. As demonstrated in Figures 1 and 2, the same samples processed via 16S rRNA amplicon sequencing and shotgun metagenomics yielded markedly different taxonomic identifications at all levels (phylum, genus, and species). While 16S rRNA sequencing captured fewer taxa overall, shotgun metagenomics uncovered a broader range of microorganisms, identifying more genera and species, many of which were missed by 16S sequencing. This aligns with previous reports emphasizing the limitations of amplicon-based methods, which often miss certain taxa and provide less comprehensive functional insights compared to shotgun metagenomics³¹⁻³³.

The observed methodological discrepancies have practical consequences for the reproducibility and comparability of microbiome research. As seen in our study, some key genera, such as *Lactobacillus*, and several phyla, such as Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria demonstrated significantly different relative abundances depending on the sequencing method used. This variability complicates the establishment of clear associations between dietary patterns and microbial profiles, illustrating how methodological choices can influence the interpretation of microbial community composition^{34,35}. These findings suggest that results from different studies using varied methods may not always be directly comparable, even when the same microbes are being investigated²⁴. Therefore, a clear methodological consensus is needed to ensure consistency in future studies.

DIETARY PATTERNS AND MICROBIAL ABUNDANCE

The association between diet and microbial composition remains an area of active investigation. In our analysis, both case studies of plant-based and western diets demonstrated a dependency of microbial abundances on sequencing methodology, raising questions about the generalizability of potential associations made in the literature across dietary patterns. Due to systematic reviews and meta-analyses constituting the top of the evidence hierarchy, reporting the consensus about food intake according to dietary patterns and their impact on gut microbiome is crucial³⁶. With the growing body of evidence that diet impacts the human gut¹²⁻¹⁴, its relation with human health, controlled feeding studies to identify their effects on the gut microbiota are needed^{13,37}, taking into account these methodological biases herein presented. The limited sample size in this study

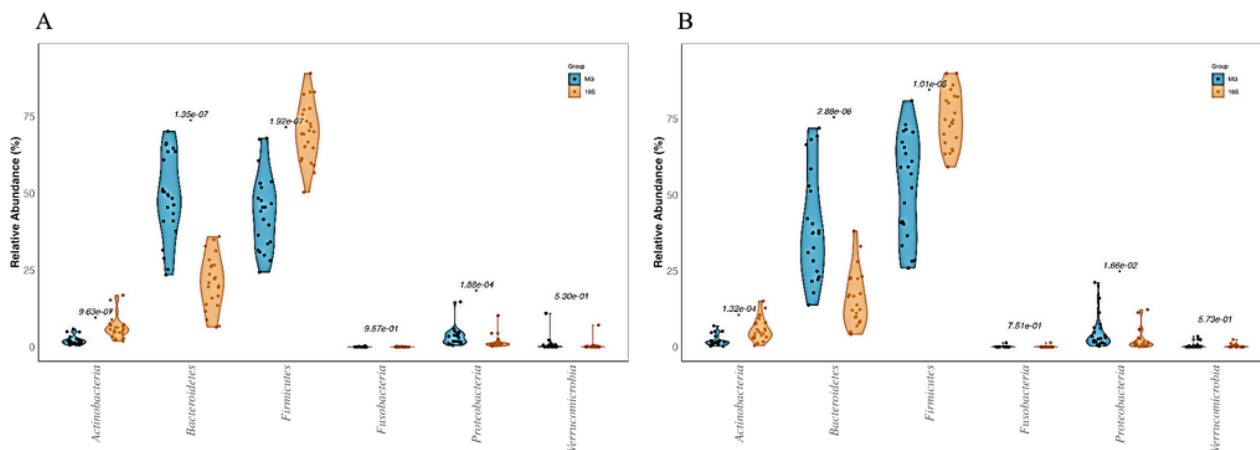


FIGURE 1. VIOLIN PLOTS FOR THE MOST ABUNDANT PHYLA IN SAMPLES FROM (A) PEOPLE FOLLOWING A WESTERN AND (B) FOLLOWING A PLANT-BASED DIETARY PATTERN MEASURED THROUGH SHOTGUN METAGENOMICS (BLUE) AND 16S RRNA SEQUENCING (ORANGE), N = 24. P VALUES LOWER THAN 0.05 ARE SHOWN WITH AN ASTERISK.

constrains our ability to conduct a comprehensive analysis of differences in microbial composition between the two dietary groups, making it challenging to draw definitive conclusions regarding their comparative effects. To address this limitation, future studies should increase the sample size and focus on the contributions and relationships of specific microbes with the dietary patterns investigated.

FUNCTIONAL INSIGHTS: SHOTGUN METAGENOMICS VS METATRANSCRIPTOMICS

While shotgun metagenomics provides a comprehensive view of the potential metabolic capabilities of the gut microbiome, it is the metatranscriptomic data that offers insight into active metabolic pathways. Our findings show a clear distinction between the functional potential of the microbiota, as detected by shotgun metagenomics, and the actual expression of those functions, as revealed by metatranscriptomics. In both diet groups, a subset of metabolic pathways appeared inactive at the genomic level but were active when analyzed through RNA transcripts, underscoring the importance of incorporating functional metatranscriptomics into microbiome research³⁸. For example, pathways associated with the production of short-chain fatty acids (SCFAs) were prevalent in both western and plant-based diets, but only a portion of these pathways herein considered were actively transcribed. This emphasizes the importance of evaluating both potential and expressed functions in the gut microbiome, as functional

discrepancies could lead to misinterpretations if only DNA-based methods are used. Furthermore, the pronounced differences in expression observed in key metabolic pathways highlight the necessity of integrating multi-omics approaches to gain a more complete understanding of the microbiome’s functional contributions to human health^{39,40}.

LIMITATIONS AND FUTURE DIRECTIONS

While this study provides a comprehensive analysis of the methodological sequencing challenges in microbiome research, several limitations warrant consideration. First, our sample size, though sufficient to reveal methodological sequencing biases, remains relatively small. In this sense, larger cohort studies are necessary for the impact of diet on microbial profiles, and other factors such as host genetics, lifestyle, and environmental influences that were not considered in this study but are known to play a role in shaping the gut microbiota.

Moving forward, it will be essential to establish standardized protocols for microbiome sampling, sequencing, and data analysis. This will enable more consistent, comparable and reproducible results across studies, facilitating the development of robust biomarkers for personalized nutrition and health interventions. Furthermore, expanding the use of multi-omics approaches, particularly metatranscriptomics, will enhance our understanding of the functional dynamics of the microbiome. While these methods may not

currently yield immediate actionable data for clinical and nutritional applications, they will pave the way for future insights. Achieving this goal requires addressing methodological biases and deepening our comprehension of the microbiota, its environment, dietary influences, and the host-microbiome relationship.

CONCLUSION:

This study reinforces the importance of considering methodological biases in microbiome research and highlights the critical need for standardized protocols. The significant differences in microbial abundances observed between 16S rRNA and shotgun metagenomics emphasize the challenges of comparing studies that employ different methods. Additionally, our findings underscore the potential of multi-omics approaches to uncover functional insights that may be missed by genomic analyses alone. As research continues to explore the relationship between diet and the gut microbiome, the integration of these advanced methodologies will be key to achieving more accurate and actionable results for personalized nutrition and health.

ACKNOWLEDGEMENTS

We are deeply thankful to MicrobiAr's (ClinicalTrials.gov ID NCT05372445) colleagues and community for their essential collaboration and contributions.

We would also like to express our sincere gratitude to Zymo Research Corporation and to Illumina, Inc. for their invaluable support in this study by providing us with processing and sequencing kits, which were essential for the development of our research. These collaborations had been key to advancing our understanding of the interactions between microbiota and sequencing methods, and its impact on scientific research.

This work was carried out with Rodrigo Peralta as a beneficiary of the Centro de Investigaciones Científicas (CIC) PhD scholarship.

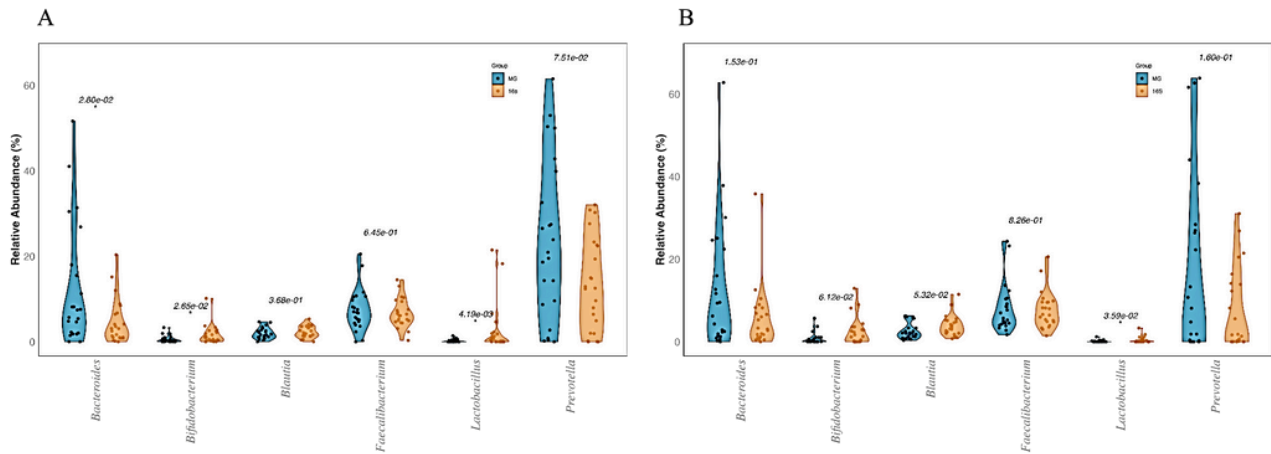


FIGURE 2. VIOLIN PLOTS FOR THE MOST ABUNDANT GENERA IN SAMPLES FROM (A) PEOPLE FOLLOWING A WESTERN AND (B) FOLLOWING A PLANT-BASED DIETARY PATTERN MEASURED THROUGH SHOTGUN METAGENOMICS (BLUE) AND 16S RRNA SEQUENCING (ORANGE), N = 24. P VALUES LOWER THAN 0.05 ARE SHOWN WITH AN ASTERISK.

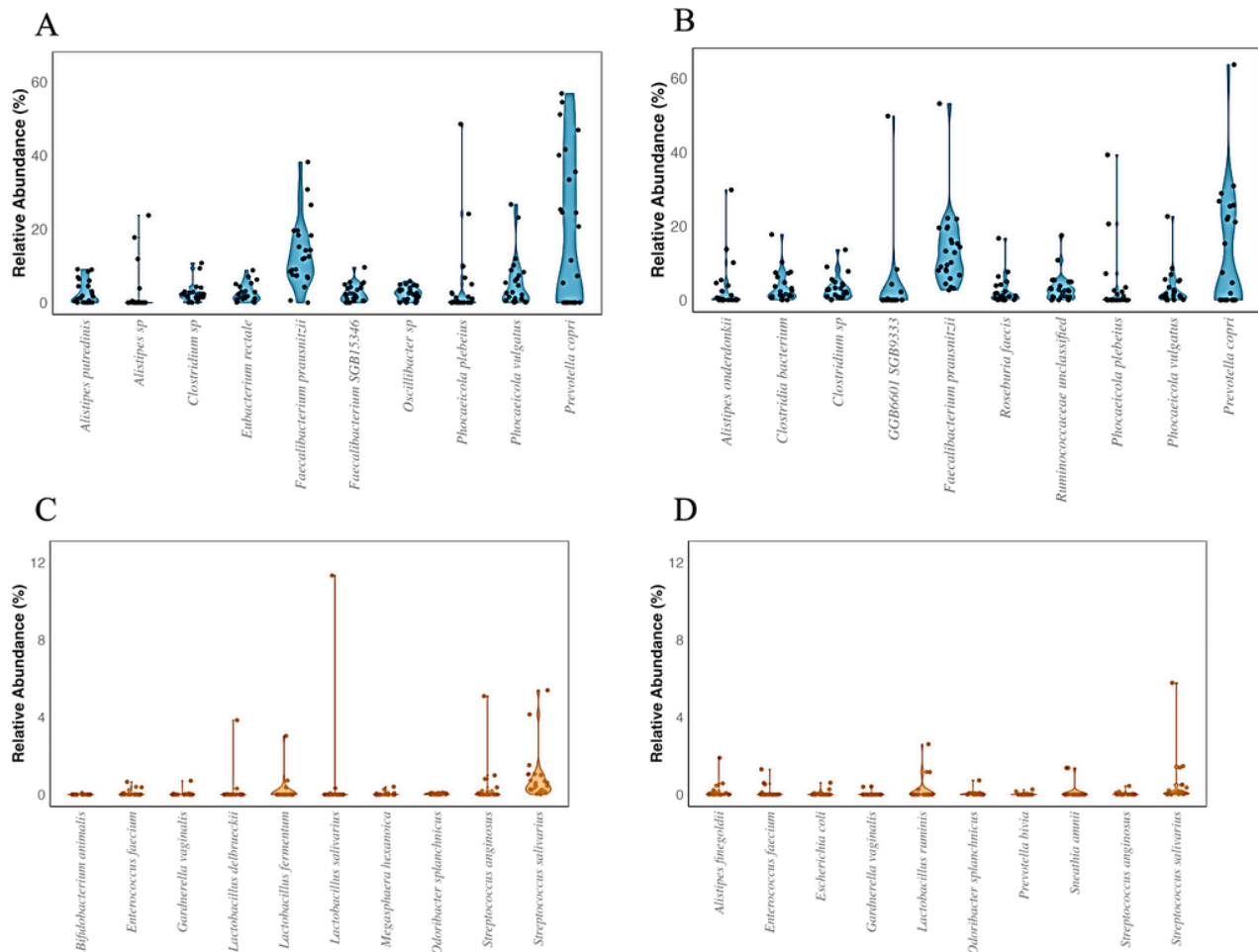


FIGURE 3. VIOLIN PLOTS REPRESENTING THE 10 MOST ABUNDANT SPECIES IN SAMPLES FROM INDIVIDUALS FOLLOWING A (A, C) WESTERN DIET AND (B, D) PLANT-BASED DIET, AS MEASURED BY SHOTGUN METAGENOMICS (BLUE) AND 16S RRNA SEQUENCING (ORANGE), N = 24.

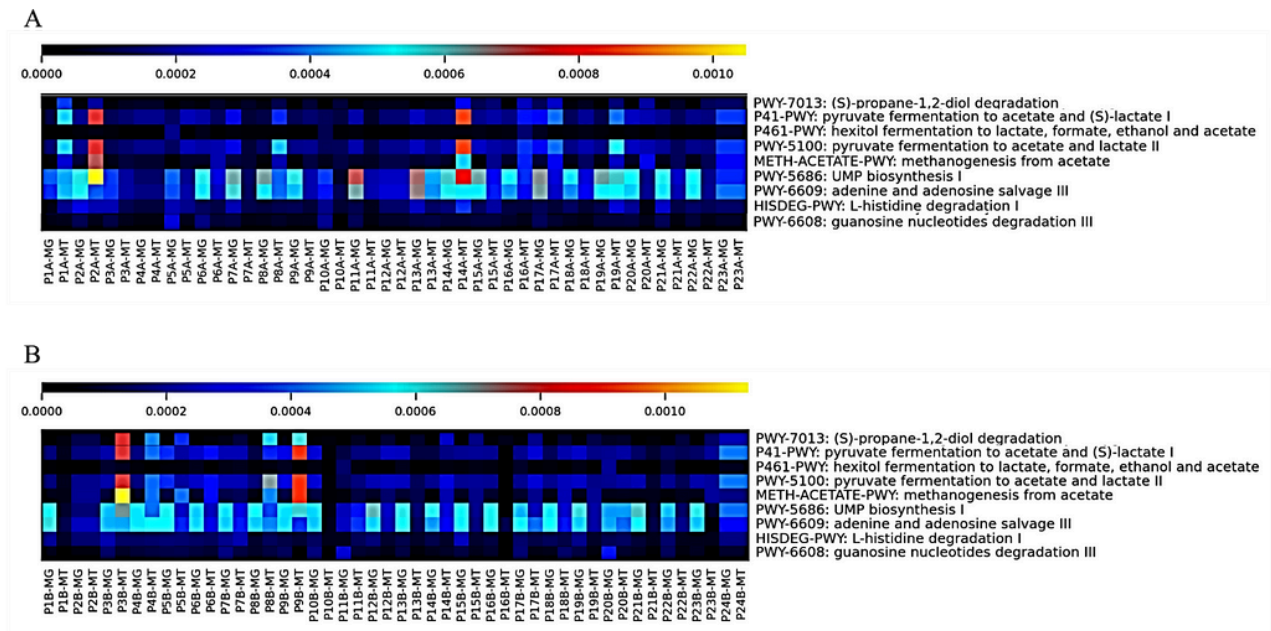


FIGURE 4. HEAT MAPS SHOWING 9 METABOLIC PATHWAYS OBSERVED BY CORRESPONDING GENES AS POTENTIAL FUNCTIONS (VIA SHOTGUN METAGENOMICS, MG) AND THOSE ACTUALLY EXPRESSED, AS MEASURED BY METATRANSCRIPTOMICS (MT), FOR INDIVIDUALS FOLLOWING A WESTERN (A) AND A PLANT-BASED DIET (B). SAMPLES FROM PARTICIPANTS UNDER A WESTERN DIET (A) AND PLANT-BASED DIET (B) ARE SHOWN. THE X-AXIS DISPLAYS THE RESULTS FOR EACH INDIVIDUAL, LABELED FOR MG AND MT IN AN INTERLEAVED FORMAT.

REFERENCES

1. Yue, Y.-Y. et al. Bibliometric analysis of subject trends and knowledge structures of gut microbiota. *World J. Clin. Cases* 8, 2817–2832 (2020).
2. Lu, G. et al. Global Trends in Research of Pain-Gut-Microbiota Relationship and How Nutrition Can Modulate This Link. *Nutrients* 15, 3704 (2023).
3. Zhang, B. et al. Global research trends on the links between the gut microbiota and diabetes between 2001 and 2021: A bibliometrics and visualized study. *Front. Microbiol.* 13, 1011050 (2022).
4. Cani, P. D. Human gut microbiome: hopes, threats and promises. *Gut* 67, 1716–1725 (2018).
5. Ojo, O., Ojo, O. O., Zand, N. & Wang, X. The Effect of Dietary Fibre on Gut Microbiota, Lipid Profile, and Inflammatory Markers in Patients with Type 2 Diabetes: A Systematic Review and Meta-Analysis of Randomised Controlled Trials. *Nutrients* 13, 1805 (2021).
6. Thursby, E. & Juge, N. Introduction to the human gut microbiota. *Biochem. J.* 474, 1823–1836 (2017).
7. The Human Microbiome Project Consortium. Structure, Function and Diversity of the Healthy Human Microbiome. *Nature* 486, 207–214 (2012).
8. Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R. & Gordon, J. I. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat. Rev. Microbiol.* 6, 776–788 (2008).
9. Medawar, E., Huhn, S., Villringer, A. & Veronica Witte, A. The effects of plant-based diets on the body and the brain: a systematic review. *Transl. Psychiatry* 9, 226 (2019).
10. Krishnan, S., Alden, N. & Lee, K. Pathways and functions of gut microbiota metabolism impacting host physiology. *Curr. Opin. Biotechnol.* 36, 137–145 (2015).
11. Moreno-del Castillo, M. C., Valladares-García, J. & Halabe-Cherem, J. Microbioma humano. *Rev. Fac. Med. Univ. Nac. Autónoma México* 61, 7–19 (2018).
12. Beam, A., Clinger, E. & Hao, L. Effect of Diet and Dietary Components on the Composition of the Gut Microbiota. *Nutrients* 13, 2795 (2021).
13. David, L. A. et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 505, 559–563 (2014).
14. Hills, R. D. et al. Gut Microbiome: Profound Implications for Diet and Disease. *Nutrients* 11, 1613 (2019).
15. Laudadio, I. et al. Quantitative Assessment of Shotgun Metagenomics and 16S rDNA Amplicon Sequencing in the Study of Human Gut Microbiome. *Omic J. Integr. Biol.* 22, 248–254 (2018).
16. Ranjan, R., Rani, A., Metwally, A., McGee, H. S. & Perkins, D. L. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem. Biophys. Res. Commun.* 469, 967–977 (2016).
17. Kuczynski, J. et al. Experimental and analytical tools for studying the human microbiome. *Nat. Rev. Genet.* 13, 47–58 (2011).
18. Peterson, D. et al. Comparative Analysis of 16S rRNA Gene and Metagenome Sequencing in Pediatric Gut Microbiomes. *Front. Microbiol.* 12, (2021).
19. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* 17, 10–12 (2011).
20. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinforma. Oxf. Engl.* 34, (2018).
21. Amir, A. et al. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *mSystems* 2, e00191-16 (2017).
22. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596 (2013).
23. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4, e2584 (2016).
24. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma. Oxf. Engl.* 30, 2114–2120 (2014).

25. Blanco-Míguez, A. et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat. Biotechnol.* 41, 1633–1644 (2023).
26. Beghini, F. et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife* 10, e65088 (2021).
27. SciPy -. <https://scipy.org/>.
28. pandas - Python Data Analysis Library. <https://pandas.pydata.org/>.
29. Matplotlib — Visualization with Python. <https://matplotlib.org/>.
30. Waskom, M. seaborn: statistical data visualization. *J. Open Source Softw.* 6, 3021 (2021).
31. Durazzi, F. et al. Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. *Sci. Rep.* 11, 3030 (2021).
32. Campanaro, S., Treu, L., Kougias, P., Zhu, X. & Angelidaki, I. Taxonomy of anaerobic digestion microbiome reveals biases associated with the applied high throughput sequencing strategies. *Sci. Rep.* 8, (2018).
33. Bikel, S. et al. Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: towards a systems-level understanding of human microbiome. *Comput. Struct. Biotechnol. J.* 13, 390–401 (2015).
34. Pollock, J., Glendinning, L., Wisedchanwet, T. & Watson, M. The Madness of Microbiome: Attempting To Find Consensus “Best Practice” for 16S Microbiome Studies. *Appl. Environ. Microbiol.* 84, e02627-17 (2018).
35. Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J. & Segata, N. Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* 35, 833–844 (2017).
36. Blake, P., Durão, S., Naude, C. E. & Bero, L. An analysis of methods used to synthesize evidence and grade recommendations in food-based dietary guidelines. *Nutr. Rev.* 76, 290–300 (2018).
37. Kaczmarek, J. L. et al. Broccoli consumption affects the human gastrointestinal microbiota. *J. Nutr. Biochem.* 63, 27–34 (2019).