



Article

Development of MetaXplore: An Interactive Tool for Targeted Metagenomic Analysis

Naima Bel Mokhtar ¹, Elias Asimakis ¹, Ioannis Galiatsatos ¹, Amal Maurady ², Panagiota Stathopoulou ¹
and George Tsiamis ^{1,*}

¹ Laboratory of Systems Microbiology and Applied Genomics, Department of Sustainable Agriculture, University of Patras, 30100 Agrinio, Greece; naima1503@gmail.com (N.B.M.); eliasasim@gmail.com (E.A.); jgalia96@gmail.com (I.G.); panstath@upatras.gr (P.S.)

² Laboratory of Innovative Technologies, National School of Applied Sciences of Tangier, Abdelmalek Essaâdi University, Tanger 93000, Morocco; amal.maurady.ma@gmail.com

* Correspondence: gtsiamis@upatras.gr

Abstract: Over the last decades, the analysis of complex microbial communities by high-throughput sequencing of marker gene amplicons has become routine work for many research groups. However, the main challenges faced by scientists who want to make use of the generated sequencing datasets are the lack of expertise to select a suitable pipeline and the need for bioinformatics or programming skills to apply it. Here, we present MetaXplore, an interactive, user-friendly platform that enables the discovery and visualization of amplicon sequencing data. Currently, it provides a set of well-documented choices for downstream analysis, including alpha and beta diversity analysis, taxonomic composition, differential abundance analysis, identification of the core microbiome within a population, and biomarker analysis. These features are presented in a user-friendly format that facilitates easy customization and the generation of publication-quality graphics. MetaXplore is implemented entirely in the R language using the Shiny framework. It can be easily used locally on any system with R installed, including Windows, Mac OS, and most Linux distributions, or remotely via a web server without bioinformatic expertise. It can also be used as a framework for advanced users who can modify and expand the tool.

Keywords: microbial profiling; shiny framework; amplicon data analysis; data visualization; taxonomic composition; diversity analysis; differential abundance



Citation: Bel Mokhtar, N.; Asimakis, E.; Galiatsatos, I.; Maurady, A.; Stathopoulou, P.; Tsiamis, G. Development of MetaXplore: An Interactive Tool for Targeted Metagenomic Analysis. *Curr. Issues Mol. Biol.* **2024**, *46*, 4803–4814. <https://doi.org/10.3390/cimb46050289>

Academic Editor: Vesselin Baev

Received: 15 April 2024

Revised: 10 May 2024

Accepted: 13 May 2024

Published: 15 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The targeted metagenomic approach has been widely used to investigate the microbiota between different environmental or individual factors (e.g., treatments, geographical locations, diets, etc.). Analyzing the changes in composition and abundance of microbial communities provides valuable insight into ecosystem dynamics, including health, stability, and resilient changes, along with the ability to identify biomarkers that can serve as indicators for monitoring environmental changes, treatment responses, or disease [1]. Targeted sequencing is typically conducted based on conserved regions containing phylogenetically informative polymorphisms, such as the 16S rRNA gene for prokaryotes and the 18S rRNA or ITS genes for eukaryotes, which offer a powerful and economical way to characterize the bacterial community in large numbers of samples with affordable techniques [2–4]. The microbiome analysis pipeline includes mainly the clustering of sequences into OTUs or ASVs (Operational Taxonomic Units or Amplicon Sequence Variants) followed by taxonomic classification of the representative sequences. For instance, OTUs are typically clustered based on a sequence similarity threshold, often ranging from 97% to 99% [5]. Sequences belonging to the same OTUs represent a putative taxonomic unit at a specific taxonomic level (e.g., genus or species). Key discriminatory criteria include sequence similarity, which determines OTU membership using algorithms such as UPRASE [5] and

CD-HIT [6], and taxonomic assignment, which assigns taxonomic lineage to OTUs based on reference databases such as Greengenes [7] or SILVA [8]. Clustering could also be influenced by sequencing errors and bias. These criteria are dynamic and evolve with sequencing technologies and bioinformatics tools. Various pipelines are commonly used for these tasks, including Mothur [9], QIIME [10], and USEARCH [11]. However, despite their widespread use, they still have obvious shortcomings, such as a lack of efficient downstream statistical analyses and visualization solutions. The primary downstream analysis includes diversity analysis, taxonomic composition, differential abundance, biomarker identification, and core microbiome selection.

As the demand for comprehensive microbial analyses continues to grow, researchers have had to choose among a range of multipurpose software platforms and a wide range of analysis methods. For non-expert users, this task can be particularly intimidating, requiring a steep learning curve to effectively use the full potential of these tools. Furthermore, the increasing complexity of study designs, often involving multiple experimental conditions, necessitates the use of flexible analysis tools capable of dynamically adjusting analyses and visualizations in real time.

In the current study, we took advantage of the flexibility of R programming and the Shiny package to develop MetaXplore, an interactive, user-friendly platform that enables the discovery and visualization of microbial sequencing data. MetaXplore features an analytical pipeline for the analysis of microbial diversity and composition based on Operational Taxonomic Unit (OTU) tables. It is designed to provide microbiologists who are not skilled in programming with the ability to utilize R functionalities and conduct exploratory analyses of microbial sequencing data within a flexible and interactive GUI. In order to elucidate the functionalities of MetaXplore, we used a set of amplicon data representing the bacteria associated with the *Glossina morsitans morsitans* (*Gmm*) species. The example analysis aims to investigate the impact of a combination of supplements and antibiotics on the gut microbiota of mass-reared *Glossina* samples.

2. Methods and Materials

MetaXplore has an interactive, user-friendly interface written entirely in the R language using the Shiny framework, which is easy to use and does not require any programming skills. Like any Shiny interface, MetaXplore can be further customized or extended using HTML, CSS, and JavaScript. The current version is dependent on previously published packages, including GUnifrac [12] and APE [13] for distance calculations, vegan [14] for ordination methods, and ggplot2 [15] for visualization of results. MetaXplore is configured primarily for microbial ecology profiling using 16S rRNA gene, 18S rRNA and ITS sequencing data and requires four different files as input: (a) OTUs tables, which details the occurrence and abundance of each OTU within each sample, with taxa in rows and samples in columns, (b) a taxonomy file, a table containing the assigned taxonomy for each OTU, (c) Mapping file, containing experimental metadata in columns with a main column labeled “sample ID”, which contains sample labels matching those used in the OTU file, and at least one more column, (d) Tree file, a phylogenetic tree saved in Newick (NWK) format. Various unique or combined such as USEARCH [11], QIIME [10] pipelines can be used to obtain these files. An example pipeline using both of these tools is illustrated in Figure S1.

2.1. Operation

MetaXplore is currently available in a web-based format (<http://MetaXplore.eu>; accessed on 12 May 2024) and can be accessed remotely from any web browser. It can also be launched locally from a computer running Linux, Windows, or MacOS with an R installation (version 4.1.2 or higher) and will run on any web browser. As MetaXplore auto-installs all required R-packages, no additional software requirements are needed. The source code of MetaXplore is available at <https://github.com/nbel15/MetaXplore> (accessed on 12 May 2024). Local installation instructions are also available. All interfaces and plots of

MetaXplore are highly interactive, allowing users to visualize and download data in real time as well as to interact efficiently with the data and plots.

2.2. MetaXplore Sections

The MetaXplore interface provides different downstream analyses, classified in the following sections: (1) Import Data section to upload the dataset at the start of the analysis. (2) Alpha Diversity section to calculate alpha diversity indices among samples and visualize indices in boxplots with significant differences between the groups. (3) Beta Diversity section to calculate the beta diversity between samples and visualize groupings of samples according to a given category. (4) Relative Abundance section to calculate relative abundance matrices and produce heatmaps, stacked or standard bar charts for different taxonomy levels according to a given category. (5) Differential abundance section to test the significant variation in OTU abundance at different taxonomic levels. (6) Core microbiome section to identify the core OTUs based on the percentage of occurrence in samples and their abundance. And (7) Biomarker discovery section to identify the discriminant taxa between groups of samples.

2.2.1. Data Import and Overview

Upon uploading the input data, MetaXplore will verify the format and validity of the input files. The OTU table, metadata, and taxonomy files can be saved either in TXT or TSV format, along with the OTU phylogenetic tree file in NWK format. It will identify the column in the metadata file labeled "sample ID", match its contents with the header row in the OTU table and synchronize both files. The uploaded files are shown in different panels for easy access to the raw data.

2.2.2. Alpha Diversity

Alpha diversity, which refers to the mean diversity of species within a sample or an experimental condition, is considered a ubiquitous approach in microbiome analysis surveys [16]. Different measures exist to estimate diversity. These measures reflect species richness, which is the number of species (OTUs) present, and species evenness, which is the equitability in the distribution of OTUs within a particular sample or a given category [17]. However, species diversity involves both the richness and evenness of individuals. To capture the structure of the community rather than enumerating the parts only, the two most popular indices are used, the Simpson and Shannon indices [18]. The Simpson index represents the probability that two randomly selected samples belong to the same species, which gives more weight to the evenness, meaning that rare species have a minimal impact on community diversity. In contrast, the Shannon index puts more weight on richness. The index is derived from information theory and represents the uncertainty of identifying the species that one randomly selected individual from the community will belong to, i.e., the more species in the community, the higher the uncertainty [19,20]. Another aspect of diversity is the phylogenetic diversity (Faith's Phylogenetic Diversity; PD). PD quantifies the total branch lengths between the observed species on a rooted phylogenetic tree [21]. This metric provides a comparable evolutionary measure of biodiversity that cannot be captured by the species' richness and/or their respective abundance [21–23]. MetaXplore allows users to select experimental categories and alpha diversity parameters simultaneously. Currently, MetaXplore supports several options of alpha diversity indices, including Shannon and Simpson, in addition to their effective values, richness indices such as ACE and Chao1, Evenness, and phylogenetic diversity (PD). The graphical results are presented in boxplots with experimental groups on the x-axis and indices on the y-axis annotated with significance labels (bars with asterisks) above the boxplots. Statistical comparisons of the indices between the groups are calculated using analysis of variance (ANOVA), followed by Tukey's honestly significant difference (HSD).

2.2.3. Beta Diversity

Beta diversity quantifies the distances between different microbial profiles described by the OTU table, which allows linking the overall taxonomic diversity pattern to the experimental features [24,25]. The most common approaches to calculate the similarity of microbial profiles are the Bray–Curtis dissimilarity index and the weighted and unweighted UniFrac distances. Quantitative metrics, including Bray–Curtis and weighted UniFrac, use OTU abundance data in calculations, whereas qualitative metrics, such as unweighted UniFrac, only consider the presence or absence. Unweighted and weighted UniFrac methods are sensitive to rare and dominant OTUs, respectively. To address this issue, a balanced alternative, known as Generalized UniFrac, was proposed [12]. Phylogenetic-based methods, such as UniFrac, exploit the similarities and differences among species, which generally provide an interpretable biological pattern [25,26]. Currently, MetaXplore allows users to choose between Bray–Curtis, weighted and unweighted UniFrac, and generalized UniFrac methods for the calculation of distance matrices. For visualizing the beta diversity analysis results in two-dimensional spaces, MetaXplore provides a variety of ordination methods, including Principal Coordinate Analysis (PCoA), Multi-Dimensional Scaling (MDS), its more robust and unconstrained non-metric version (NMDS), and Canonical Analysis of Principal Coordinate (CAP). The non-parametric permutation test PERMANOVA is used to assess the significant clustering between the groups [27]. In the case where the chosen data category contains more than two experimental groups, a pairwise comparison of the beta diversity metric is conducted, and the results are plotted in a heatmap graph.

2.2.4. Relative Abundance

MetaXplore calculates the relative abundance with standard error of individual taxa within the selected experimental category. Users select a category from the metadata list, the taxonomic level(s) where all the OTUs with the same level will be combined, and the threshold of relative abundance taxon to visualize. The results can be displayed in various plots: stacked bar chart, heatmap, or basic bar chart. Additionally, users can download the average relative abundance along with the standard error table as a CSV file. All the plots can be customized in real-time by widgets, allowing the adjustment of label size, rotation of x-axis labels, graph colors, etc.

2.2.5. Differential Abundance

The differential abundance analysis allows a systematic identification of the significant variation in OTUs between the assessed experimental conditions [28]. The statistical test is performed using nonparametric Kruskal–Wallis Rank Sum tests (for group > 2) followed by Wilcoxon Rank-Sum tests for each paired group as implemented by the *microeco* R package [29]. To avoid comparisons of taxa that may not be relevant in the experimental condition of interest (e.g., taxa with low relative abundance), MetaXplore offers the ability to identify mean abundance thresholds that can be set to filter out taxa with low abundance, as the pre-filtering of the data has been found to enhance the analytical power [30,31].

2.2.6. Core Microbiome

The core microbiomes are typically quantified by the occurrence and abundance of microbial taxa across multiple samples [32]. These metrics are estimated from the standard OTU table. MetaXplore offers the ability to set a minimum abundance threshold, under which the taxon is disqualified from being part of the core microbiota. Then, the remaining taxa are evaluated based on their occurrence across samples using an assigned cutoff value. Overall, the minimum abundance threshold for OTUs to be considered for core membership typically ranges from 0.001% to 4.5%, while the occurrence cutoff ranges from 50% to 100% across different sample categories (e.g., geographical locations, host species, experimental treatments, etc.) [33,34]. Results are then presented in a presence–absence table for each qualified taxon, along with a Venn diagram comparing the different sample categories.

2.2.7. Biomarker Discovery

The biomarker discovery analysis allows for the identification of specific taxa (also known as discriminant taxa) that exhibit characteristic differences between two or more experimental factors. This feature allows for the identification of potential indicators of diseases, different treatments, environmental changes, etc. The identification of discriminant taxa in MetaXplore is based on the LefSe algorithm [35] as implemented by the MicrobiotaProcess R package [36]. The identification is based on two steps: initially, identification of taxa with significant differences in relation to the selected factor using non-parametric Kruskal–Wallis tests. Then, the significant taxa are assessed by Linear Discriminant Analysis (LDA) to estimate the effect size of each taxon. The resulting output consists of a taxonomic tree listing all the taxa along with a bar chart highlighting the discriminative taxa according to the logarithmic LDA score (Effect Size) threshold (2.0 as default).

3. Results

As a use case of MetaXplore, we describe hereafter the main features of the platform and illustrate their utility via a microbiome study examining the dynamic of the bacterial community associated with laboratory-reared *Glossina morsitans morsitans* (*Gmm*) under different diets, with or without antibiotic treatment.

3.1. Dataset

The bacterial profiles were studied using amplicon sequencing of the hypervariable V3–V4 region of the 16S rRNA gene on gastrointestinal tract samples extracted from adult flies that were fed on diets enriched with three different supplements: B vitamins (Bvit), Vanderzant vitamin mixture (Van) and Yeast, with/without treatment with the antibiotic tetracycline (treated samples are labeled T/Bvit, T/Van and T/Yeast, respectively). Sample collection, dissection, DNA extraction, and sequencing were performed as described in [37]. Briefly, adult flies from both treated and untreated diets were surface sterilized with 70% ethanol and sterile PBS. DNA extraction from a pool of five guts for each sample was performed using a modified CTAB protocol [38]. The variable V3–V4 region of the bacterial 16S rRNA sequences was amplified using fusion primers U341F-MiSeq and 805R-MiSeq. High-throughput sequencing was performed by Macrogen using a 2 × 300 bp pair-end kit on a MiSeq platform. Sequencing resulted in a total of 850,135 reads. The NCBI Bioproject accession number for the raw sequencing data reported in this study is PRJNA1088284. Analysis of raw reads was conducted as previously described by using a combination of USEARCH and QIIME pipeline [37].

First, the OTU abundance table, taxonomy, and metadata files were loaded on the “Input Data section”. Uploading the phylogenetic tree is generally optional; it is only required if the user intends to perform a phylogenetically based analysis for alpha diversity (PD analysis) or beta diversity analysis that is based on the UniFrac distance matrix. After hitting the “Read Input” button, MetaXplore creates a phyloseq object [39] displaying the common samples and/or OTUs between the loaded files. This allows the user to reduce the number of samples and OTUs to include in the analysis by modifying only the metadata file and the taxonomy file, respectively.

3.2. Alpha Diversity Changes in Response to Diet

The bacterial community within the untreated *Gmm* samples fed on the supplement had a significantly lower species richness (Chao and ACE) and diversity (Shannon and Simpson) compared to their treated counterpart (Figure 1). Among the treated samples, *Gmm* samples supplemented with yeast showed the highest species richness and diversity compared to those supplemented with B vitamins (Bvit) and Vanderzant vitamin mixture (Van) (Figure 1).

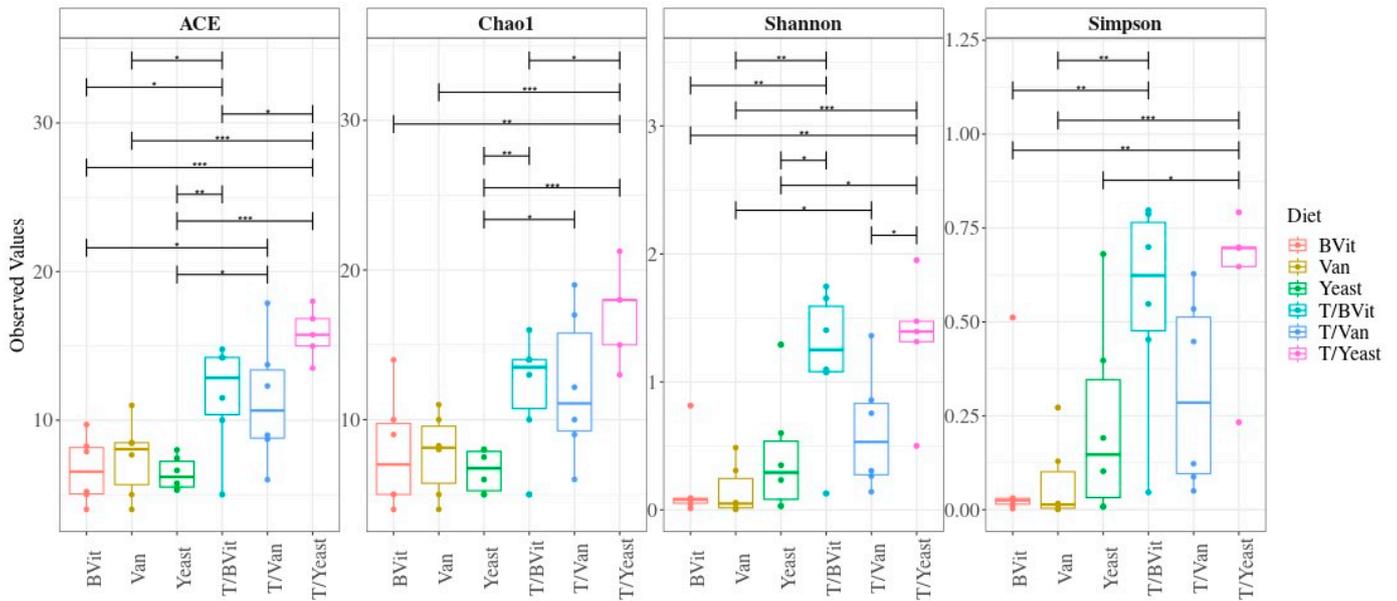


Figure 1. Species richness (ACE and Chao1) and diversity (Shannon and Simpson) within the examined diets. Diet supplemented with B vitamins (Bvit), Vanderzant vitamin mixture (Van), yeast, and their corresponding treated samples T/Bvit, T/Van, and T/Yeast, respectively. * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

3.3. Dynamic in Bacterial Diversity among Diets and Gender

Different bacterial diversity patterns were observed among communities fed on varying diets (Figure 2). Untreated *Gmm* flies fed on diets enriched with supplements form a tight cluster compared to their treated counterparts, which form separate clusters (Figure 2a). The PERMANOVA test indicated that no significant difference was detected between the untreated diets, which indicates that, to some extent, the type of supplement did not alter the bacterial community of the host (p -value > 0.05 ; Figure 2b). In contrast, significantly different bacterial communities were observed between the tetracycline-treated samples fed on different supplements (p -value < 0.05 ; Figure 2b). On the other hand, no significant difference was observed between males and females (p -values = 0.471, Figure S2).

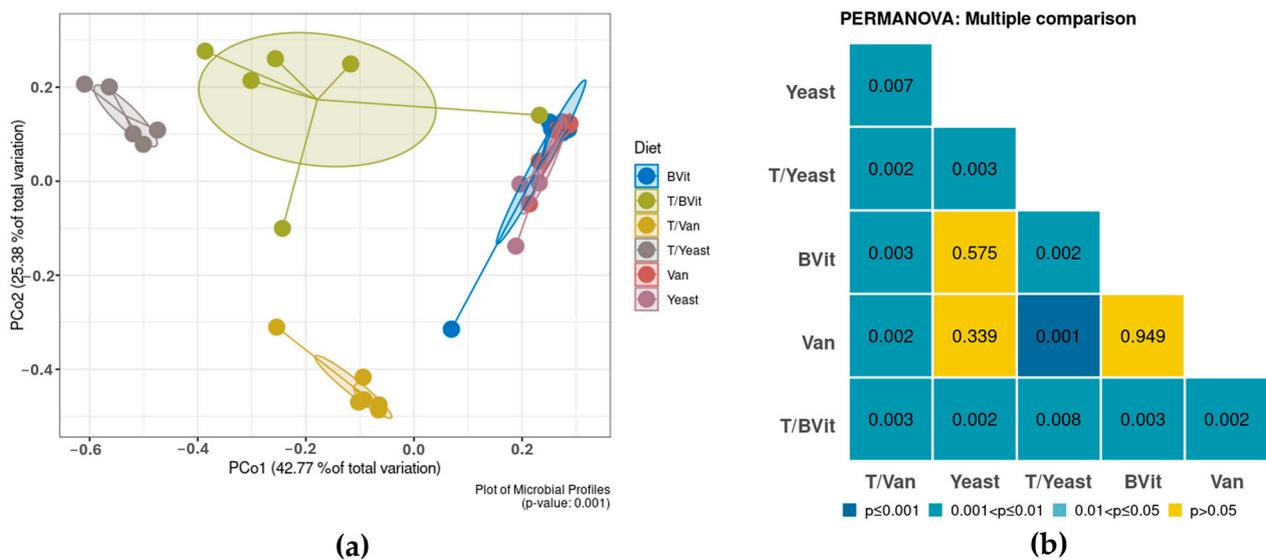


Figure 2. Bacterial diversity between the examined diets based on the GUniFrac dissimilarity matrix. (a) Principal Coordinates Analysis (PCoA). (b) PERMANOVA pairwise comparison between diets.

3.4. Taxonomic Composition in Relation to Rearing Diet

Different bacterial compositions were observed between the treated and untreated samples (Figure 3). At the class level, the guts of untreated *Gmm* samples fed on diets enriched with supplements were dominated by members of Gammaproteobacteria, representing more than 99% of the bacterial communities (Figure 3a). Treated samples supplemented with Vanderzant vitamin mixture maintain the dominance of members of Gammaproteobacteria ($96.4 \pm 2.1\%$) with a slight increase in members of Bacteroidia, Actinobacteria, and Alphaproteobacteria. In the samples supplemented with B vitamins, tetracycline treatment significantly reduced the presence of Gammaproteobacteria ($68 \pm 9.4\%$), leading to a significant increase in members of Bacteroidia and Alphaproteobacteria ($26.4 \pm 6.5\%$ and $5.4 \pm 3.7\%$, respectively) (Figure 3a,b). The samples supplemented with yeast showed the highest influence with the tetracycline treatment, where Gammaproteobacteria significantly decreased to ($8.2 \pm 2.3\%$) along with a significant increase in members of Bacteroidia ($43.9 \pm 11.9\%$), Alphaproteobacteria ($24.7 \pm 8.5\%$), Actinobacteria ($19.6 \pm 6.3\%$), and Bacilli (3.3 ± 2.7) (Figure 3a,b).

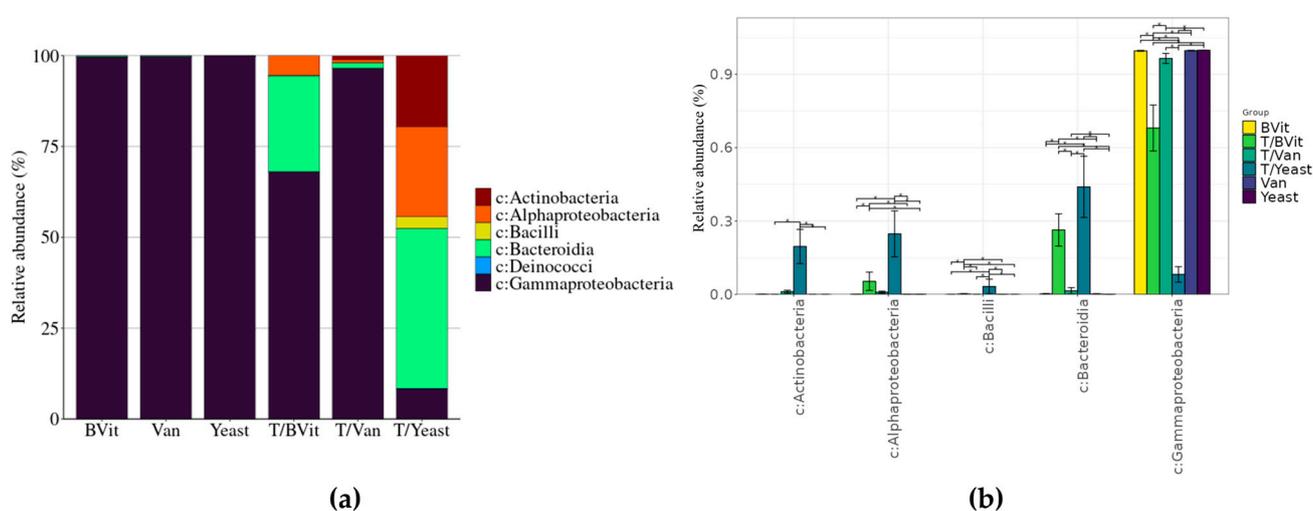


Figure 3. Composition of bacterial community associated with *Gmm* gut according to diets. (a) Relative abundance of taxa at Class level. (b) Differentially abundant classes. * $p < 0.05$.

At the Genus level, the gut bacterial communities of the untreated samples were mainly dominated by *Sodalis*, representing $88.4 \pm 10.5\%$, $95.9 \pm 2.6\%$, and $82.2 \pm 9.8\%$ in *Gmm* samples fed on a diet enriched with B vitamins, Vanderzant vitamin mixture and yeast, respectively (Figure 4a). Interestingly, *Wigglesworthia*, which is an obligatory symbiont of *Glossina* flies [40–42] was under the detection level. This indicates that the supplements assessed in this study prevent the proliferation of these bacteria. The relative abundance of *Sodalis* was significantly decreased in tetracycline-treated samples regardless of the supplement used (Figure 4a,b). On the other hand, *Empedobacter* was significantly increased in the treated samples that were fed on a diet supplemented with B vitamins and yeast ($16.2 \pm 6.9\%$ and $26.6 \pm 7.7\%$, respectively). Similarly, Vanderzant vitamin mixture supplements in the treated samples enhanced the presence of *Comamonas* ($64.4 \pm 14.2\%$) with the tetracycline treatments (Figure 4a,b).

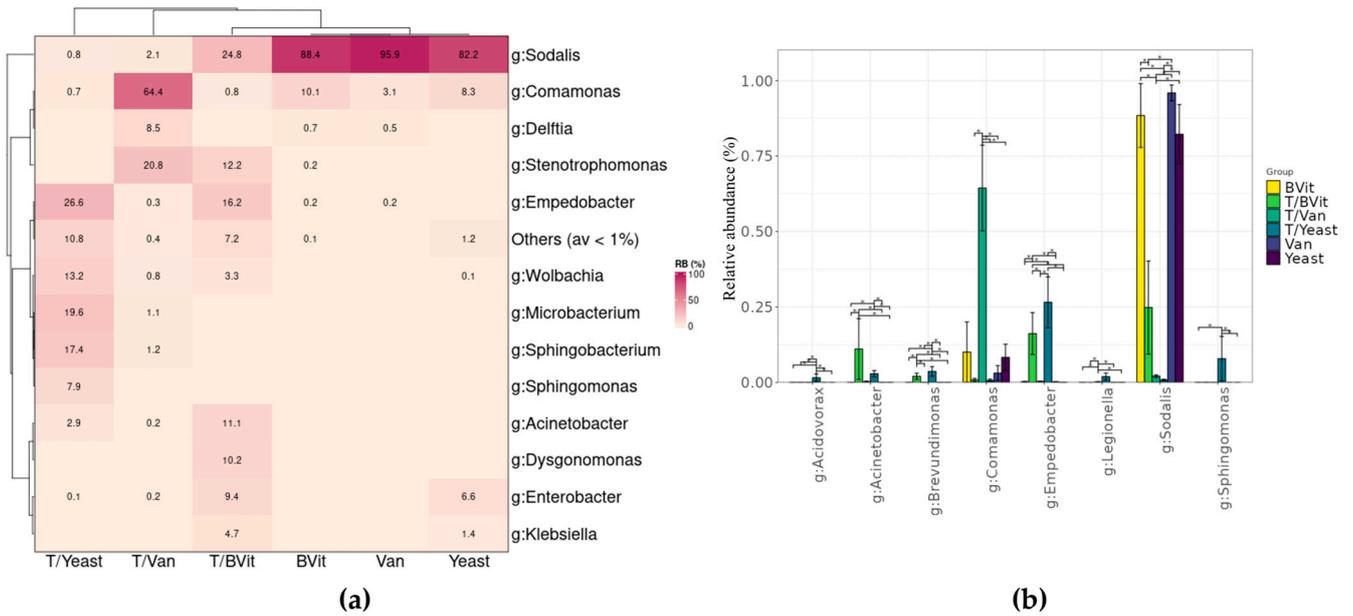


Figure 4. Composition of bacterial community associated with *Gmm* gut according to diets. (a) Relative abundance of taxa at the Genus level. (b) Differentially abundance genera. * $p < 0.05$.

3.5. Core Microbiome among Treated and Untreated Samples

At 75% prevalence and 0.01% abundance thresholds, a total of three OTUs were identified as the core bacteriome of *Gmm* samples (*Sodalis*, *Empedobacter*, and *Wolbachia*), together representing around 60% of the total bacterial community (Figure 5 and Table 1). *Comamonas* was detected as a core microbiome member in all the tetracycline-treated samples, while *Acinetobacter*, *Bacillus*, and *Brevundimonas* were detected in treated samples fed on a diet enriched with B vitamins and yeast supplement. Six core OTUs were detected in treated samples with yeast supplements that are not shared with the other diets, and these include *Microbacterium*, *Sphingomonas*, *Acidovorax*, *Cupriavidus*, *Acinetobacter*, and *Meiothermus* (Figure 5).

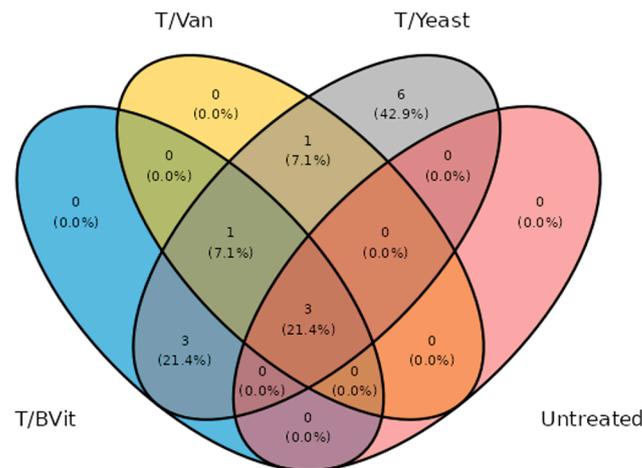


Figure 5. Venn diagram of the distribution of core taxa determined in *Gmm* gut under the assessed diets, with a threshold set at 75%.

Table 1. Core bacterial taxa determined in *Gmm* gut under the assessed diets, with a threshold set at 75%.

| OTUs | Phylum | Class | Genus | T/BVit | T/Van | T/Yeast | Untreated |
|--------|--------------------|-----------------------|-------------------------|--------|-------|---------|-----------|
| Otu1 | p:Proteobacteria | c:Gammaproteobacteria | <i>g:Sodalis</i> | 1 | 1 | 1 | 1 |
| Otu2 | p:Bacteroidota | c:Bacteroidia | <i>g:Empedobacter</i> | 1 | 1 | 1 | 1 |
| Otu6 | p:Proteobacteria | c:Alphaproteobacteria | <i>g:Wolbachia</i> | 1 | 1 | 1 | 1 |
| Otu7 | p:Proteobacteria | c:Gammaproteobacteria | <i>g:Comamonas</i> | 1 | 1 | 1 | 0 |
| Otu11 | p:Proteobacteria | c:Gammaproteobacteria | <i>g:Acinetobacter</i> | 1 | 0 | 1 | 0 |
| Otu15 | p:Firmicutes | c:Bacilli | <i>g:Bacillus</i> | 1 | 0 | 1 | 0 |
| Otu3 | p:Proteobacteria | c:Alphaproteobacteria | <i>g:Brevundimonas</i> | 1 | 0 | 1 | 0 |
| Otu4 | p:Proteobacteria | c:Gammaproteobacteria | <i>g:Legionella</i> | 0 | 1 | 1 | 0 |
| Otu19 | p:Actinobacteriota | c:Actinobacteria | <i>g:Microbacterium</i> | 0 | 0 | 1 | 0 |
| Otu22 | p:Proteobacteria | c:Alphaproteobacteria | <i>g:Sphingomonas</i> | 0 | 0 | 1 | 0 |
| Otu23 | p:Proteobacteria | c:Gammaproteobacteria | <i>g:Acidovorax</i> | 0 | 0 | 1 | 0 |
| Otu26 | p:Proteobacteria | c:Gammaproteobacteria | <i>g:Cupriavidus</i> | 0 | 0 | 1 | 0 |
| Otu27 | p:Deinococcota | c:Deinococci | <i>g:Meiothermus</i> | 0 | 0 | 1 | 0 |
| Otu991 | p:Proteobacteria | c:Gammaproteobacteria | <i>g:Acinetobacter</i> | 0 | 0 | 1 | 0 |

3.6. Discriminant Taxa between Treated and Untreated Samples

Based on the LEfSe approach, the untreated samples were significantly characterized by the *Sodalis* genus, treated samples fed on Vanderzant vitamin mixture with *Comamonas*, and those fed on yeast were characterized by various genera, including *Sphingomonas*. In contrast, no discriminative taxon was found for the treated samples supplemented with B vitamins (Figure 6).

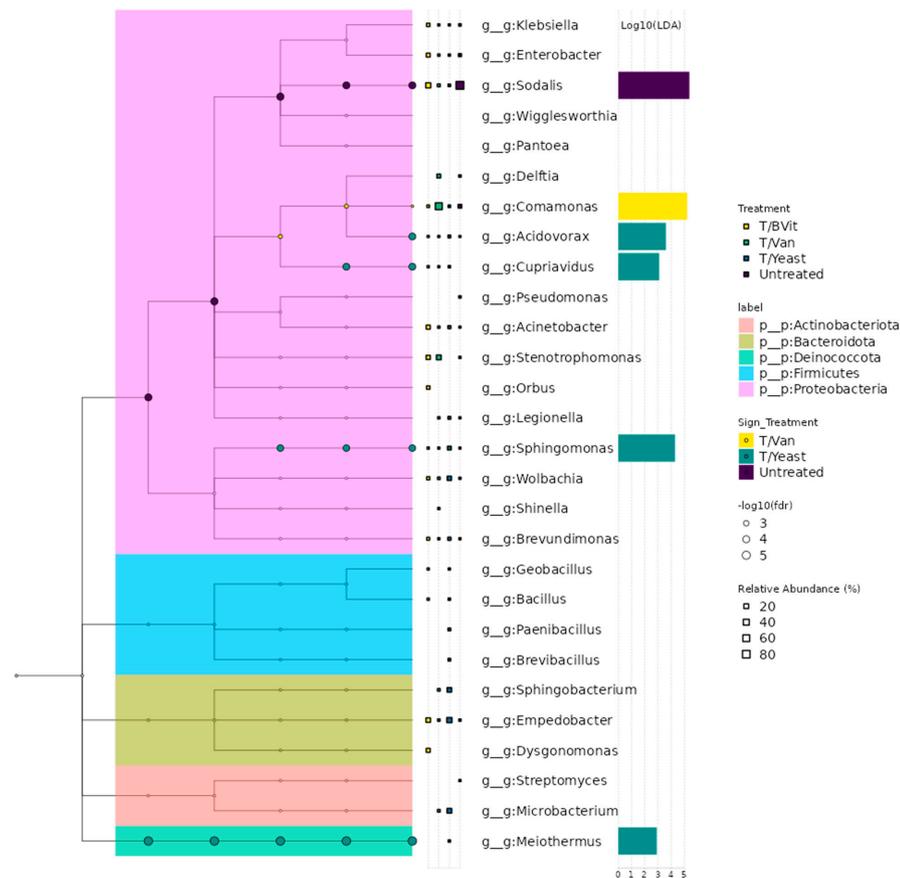


Figure 6. Discriminant taxa between treated and untreated diets at the Genus level. The tree represents the phylogenetic relationships of the identified genera. The horizontal bars represent the logarithmic LDA score of the discriminant genera.

4. Discussion and Conclusions

The generation of amplicon sequencing data to explore microbial communities has become routine in numerous research laboratories, using technologies such as Illumina and Oxford Nanopore. Research institutes and scientists are consistently focusing on the adoption of open-source software for the analysis of these data. R programming environment stands out as the most used tool, offering a free environment for statistical computing, flexible and continuously supported by a large community of active users, developers, and researchers. The development of the shiny and shin-dashboard packages facilitated the development of web applications within the R framework. Consequently, MetaXplore has been implemented using both tools along with functions from various R packages and statistical programs. MetaXplore was created to provide dynamic experience for rapid data analysis of complex microbial communities generated by high-throughput sequencing of community markers such as the 16S and 18 rRNA genes or the ITS region. The current release provides the fundamental downstream analyses required for publications and is rooted in the most up-to-date methods used in the field of microbial ecology. The use of the R language for the development of the entire platform makes MetaXplore a suitable framework for advanced users who can modify and expand the tool.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/cimb46050289/s1>. Figure S1: Exemplar pipeline to generate MetaXplore input files, Figure S2: Principal Coordinates Analysis (PCoA) of bacterial diversity between males and females.

Author Contributions: Conceptualization, G.T.; methodology, N.B.M. and G.T.; software, N.B.M.; formal analysis, N.B.M.; visualization, N.B.M.; resources, A.M., P.S. and G.T.; data curation, N.B.M. and G.T.; writing—original draft preparation, N.B.M., E.A., I.G. and P.S.; writing—review and editing, N.B.M., E.A., I.G., A.M., P.S. and G.T.; supervision, G.T. All authors have read and agreed to the published version of the manuscript.

Funding: The research was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the “First Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant” (Project Number: 423).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: MetaXplore is available at <http://metaxplore.eu> website (accessed on 12 May 2024). The complete source code is available at <https://github.com/nbel15/MetaXplore> (accessed on 12 May 2024). Raw sequencing data reported in this study are available under the NCBI accession number PRJNA1088284.

Acknowledgments: The authors are thankful to the Institute of Zoology, Section of Molecular and Applied Zoology, Slovak Academy of Science, Bratislava, Slovakia, for providing the *Glossina* samples used in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Konopka, A.; Lindemann, S.; Fredrickson, J. Dynamics in Microbial Communities: Unraveling Mechanisms to Identify Principles. *ISME J.* **2015**, *9*, 1488–1495. [[CrossRef](#)] [[PubMed](#)]
2. Sinclair, L.; Osman, O.A.; Bertilsson, S.; Eiler, A. Microbial Community Composition and Diversity via 16S rRNA Gene Amplicons: Evaluating the Illumina Platform. *PLoS ONE* **2015**, *10*, e0116955. [[CrossRef](#)]
3. Matsuo, Y.; Komiya, S.; Yasumizu, Y.; Yasuoka, Y.; Mizushima, K.; Takagi, T.; Kryukov, K.; Fukuda, A.; Morimoto, Y.; Naito, Y.; et al. Full-Length 16S rRNA Gene Amplicon Analysis of Human Gut Microbiota Using MinION™ Nanopore Sequencing Confers Species-Level Resolution. *BMC Microbiol.* **2021**, *21*, 35. [[CrossRef](#)]
4. Kracmarova, M.; Karpiskova, J.; Uhlik, O.; Strejcek, M.; Szakova, J.; Balik, J.; Demnerova, K.; Stiborova, H. Microbial Communities in Soils and Endosphere of *Solanum tuberosum* L. and Their Response to Long-Term Fertilization. *Microorganisms* **2020**, *8*, 1377. [[CrossRef](#)] [[PubMed](#)]

5. Edgar, R.C. UPARSE: Highly Accurate OTU Sequences from Microbial Amplicon Reads. *Nat. Methods* **2013**, *10*, 996–998. [[CrossRef](#)]
6. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data. *Bioinformatics* **2012**, *28*, 3150–3152. [[CrossRef](#)]
7. DeSantis, T.Z.; Hugenholtz, P.; Larsen, N.; Rojas, M.; Brodie, E.L.; Keller, K.; Huber, T.; Dalevi, D.; Hu, P.; Andersen, G.L. Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Appl. Environ. Microbiol.* **2006**, *72*, 5069–5072. [[CrossRef](#)] [[PubMed](#)]
8. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F.O. The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools. *Nucleic Acids Res.* **2013**, *41*, D590–D596. [[CrossRef](#)] [[PubMed](#)]
9. Schloss, P.D.; Westcott, S.L.; Ryabin, T.; Hall, J.R.; Hartmann, M.; Hollister, E.B.; Lesniewski, R.A.; Oakley, B.B.; Parks, D.H.; Robinson, C.J.; et al. Introducing Mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl. Environ. Microbiol.* **2009**, *75*, 7537–7541. [[CrossRef](#)]
10. Bolyen, E.; Rideout, J.R.; Dillon, M.R.; Bokulich, N.A.; Abnet, C.C.; Al-Ghalith, G.A.; Alexander, H.; Alm, E.J.; Arumugam, M.; Asnicar, F. Reproducible, Interactive, Scalable and Extensible Microbiome Data Science Using QIIME 2. *Nat. Biotechnol.* **2019**, *37*, 852–857. [[CrossRef](#)]
11. Edgar, R. *Usearch*; Lawrence Berkeley National Lab (LBNL): Berkeley, CA, USA, 2010.
12. Chen, J.; Zhang, X.; Zhou, H. GUniFrac: Generalized UniFrac Instances, Distance-Based Multivariate Methods and Feature-Based Univariate Methods for Microbiome Data Analysis. 2023. Available online: <https://CRAN.R-project.org/package=GUniFrac> (accessed on 12 May 2024).
13. Paradis, E.; Claude, J.; Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R Language. *Bioinformatics* **2004**, *20*, 289–290. [[CrossRef](#)] [[PubMed](#)]
14. Dixon, P. VEGAN, A Package of R Functions for Community Ecology. *J. Veg. Sci.* **2003**, *14*, 927–930. [[CrossRef](#)]
15. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016; ISBN 978-3-319-24277-4.
16. Whittaker, R.H. Evolution and Measurement of Species Diversity. *TAXON* **1972**, *21*, 213–251. [[CrossRef](#)]
17. Bunge, J.; Willis, A.; Walsh, F. Estimating the Number of Species in Microbial Diversity Studies. *Annu. Rev. Stat. Its Appl.* **2014**, *1*, 427–445. [[CrossRef](#)]
18. Magurran, A.E.; McGill, B.J.; Magurran, A.E.; McGill, B.J. (Eds.) *Biological Diversity: Frontiers in Measurement and Assessment*; Oxford University Press: Oxford, NY, USA, 2010; ISBN 978-0-19-958067-5.
19. Herrmann, B.; Cerbule, K.; Brčić, J.; Grimaldo, E.; Geoffroy, M.; Daase, M.; Berge, J. Accounting for Uncertainties in Biodiversity Estimations: A New Methodology and Its Application to the Mesopelagic Sound Scattering Layer of the High Arctic. *Front. Ecol. Evol.* **2022**, *10*, 775759. [[CrossRef](#)]
20. Hurlbert, S.H. The Nonconcept of Species Diversity: A Critique and Alternative Parameters. *Ecology* **1971**, *52*, 577–586. [[CrossRef](#)] [[PubMed](#)]
21. Grover, S.; Markin, A.; Anderson, T.K.; Eulenstein, O. Phylogenetic Diversity Statistics for All Clades in a Phylogeny. *Bioinformatics* **2023**, *39*, i177–i184. [[CrossRef](#)] [[PubMed](#)]
22. Miller, J.T.; Jolley-Rogers, G.; Mishler, B.D.; Thornhill, A.H. Phylogenetic Diversity Is a Better Measure of Biodiversity than Taxon Counting. *J. Syst. Evol.* **2018**, *56*, 663–667. [[CrossRef](#)]
23. Lozupone, C.A.; Knight, R. Species Divergence and the Measurement of Microbial Diversity. *FEMS Microbiol. Rev.* **2008**, *32*, 557. [[CrossRef](#)]
24. Su, X. Elucidating the Beta-Diversity of the Microbiome: From Global Alignment to Local Alignment. *mSystems* **2021**, *6*, e00363-21. [[CrossRef](#)]
25. Knight, R.; Vrbanac, A.; Taylor, B.C.; Aksenov, A.; Callewaert, C.; Debelius, J.; Gonzalez, A.; Kosciolek, T.; McCall, L.-I.; McDonald, D.; et al. Best Practices for Analysing Microbiomes. *Nat. Rev. Microbiol.* **2018**, *16*, 410–422. [[CrossRef](#)]
26. Hamady, M.; Lozupone, C.; Knight, R. Fast UniFrac: Facilitating High-Throughput Phylogenetic Analyses of Microbial Communities Including Analysis of Pyrosequencing and PhyloChip Data. *ISME J.* **2010**, *4*, 17–27. [[CrossRef](#)] [[PubMed](#)]
27. Anderson, M.J.; Walsh, D.C.I. PERMANOVA, ANOSIM, and the Mantel Test in the Face of Heterogeneous Dispersions: What Null Hypothesis Are You Testing? *Ecol. Monogr.* **2013**, *83*, 557–574. [[CrossRef](#)]
28. Nearing, J.T.; Douglas, G.M.; Hayes, M.G.; MacDonald, J.; Desai, D.K.; Allward, N.; Jones, C.M.A.; Wright, R.J.; Dhanani, A.S.; Comeau, A.M.; et al. Microbiome Differential Abundance Methods Produce Different Results across 38 Datasets. *Nat. Commun.* **2022**, *13*, 342. [[CrossRef](#)] [[PubMed](#)]
29. Liu, C.; Cui, Y.; Li, X.; Yao, M. Microeco: An R Package for Data Mining in Microbial Community Ecology. *FEMS Microbiol. Ecol.* **2021**, *97*, fiae255. [[CrossRef](#)] [[PubMed](#)]
30. Bourgon, R.; Gentleman, R.; Huber, W. Independent Filtering Increases Detection Power for High-Throughput Experiments. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 9546–9551. [[CrossRef](#)] [[PubMed](#)]
31. Cao, Q.; Sun, X.; Smirnova, E. Effects of Rare Microbiome Taxa Filtering on Statistical Analysis. *Front. Microbiol.* **2021**, *11*, 607325. [[CrossRef](#)] [[PubMed](#)]
32. Custer, G.F.; Gans, M.; van Diepen, L.T.A.; Dini-Andreote, F.; Buerkle, C.A. Comparative Analysis of Core Microbiome Assignments: Implications for Ecological Synthesis. *mSystems* **2023**, *8*, e01066-22. [[CrossRef](#)] [[PubMed](#)]

33. Leis, M.L.; Costa, M.O. Initial Description of the Core Ocular Surface Microbiome in Dogs: Bacterial Community Diversity and Composition in a Defined Canine Population. *Vet. Ophthalmol.* **2019**, *22*, 337–344. [[CrossRef](#)]
34. Neu, A.T.; Allen, E.E.; Roy, K. Defining and Quantifying the Core Microbiome: Challenges and Prospects. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2104429118. [[CrossRef](#)]
35. Segata, N.; Izard, J.; Waldron, L.; Gevers, D.; Miropolsky, L.; Garrett, W.S.; Huttenhower, C. Metagenomic Biomarker Discovery and Explanation. *Genome Biol.* **2011**, *12*, R60. [[CrossRef](#)] [[PubMed](#)]
36. Xu, S.; Zhan, L.; Tang, W.; Wang, Q.; Dai, Z.; Zhou, L.; Feng, T.; Chen, M.; Wu, T.; Hu, E.; et al. MicrobiotaProcess: A Comprehensive R Package for Deep Mining Microbiome. *The Innovation* **2023**, *4*, 100388. [[CrossRef](#)] [[PubMed](#)]
37. Bel Mokhtar, N.; Catalá-Oltra, M.; Stathopoulou, P.; Asimakis, E.; Remmal, I.; Remmas, N.; Maurady, A.; Britel, M.R.; García de Oteyza, J.; Tsiamis, G.; et al. Dynamics of the Gut Bacteriome During a Laboratory Adaptation Process of the Mediterranean Fruit Fly, *Ceratitis Capitata*. *Front. Microbiol.* **2022**, *13*, 919760. [[CrossRef](#)]
38. Augustinos, A.A.; Santos-Garcia, D.; Dionyssopoulou, E.; Moreira, M.; Papapanagiotou, A.; Scarvelakis, M.; Doudoumis, V.; Ramos, S.; Aguiar, A.F.; Borges, P.A.V.; et al. Detection and Characterization of Wolbachia Infections in Natural Populations of Aphids: Is the Hidden Diversity Fully Unraveled? *PLoS ONE* **2011**, *6*, e28695. [[CrossRef](#)]
39. McMurdie, P.J.; Holmes, S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE* **2013**, *8*, e61217. [[CrossRef](#)] [[PubMed](#)]
40. Doudoumis, V.; Blow, F.; Saridaki, A.; Augustinos, A.; Dyer, N.A.; Goodhead, I.; Solano, P.; Rayaisse, J.-B.; Takac, P.; Mekonnen, S.; et al. Challenging the Wigglesworthia, Sodalis, Wolbachia Symbiosis Dogma in Tsetse Flies: Spiroplasma Is Present in Both Laboratory and Natural Populations. *Sci. Rep.* **2017**, *7*, 4699. [[CrossRef](#)] [[PubMed](#)]
41. Weiss, B.L.; Rio, R.V.M.; Aksoy, S. Microbe Profile: Wigglesworthia Glossinidia: The Tsetse Fly's Significant Other. *Microbiology* **2022**, *168*, 001242. [[CrossRef](#)]
42. Doudoumis, V.; Augustinos, A.; Saridaki, A.; Parker, A.; Abd-Alla, A.M.M.; Bourtzis, K.; Tsiamis, G. Different Laboratory Populations Similar Bacterial Profile? The Case of Glossina Palpalis Gambiense. *BMC Microbiol.* **2018**, *18*, 148. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.