

## **AT YOUR FINGERTIPS: RAPID ACCESS TO METATAXONOMIC DATA ON THE COMPOSITION OF FOOD BACTERIAL COMMUNITIES WITH THE FOODMICROBIONET DATABASE**

**Main author:** Eugenio Parente (University of Basilicata - UNIBAS)

**Co-authors:** Teresa Zotta, Annamaria Ricciardi

### **INTRODUCTION**

Since the publication in 2009 of the first paper using pyrosequencing of the 16S rRNA gene for the characterisation of pearl millet fermentation (<https://doi.org/10.1128/aem.00451-09>), amplicon targeted high-throughput sequencing (AT-HTS) has become the standard in culture-independent studies of food microbial communities. To date, more than 500 studies using metataxonomic approaches targeting the 16S rRNA gene for prokaryotes or, less frequently, 18S rRNA/ITS for eukaryotic communities have been published, and that number is increasing steadily. This wealth of information is dispersed in hundreds of research papers and a limited number of public resources provide direct access to processed information. Three online tools (QIITA, <https://qiita.ucsd.edu>; MGnify <https://www.ebi.ac.uk/metagenomics/>; IMNGS <https://www.imngs.org>) provide, to different degrees, a variety of services using public or private sequence data sets, from bioinformatic analysis to the possibility of searching available studies and combining them in metastudies. Unfortunately, the number of public studies on foods available through these tools is limited and the metadata structure is not optimised.

### **METHODOLOGY**

With the objective of addressing the needs of scientists wishing to access metataxonomic data, we developed the FoodMicrobionet database. In its first iteration (<https://doi.org/10.1016/j.ijfoodmicro.2015.12.001>) it was available as a web visualisation of the bipartite microbe-food network. This offered limited possibilities and soon a different approach based on a relational database structure was developed (<https://doi.org/10.1016/j.ijfoodmicro.2019.108249>). New studies are now added by downloading raw sequences and metadata from the NCBI SRA archive and processing the sequences using a modified version of the Bioconductor DADA2 pipeline ([https://benjjneb.github.io/dada2/tutorial\\_1\\_8.html](https://benjjneb.github.io/dada2/tutorial_1_8.html)). Taxonomic assignment is carried out using the SILVA v138.1 database, Amplicon Sequence Variants are then aggregated at the lowest possible taxonomic level, and sample and study metadata are curated. The database is organised in four related tables, three of which contain metadata (studies, samples, taxa)

and the latter (edges) quantitative information on taxa abundance in individual samples. Access to the database is via an interactive Shiny app developed with R.

## RESULTS

FoodMicrobionet is arguably the largest and best-annotated database of metataxonomic data on food bacterial communities. Version 4 includes (September 2021) 9700 samples from 167 published and unpublished studies, from 15 different food groups. Foods are classified using the FoodEx2 scheme developed by EFSA (<https://www.efsa.europa.eu/en/supporting/pub/en-6507>), integrated with further information on gene targets and sample nature. Smooth integration with other databases (NCBI SRA for samples and studies; publications via DOI for studies; LPSN, NCBI taxonomy and Florilege for taxa) is provided by automatically generated links. Experienced users can carry out complex searches and data processing using the raw data tables, but an interactive app facilitates searches on studies and samples, optional aggregation of samples (taking advantage of the hierarchical structure of the FoodEx2 classification) and taxa, data export in a variety of formats, and limited numerical and graphical analyses.

## DISCUSSION

FoodMicrobionet differs in some notable respects from other online tools. While platforms like QIITA, MGnify or IMNGS always provide state of the art bioinformatic pipelines, they do have some shortcomings. Both QIITA and IMNGS may require significant process time to generate an output. MGnify provides access to processed data but does not allow merging of studies. None of these tools uses a metadata structure comparable to the FoodEx2 classification used in FoodMicrobionet, which, in addition, provides the possibility of aggregating samples and taxa, and provides a smooth integration with other databases or sources of information. In addition, the ShinyFMBN app offers several different export formats, which can be directly used for further processing in graph visualisation tools, for graphical and statistical processing and for inference of microbial association networks.

In conclusion, FoodMicrobionet is a flexible tool for rapid access to metataxonomic data, which can be used both by scientists for metastudies and reviews, and by the industry to rapidly access information for product and process development, as well as by regulatory agencies for risk assessment.