

Microbiota plays a crucial role in human health. Recently, NGS technologies have enabled the exploration of the microbiome without isolation and culturing. However, analyzing and translating microbiome data into meaningful biological insights is challenging due to the data's compositional nature, high dimensionality, sparseness, and over-dispersion. The gut microbiome can vary from individual to individual, and microbiome communities can be grouped to identify community types linked to environmental or health conditions. Different data features, such as individual profiles, community-based descriptors, or genera interactions within a community, provide different perspectives on microbiome complexity.

The goal of this research project, which is part of the P6 Heal Italia Partnership of PNRR, is to study methodological solutions aimed at addressing microbiome classification according to the different perspectives. The available methods fail in many cases to produce meaningful results because they are unable to deal with data with such a high dimensionality. We will mainly focus on methods that approach classification according to a model-based perspective and will include latent variables to allow for dependence between the abundances of the different genera. The research project involves:

- literature review
- methodological developments
- code development
- real data analysis.

Attention will be also paid to variable selection issues (which are complicated by the compositional nature of microbiome data) and to simulation strategies for microbiome data.

The project will be carried out in collaboration with the Oslo Center for Biostatistics and Epidemiology and with researchers from the Department of Pharmacy and Biotechnology of Unibo.