**BIO5202 Analysis Tutorial Overview**

Changes in the gut microbiome composition and function have been connected to various factors, including antibiotics, diet, pollutants, among others (Peeters et al., 2021, Dudek-Wicher et al., 2018). These changes have also been linked to numerous inflammatory and metabolic diseases such as inflammatory bowel diseases (IBD), obesity, cancer, and others (Szilagyi, 2020, García-Mazcorro et al., 2017). Understanding changes in microbiome composition which mostly influence functional dysbiosis is an important aspect of microbial ecology. Usually, in a gut sample, we would want to understand if there are variations in the relative composition and abundance of the microbial groups as well how dissimilar they are. In Microbial ecology, heatmaps are a powerful graphical representation tool that has been widely used to visualize complex datasets (Perrone et al., 2022, Xu et al., 2017). They enable researchers to examine the presence, absence, and abundance of various microbial taxa across different samples or conditions, using color gradients to represent numerical values. This visualization approach is particularly useful for highlighting patterns, such as similarities and differences between microbial communities, which can be essential for understanding ecological dynamics and microbial interactions necessary to infer functions (Williams et al., 2019). This analysis tutorial titled “Using R code to create a heatmap for the visualization of microbiome composition of samples” is intended to help researchers through the process of analyzing microbiome data using R (R Core Team 2024). It featured interesting data transformation techniques and visualization methods to better understand microbial diversity and abundance across samples. I used several specialized R packages such as tidyverse (a collection R packages including ggplot, dplr, etc.), RcolorBrewer (for color schemes), vegan (for ecological diversity analysis) and devtools (that makes package development easier) (Oksanen et al., 2022; A[lboukadel,](https://www.datanovia.com/en/blog/author/kassambara/) 2019). For data handling, I loaded data into R directly from my GitHub repository using a URL. I prepared this data by previewing it and removing the unnecessary column (i.e. the index column) to focus solely on the sequence count data. The raw counts were transformed into microbial proportions within each sample to account for varying total counts across samples (normalization). Next, I created the first heatmap that visualized the transformed microbiome data (proportions) using the “heatmap” function with a color palette ranging from light yellow to magenta. I then calculated the maximum relative abundance based off of the proportions and filtered out taxa with a minimum relative abundance of less than 1% to focus on more prevalent taxa (genera or class). I generated the second heatmap using the filtered data, ensuring better clarity and focus on significant taxa. After this, I then implemented hierarchical clustering of samples using the Bray-Curtis dissimilarity and the average linkage method to explore sample similarities. I also carried out ‘genera’ clustering based on their co-occurrence across samples, using a transposed data matrix. The final heatmap which combined row and column dendrograms with the heatmap provided a comprehensive view of both the sample and genera clustering, revealing the bacteria composition, abundance and relationships essential for function inference.