16S rRNA Gene Sequencing and Microbiome Analysis

Raw 16S rRNA gene sequencing data were processed using the **DADA2** pipeline (v1.30) in **R**. Only forward reads were retained for downstream analysis. Primer sequences and low-quality regions were removed by truncating reads at 250 bp and trimming 20 bases from the 5′ end. Reads exceeding a maximum expected error threshold of 2 (maxEE = 2) were discarded. Following quality filtering, reads were dereplicated, and chimeric sequences were removed. Amplicon sequence variants (ASVs) were inferred and taxonomically classified using the **SILVA** reference database (release 138.2).

ASVs were aligned using **MAFFT** (v7.149), and a phylogenetic tree was constructed with **FastTree** (v2.1). The resulting ASV table and associated metadata were integrated into a **phyloseq** object (v1.46.0). Non-bacterial sequences and those annotated as Chloroplast or Mitochondria were removed. ASVs present in fewer than 25% of samples were also excluded to reduce sparsity.

To account for differences in sequencing depth, the ASV table was rarefied to an even sampling depth of 12,487 reads per sample. Relative abundances were computed by normalizing ASV counts to the total reads per sample.

For microbiome community composition analysis, the relative abundance profiles of microbial families in Fecal Buffer Blanks at time 0 (FBB0) and 16 hours (FBB16) were visualized using **ggplot2** (v3.5.2). Community dissimilarity was assessed using four distance metrics: Bary-Curtis, Jaccard, Unweighted UniFrac, and Weighted UniFrac. Principal coordinate analysis (PCoA) was performed for each distance method, and the first two components were used to compute centroid positions for each category within a given microbiome. FBB0 coordinates served as reference starting points in visualizations depicting the dissimilarity trajectories across conditions.

Boxplots were generated to show the distribution of PCoA distances between category levels and the FBB0 baseline within each microbiome and at the individual sample level. All visualizations were produced using **ggplot2** (v3.5.2).

To assess statistical differences in microbial community structure across sample and category groups, **Analysis of Molecular Variance (AMOVA)** was conducted on the Weighted UniFrac distance matrix using the **vegan** package (v2.6.4). Dose-response relationships were evaluated by correlating sample dosage levels with PCoA distances from FBB0 centroids, based on Weighted UniFrac dissimilarity. The resulting associations were visualized per sample and microbiome.

Permutational Multivariate Analysis of Variance (PERMANOVA) based on Bray-Curtis distance matrices was used to assess genus-level compositional differences across sample groups. P-values were adjusted using the Benjamini-Hochberg procedure to control the false discovery rate. Genera exhibiting significant variation between groups (adjusted *p* < 0.05) were retained for downstream analysis. The relative abundances of these significant taxa were log₁₀-transformed, and fold changes were visualized using heatmaps, stratified by treatment group and individual microbiome.

A subset of health-relevant taxa, including *Bacillus, Bifidobacterium, Clostridium, Faecalibacterium, Fusobacterium, Lachnospiracease UCG-010, and the family Enterobacteriaceae,* were selected for dose-response analysis. Relative abundance changes for these taxa were plotted against sample dosage using ggplot2 (v3.5.2).

To explore correlation patterns among taxa and treatments, **parallel coordinate plots** (via GGally, ggplot2, RColorBrewer) and **trellis plots** were employed. Based on the criteria described in Korth et al. (2024) [<https://doi.org/10.1093/g3journal/jkae145>], taxa were further classified as health-promoting or potentially harmful. These classifications were then used to compute the **PPIa index**, a composite metric of gut microbiome health.