📊 **MaAsLin2 Analysis: Aerobic vs Anaerobic Microbiome Samples**

This R script performs performs a microbiome differential abundance analysis using **MaAsLin2** to compare aerobic, anaerobic, and uncultured caecal microbiome samples from D7 HMA mice. The analysis filters for species-level taxa, applies **log transformation**, and outputs visual and tabular summaries of significant associations.

📁 **Directory Structure**

Ensure your working directory includes the following input files:

* **merged\_abundance\_table.txt** — OTU abundance table
  + exported from the tool MetaPhlAn
  + taxonomic labels follow GTDB-style formatting (e.g., d\_\_Bacteria;p\_\_Firmicutes;...)
  + taxonomic lineages should be annotated in clade\_name
  + tab-delimited
* **AerobevsAnaerobe\_scraping\_metadata.txt** — Sample metadata file
  + create it based on Supplementary Table 17
  + include grouping variables such as annot1: experimental group (WT/ D2 HMA/ D7 HMA) and annot2: condition (uncultured/ aerobic/ anaerobic)
* Update the setwd() path in the script as needed.

**🔧 Requirements**

Install the following R packages if not already installed:

install.packages(c("BiocManager", "vegan", "devtools"))

BiocManager::install("Maaslin2")

devtools::install\_github("leylabmpi/metagMisc")

🚀 **How to Run**

* Set your working directory inside the R script to the folder where the data files are located: setwd("/path/to/your/project\_directory")
* Run the script AervsAnaer\_maaslin2.R using your preferred R environment.
* MaAsLin2 results will be saved in the specified output directory.

**🧬 Overview of Analysis**

**Step 1: Load Libraries and Data**

* Loads necessary R packages for microbiome analysis and metadata handling.
* Loads the OTU table and metadata file.
* Filters OTU table to include only species-level taxonomic entries (s\_\_, excluding t\_\_).
* Filters metadata to exclude other experimental groups (e.g., "WT" and "D2 HMA").

**Step 2: Configure and Run MaAsLin2**

* Compares groups based on the annot2 variable (aerobic, anaerobic, uncultured).
* Sets "Uncultured" as the reference level.
* Applies:
  + **No normalization**
  + **Log transformation**
  + **Benjamini-Hochberg** multiple testing correction
* Filters features by:
  + Minimum prevalence of **1%**
  + Maximum adjusted p-value of **0.05**

**Step 3: Output and Visualization**

* Outputs results to the folder ‘maaslin2\_AervsAnaer\_D7HMA\_uncultured’

**📈 Output**

* Significant associations (significant\_results.tsv) corresponding to Figure 6C or Supplementary Figure 2D and 6D in the manuscript.
* All model results (all\_results.tsv)
* Heatmap and scatter plots of top taxa

**⚠️ Notes**

* The final analysis includes only **D7 HMA** samples.
  + To analyse other experimental groups (e.g., **WT or D2 HMA**),adjust **lines 29–30** in the script.