📊 **Heatmap Visualization: Top Differential Species in Aerobic vs Anaerobic Microbiome Samples**

This R script visualizes the top 30 most significantly different species (identified by MaAsLin2) across **aerobic**, **anaerobic**, and **uncultured** D7 HMA mouse caecal microbiome samples using a **log-transformed heatmap**. It includes family-level annotations and color-coded sample metadata.

📁 **Directory Structure**

Ensure your working directory includes the following input files:

* **significant\_results.tsv** — Output from MaAsLin2
  + located at: maaslin2\_AervsAnaer\_D7HMA\_uncultured/significant\_results.tsv
  + includes a feature column with GTDB-formatted species names
* **merged\_abundance\_table.txt** — OTU abundance table
  + exported from MetaPhlAn
  + taxonomic labels follow GTDB-style formatting (e.g., d\_\_Bacteria;p\_\_Firmicutes;...)
  + species taxonomy must be in the clade\_name column
  + 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("pheatmap", "stringr", "ggplot2", "grid"))

🚀 **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\_pheatmap.R using your preferred R environment.
* A heatmap will be generated and displayed.

**🧬 Overview of Analysis**

**Step 1: Load Libraries and Input Data**

* Loads necessary R packages
* Loads metadata, full abundance data, and MaAsLin2 significant species.
* Cleans and reformats species taxonomy for readability.
* Filters the abundance table to include only the significant species.
* Filters metadata to exclude other experimental groups (e.g., "WT" and "D2 HMA").

**Step 2: Data Preprocessing for Visualization**

* Selects the top 30 significant species.
* Extracts and formats taxonomic lineage for row annotations (focus on family and species).
* Constructs a custom sample annotation (annot2) to be shown as a color bar.
* Uses the Okabe-Ito colorblind-friendly palette for consistency and accessibility.
* Applies a custom sample ordering for interpretability.

**Step 3: Plot Heatmap**

* Plots a heatmapof **log10-transformed** abundance values.
* Adds sample group (column) and taxonomic family (row) annotations.
* Highlights species names as row labels with increased font size**.**
* Disables column name display.

**📈 Output**

* A customized **heatmap** visualizing the top 30 significant microbial species across samples corresponding to Figure 6C or Supplementary Figure 2D and 6D in the manuscript.

**⚠️ Notes**

* The heatmap includes only **D7 HMA** samples.
  + To visualize other experimental groups (e.g., **WT or D2 HMA**), adjust **lines 21–22** in the script.
* Heatmap row annotations use taxonomic **Family**. Modify to **Genus or Order** if needed.