📊 **UpSet Plot for Species-Level Gut Microbiome Analysis**

This script generates **UpSet plot**s to visualize species-level intersections across experimental groups in a gut microbiome dataset (Donor 2 (D2) and Donor 7 (D7) Human Microbiome Associated (HMA) vs. wild-type). Taxonomic annotations are included to highlight the **top 12 most prevalent bacterial families**.

📁 **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
* metadata\_WC\_HMAvsWT.txt — Sample metadata file,
  + create it based on Supplementary Table 17
* Update the setwd() path in the script as needed.

**🔧 Requirements**

Install the following R packages if not already installed:

install.packages(c("reshape2", "tidyr", "ggplot2", "RColorBrewer", "stringr", "writexl", "ggprism", “ComplexUpset”))

🚀 **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 wc\_upset.R using your preferred R environment.

**🧬 Overview of Analysis**

**Step 1: Load Libraries and Data**

* Loads necessary R packages for ecological and statistical analysis.
* Loads the OTU table and metadata file.
* Filters OTU data to include only **species-level** taxonomic entries (lines matching "s\_\_" and excluding "t\_\_").

**Step 2: Metadata Alignment**

Matches OTU data with sample metadata by run accession.

**Step 3: Data Transformation**

* Converts raw abundance values to **presence/absence (binary) format**.
* Aggregates presence across groups (WT, D2 HMA, D7 HMA).

**Step 4:** **UpSet Plot**

* Generates an initial **UpSet plot** using ComplexUpset.

**Step 5:** **Taxonomic Annotation**

* Adds taxonomic lineage information (Domain → Species).
* Filters for the **top 12 families** by frequency.

**Step 6: Upset Plot with Taxonomic Annotations**

* Constructs an annotated UpSet plot with taxonomic bars and custom colors.

**📈 Output**

* **UpSet plot** displaying species-level overlap across conditions corresponding to Figure 3C.
* **Color-coded bar annotations** for taxonomic family frequencies corresponding to Figure 3C.