📊 **Beta Diversity Analysis Using Bray-Curtis and PCoA**

This R script performs **beta diversity analysis** on microbiome data using **Bray-Curtis distance** and **Principal Coordinates Analysis (PCoA)**, and visualizes the results with ggplot2.

📁 **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("ape", "vegan", "ggplot2", "tidyr", "data.table"))

remotes::install\_github("csdaw/ggprism") # for ggprism

🚀 **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\_beta\_div.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: Beta Diversity Analysis**

* **Bray-Curtis dissimilarity** is calculated between samples using the vegan::vegdist() function.
* **PCoA** is performed using the ape::pcoa() function.

**Step 3: Visualization**

* The first two PCoA axes are extracted and plotted with ggplot2.
* Points are colored by sample group (sample title).
* A 95% confidence ellipse is drawn around each group using stat\_ellipse().

**📈 Output**

* A **PCoA plot** of Figure 3A is generated, showing the distribution of samples in reduced dimensional space.