

Gut Microbiome Analysis WP2020-0022-015

Rogério Castro

2025-07-31

- 1 Start
- 2 Alpha diversity
 - 2.1 Alpha diversity calculations
 - 2.2 Add Sex and Genotype to Metadata
 - 2.3 Alpha diversity statistical analysis
- 3 Beta diversity
 - 3.1 MDS
 - 3.2 MDS centroids and elyseses
 - 3.3 CAP
 - 3.4 CAP centroids and elyseses
 - 3.5 PERMANOVA
 - 3.6 PAIRWISE.ADONIS
- 4 Genera abundance and prevalence
 - 4.1 MaAsLin3
 - 4.2 Significant genera
- 5 Visualization
 - 5.1 Group averages
 - 5.2 Individual samples
 - 5.3 Density plots

1 Start

```
# Load packages
library(MultiAssayExperiment)
library(mia)
library(dplyr)
library(knitr)
library(miaViz)
library(ggcorrplot)

# Read rds file
RDS_015 <- readRDS("C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut microbiome analysis/Preprocessed
```

2 Alpha diversity

2.1 Alpha diversity calculations

```
# Transform assay to relative abundance
RDS_015 <- mia::transformAssay(RDS_015, assay.type = 'counts', method = 'relabundance')

# Estimate alpha diversity indices
RDS_015 <- addAlpha(RDS_015, assay.type = 'counts', index = c('coverage', 'faith', 'fisher', 'gini_simpson', 'inverse

# Estimate alpha dominance indices
RDS_015 <- addAlpha(RDS_015, assay.type = 'counts', index = c('absolute', 'dbp', 'core_abundance', 'gini', 'dmn', 're

# Estimate alpha evenness indices
RDS_015 <- addAlpha(RDS_015, assay.type = 'counts', index = c('camargo', 'simpson_evenness', 'pielou', 'evan', 'bulla
```

```
# Estimate alpha richness indices
RDS_015 <- addAlpha(RDS_015, assay.type = 'counts', index = c('ace', 'chao1', 'hill', 'observed'))

# Export alpha diversity calculations
write.csv(RDS_015@colData, "C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut microbiome analysis/Prepr
```

2.2 Add Sex and Genotype to Metadata

```
# Load packages
library(SummarizedExperiment)
library(phyloseq)
library(vegan)
library(ggplot2)
library(compositions)
library(edgeR)
library(microbiome)
library(ggsci)
library(TreeSummarizedExperiment)
library(readxl)

# Read the Excel metadata file
metadata_015 <- read_excel("C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut microbiome analysis/Prepr

# Create a new metadata table using values from the file
group_metadata_015 <- data.frame(SampleID = colnames(RDS_015))

# Merge in metadata from the Excel file based on SampleID
metadata_015 <- as.data.frame(metadata_015)
group_metadata_015 <- merge(group_metadata_015, metadata_015, by.x = "SampleID", by.y = "Behavioral ID", all.x = TRUE)

# Set rownames and clean up
rownames(group_metadata_015) <- group_metadata_015$SampleID
group_metadata_015$SampleID <- NULL

# Merge with existing colData
colData(RDS_015) <- cbind(colData(RDS_015), group_metadata_015)
```

2.3 Alpha diversity statistical analysis

```
# Extract colData into a dataframe
alpha_df <- as.data.frame(colData(RDS_015))

# List of alpha diversity indices to analyze
alpha_cols <- c('coverage', 'faith', 'fisher', 'gini_simpson', 'inverse_simpson', 'log_modulo_skewness', 'shannon', '

# Ensure Sex and Genotype are factors
alpha_df$Sex <- as.factor(alpha_df$Sex)
alpha_df$Genotype <- as.factor(alpha_df$Genotype)

# Function to safely run 2-way ANOVA and extract p-values
run_anova <- function(index) {
  formula <- as.formula(paste(index, "~ Sex * Genotype"))
  result <- tryCatch({
    model <- aov(formula, data = alpha_df)
```

```

smry <- summary(model)[[1]]
pvals <- smry[["Pr(>F)"]][1:3]
names(pvals) <- c("Sex", "Genotype", "Interaction")
data.frame(Index = index, t(pvals))
}, error = function(e) {
  data.frame(Index = index, Sex = NA, Genotype = NA, Interaction = NA)
})
}

# Run ANOVA for all alpha indices
raw_anova_results <- do.call(rbind, lapply(alpha_cols, run_anova))

# Remove rows with NA to apply p-value correction only on valid results
valid_results <- na.omit(raw_anova_results)

# Apply BH correction
valid_results$Sex_adj <- p.adjust(valid_results$Sex, method = "BH")
valid_results$Genotype_adj <- p.adjust(valid_results$Genotype, method = "BH")
valid_results$Interaction_adj <- p.adjust(valid_results$Interaction, method = "BH")

# Merge corrected p-values back into full result set
anova_results_corrected <- merge(raw_anova_results, valid_results[, c("Index", "Sex_adj", "Genotype_adj", "Interaction_adj")], by = "Index")

# Reorder for clarity
anova_results_corrected <- anova_results_corrected[, c("Index", "Sex", "Sex_adj", "Genotype", "Genotype_adj", "Interaction", "Interaction_adj")]

# View results
print(anova_results_corrected)

```

	Index	Sex	Sex_adj	Genotype	Genotype_adj	Interaction
## 1	absolute	0.9484756	0.9899383	0.895553832	0.89555383	0.7149315
## 2	ace	0.8994078	0.9899383	0.004408682	0.02538956	0.2433116
## 3	bull	0.8742798	0.9899383	0.128087405	0.26781912	0.6487432
## 4	camargo	0.7673063	0.9899383	0.184759727	0.32688259	0.7203227
## 5	chao1	0.8995832	0.9899383	0.004415436	0.02538956	0.2430112
## 6	core_abundance	0.7931300	0.9899383	0.415620492	0.50311954	0.0984459
## 7	coverage	0.9328166	0.9899383	0.021188950	0.07276867	0.5588819
## 8	dbp	0.6312899	0.9899383	0.763390661	0.79809024	0.3497435
## 9	dmn	0.7836698	0.9899383	0.392667408	0.50311954	0.2429101
## 10	evan	0.7915408	0.9899383	0.491787277	0.56555537	0.6102353
## 11	faith	0.1783910	0.9899383	0.093541743	0.21514601	0.2726976
## 12	fisher	0.9899383	0.9899383	0.003715487	0.02538956	0.2775166
## 13	gini	0.7742899	0.9899383	0.012507440	0.05753423	0.6035385
## 14	gini_simpson	0.7283458	0.9899383	0.408939255	0.50311954	0.2257520
## 15	hill	0.5144230	0.9899383	0.022146987	0.07276867	0.7172792
## 16	inverse_simpson	0.3001505	0.9899383	0.055602749	0.14209592	0.7296343
## 17	log_modulo_skewness	0.1434745	0.9899383	0.291656461	0.47914990	0.6821158
## 18	observed	0.8995735	0.9899383	0.004415575	0.02538956	0.2430073
## 19	pielou	0.9818507	0.9899383	0.149693929	0.28691336	0.4698614
## 20	relative	0.6312899	0.9899383	0.763390661	0.79809024	0.3497435
## 21	shannon	0.9170140	0.9899383	0.046917395	0.13488751	0.3954059
## 22	simpson_evenness	0.3205675	0.9899383	0.384197272	0.50311954	0.6939565
## 23	simpson_lambda	0.7283458	0.9899383	0.408939255	0.50311954	0.2257520
##	Interaction_adj					
## 1		0.7296343				
## 2		0.7092092				
## 3		0.7296343				
## 4		0.7296343				

```
## 5      0.7092092
## 6      0.7092092
## 7      0.7296343
## 8      0.7296343
## 9      0.7092092
## 10     0.7296343
## 11     0.7092092
## 12     0.7092092
## 13     0.7296343
## 14     0.7092092
## 15     0.7296343
## 16     0.7296343
## 17     0.7296343
## 18     0.7092092
## 19     0.7296343
## 20     0.7296343
## 21     0.7296343
## 22     0.7296343
## 23     0.7092092
```

3 Beta diversity

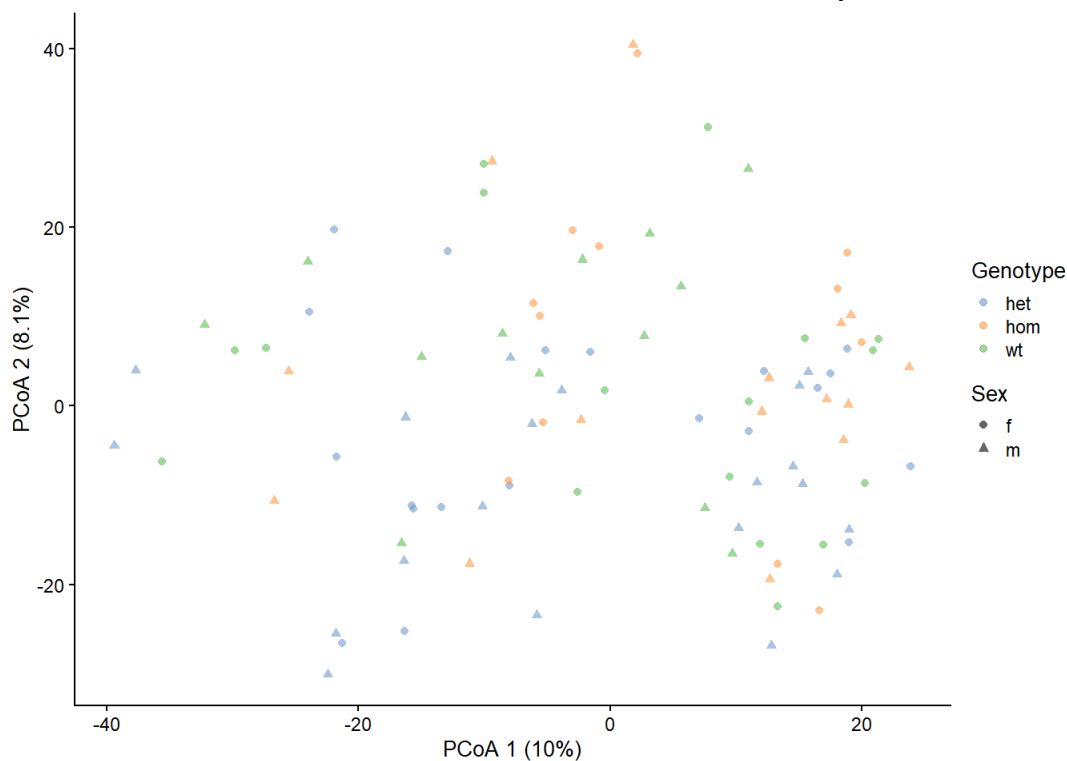
3.1 MDS

```
set.seed(123) # Ensures reproducibility

# Run MDS
RDS_015 <- transformAssay(RDS_015, assay.type = 'counts', method = 'clr', pseudocount = 1)
RDS_015 <- scatter::runMDS(RDS_015, FUN = vegan::vegdist, method = "euclidean", assay.type = "clr", name = "MDS_aitchison")

# Extract eigenvalues for percentage variation explained
e <- attr(reducedDim(RDS_015, "MDS_aitchison"), "eig")
rel_eig <- e / sum(e[e > 0])

# Plot MDS results
scatter::plotReducedDim(RDS_015, "MDS_aitchison", ncomponents = 2, colour_by = "Genotype", shape_by = "Sex") + labs(x
```



```
# Extract PCoA1 and PCoA2 coordinates
mds_coords_asv <- as.data.frame(reducedDim(RDS_015, "MDS_aitchison"))

# Add sample names
mds_coords_asv$SampleID <- rownames(mds_coords_asv)

# Rename columns for clarity
colnames(mds_coords_asv)[1:2] <- c("PCoA1", "PCoA2")

# Export to CSV
write.csv(mds_coords_asv, "C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut microbiome analysis/Prepr
```

3.2 MDS centroids and elyptes

```
## Beta diversity calculations - Aitchison distance (MDS with scater) - centroid and elyptes

#Load packages
library(scater)
library(MASS)
library(tibble)

set.seed(123)

# ----- MDS -----
RDS_015 <- transformAssay(RDS_015, assay.type = 'counts', method = 'clr', pseudocount = 1)
RDS_015 <- scater::runMDS(
  RDS_015,
  FUN = vegan::vegdist,
  method = "euclidean",
  assay.type = "clr",
  name = "MDS_aitchison"
)
```

```
e <- attr(reducedDim(RDS_015, "MDS_aitchison"), "eig")
rel_eig <- e / sum(e[e > 0])

# ----- Extract coordinates and group info -----
mds_coords <- as.data.frame(reducedDim(RDS_015, "MDS_aitchison"))
colnames(mds_coords)[1:2] <- c("PCoA1", "PCoA2")
mds_coords$SampleID <- rownames(mds_coords)
mds_coords$Genotype <- colData(RDS_015)$Genotype
mds_coords$Sex <- colData(RDS_015)$Sex
mds_coords$Group <- paste(mds_coords$Genotype, mds_coords$Sex, sep = "_")

# ----- Centroids -----
centroids <- mds_coords %>%
  group_by(Group) %>%
  summarise(
    PCoA1 = mean(PCoA1),
    PCoA2 = mean(PCoA2),
    Genotype = first(Genotype),
    Sex = first(Sex),
    .groups = 'drop'
  )

# ----- Ellipse points -----
# Function to compute ellipse points
compute_ellipse <- function(df, level = 0.95, npoints = 100) {
  group_name <- unique(df$Group)
  mu <- colMeans(df[, c("PCoA1", "PCoA2")])
  sigma <- cov(df[, c("PCoA1", "PCoA2")])

  ell <- as.data.frame(ellipse::ellipse(sigma, centre = mu, level = level, npoints = npoints))
  ell$Group <- group_name
  return(ell)
}

# Compute ellipse for each group
ellipse_coords <- mds_coords %>%
  group_by(Group) %>%
  group_split() %>%
  lapply(compute_ellipse) %>%
  bind_rows()

# Export centroids
write.csv(centroids, "C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut microbiome analysis/Preprocess

# Export ellipse points
write.csv(ellipse_coords, "C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut microbiome analysis/Prepr
```

3.3 CAP

```
set.seed(123) # Ensures reproducibility

# Extract data
otu_asv <- assay(RDS_015, "counts")
tax_asv <- rowData(RDS_015)
meta_asv <- colData(RDS_015)[, c("Genotype", "Sex")]

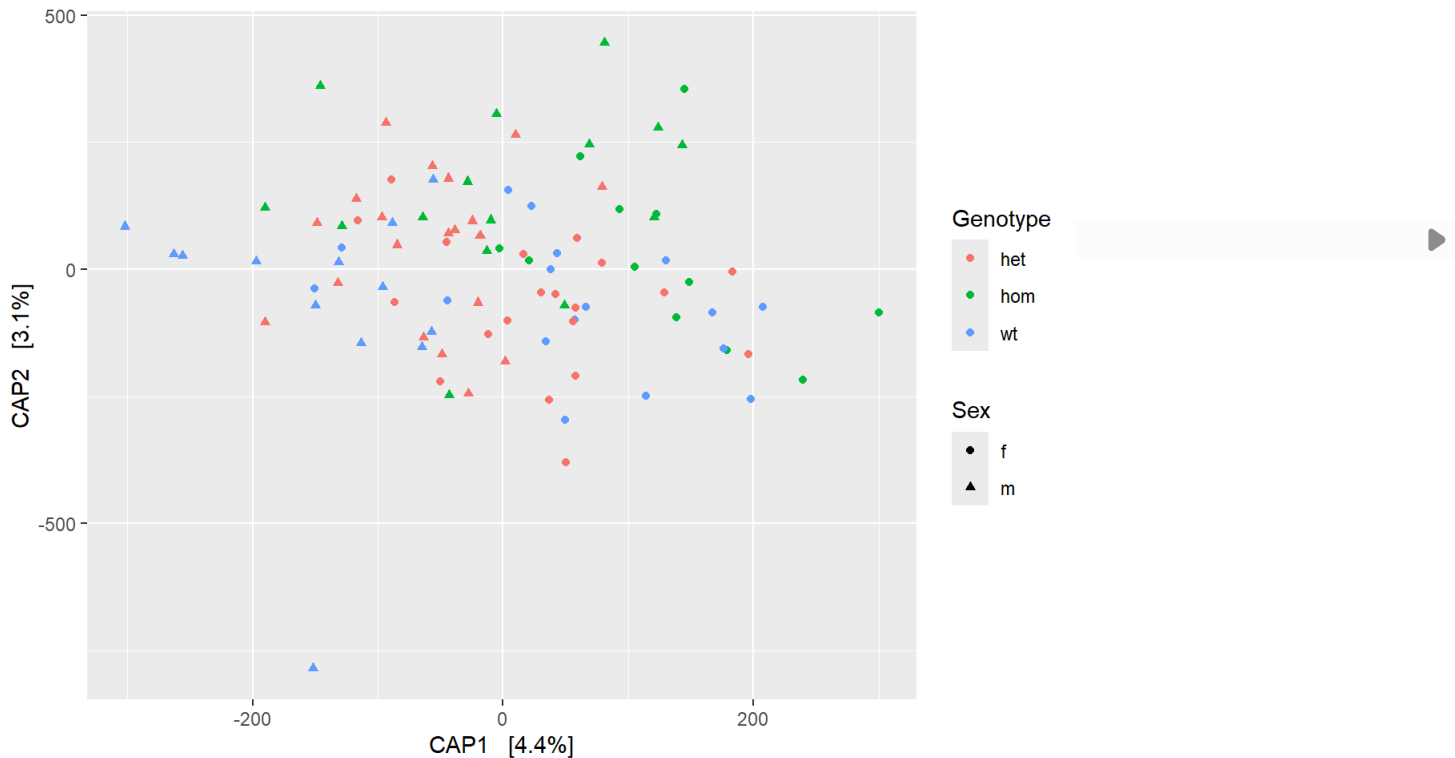
# Convert to phyloseq format
```

```
RDS_015_asv.phylo <- phyloseq(otu_table(otu_asv, taxa_are_rows = TRUE), tax_table(as.matrix(tax_asv)), sample_data(asv_metadata))

# Convert metadata variables to factors
sample_data(RDS_015_asv.phylo)[["Sex"]] <- as.factor(sample_data(RDS_015_asv.phylo)[["Sex"]])
sample_data(RDS_015_asv.phylo)[["Genotype"]] <- as.factor(sample_data(RDS_015_asv.phylo)[["Genotype"]])

# CAP ordination
capdiag_factor_asv <- phyloseq::ordinate(RDS_015_asv.phylo, method = "CAP", distance = "euclidean", formula = ~ Genotype + Sex)

# Plot CAP results
plotgroupstat_asv <- plot_ordination(RDS_015_asv.phylo, capdiag_factor_asv, color = "Genotype", shape = "Sex")
plotgroupstat_asv
```



3.4 CAP centroids and elypses

```
#Load packages
library(ellipse)

set.seed(123) # Ensures reproducibility

# Extract data
otu_asv <- assay(RDS_015, "counts")
tax_asv <- rowData(RDS_015)
meta_asv <- colData(RDS_015)[, c("Genotype", "Sex")]

# Convert to phyloseq format
RDS_015_asv.phylo <- phyloseq(
  otu_table(otu_asv, taxa_are_rows = TRUE),
  tax_table(as.matrix(tax_asv)),
  sample_data(as.data.frame(meta_asv))
)

# Convert metadata to factors
```

```

sample_data(RDS_015_asv.phylo)[["Sex"]] <- as.factor(sample_data(RDS_015_asv.phylo)[["Sex"]])
sample_data(RDS_015_asv.phylo)[["Genotype"]] <- as.factor(sample_data(RDS_015_asv.phylo)[["Genotype"]])

# CAP ordination
capdiag_factor_asv <- ordinate(
  RDS_015_asv.phylo,
  method = "CAP",
  distance = "euclidean",
  formula = ~ Genotype * Sex + Genotype + Sex,
  weighted = TRUE,
  na.action = na.exclude
)

# Extract ordination coordinates
plotgroupstat_asv <- plot_ordination(RDS_015_asv.phylo, capdiag_factor_asv, color = "Genotype", shape = "Sex")
ord_data <- plotgroupstat_asv$data
ord_data$Group <- paste(ord_data$Genotype, ord_data$Sex, sep = "_")

# ----- Compute centroids -----
centroids <- ord_data %>%
  group_by(Group) %>%
  summarise(
    CAP1 = mean(CAP1),
    CAP2 = mean(CAP2),
    Genotype = first(Genotype),
    Sex = first(Sex),
    .groups = 'drop'
  )

# ----- Compute ellipse coordinates -----
compute_ellipse <- function(df, level = 0.95, npoints = 100) {
  group_name <- unique(df$Group)
  mu <- colMeans(df[, c("CAP1", "CAP2")])
  sigma <- cov(df[, c("CAP1", "CAP2")])
  ell <- as.data.frame(ellipse::ellipse(sigma, centre = mu, level = level, npoints = npoints))
  ell$Group <- group_name
  return(ell)
}

ellipse_coords <- ord_data %>%
  group_by(Group) %>%
  group_split() %>%
  lapply(compute_ellipse) %>%
  bind_rows()

# Export to CSV
write.csv(ord_data, "C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut microbiome analysis/Preprocess
write.csv(centroids, "C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut microbiome analysis/Preprocess
write.csv(ellipse_coords, "C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut microbiome analysis/Prepr

```

3.5 PERMANOVA

```

set.seed(123) # Ensures reproducibility

mddperm0 <- vegan::adonis2(t(assay(RDS_015, 'clr')) ~ Genotype * Sex, by = 'terms', data = colData(RDS_015), method =
mddperm0

```

```
## Permutation test for adonis under reduced model
## Terms added sequentially (first to last)
## Permutation: free
## Number of permutations: 999
##
## vegan::adonis2(formula = t(assay(RDS_015, "clr")) ~ Genotype * Sex, data = colData(RDS_015), permutations = 999,
method = "euclidean", by = "terms", na.action = na.exclude)
##
```

	Df	SumOfSqs	R2	F	Pr(>F)
Genotype	2	21086	0.04376	2.2228	0.001 ***
Sex	1	13735	0.02851	2.8958	0.001 ***
Genotype:Sex	2	15366	0.03189	1.6198	0.001 ***
Residual	91	431621	0.89584		
Total	96	481808	1.00000		

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As a sensitivity analysis

set.seed(123) # Ensures reproducibility

```
mddperm1 <- vegan::adonis2(t(assay(RDS_015, 'clr')) ~ Sex * Genotype, by = 'terms', data = colData(RDS_015), method =
mddperm1
```

```
## Permutation test for adonis under reduced model
## Terms added sequentially (first to last)
## Permutation: free
## Number of permutations: 999
##
## vegan::adonis2(formula = t(assay(RDS_015, "clr")) ~ Sex * Genotype, data = colData(RDS_015), permutations = 999,
method = "euclidean", by = "terms", na.action = na.exclude)
##
```

	Df	SumOfSqs	R2	F	Pr(>F)
Sex	1	13766	0.02857	2.9023	0.001 ***
Genotype	2	21055	0.04370	2.2196	0.001 ***
Sex:Genotype	2	15366	0.03189	1.6198	0.001 ***
Residual	91	431621	0.89584		
Total	96	481808	1.00000		

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

set.seed(123) # Ensures reproducibility

```
mddperm2 <- vegan::adonis2(t(assay(RDS_015, 'clr')) ~ Genotype + Sex, by = 'margin', data = colData(RDS_015), method =
mddperm2
```

```
## Permutation test for adonis under reduced model
## Marginal effects of terms
## Permutation: free
## Number of permutations: 999
##
## vegan::adonis2(formula = t(assay(RDS_015, "clr")) ~ Genotype + Sex, data = colData(RDS_015), permutations = 999,
method = "euclidean", by = "margin", na.action = na.exclude)
##
```

	Df	SumOfSqs	R2	F	Pr(>F)
Genotype	2	21055	0.04370	2.1904	0.001 ***
Sex	1	13735	0.02851	2.8577	0.001 ***

```
## Residual 93    446987 0.92773
## Total    96    481808 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
set.seed(123) # Ensures reproducibility
```

```
mddperm3 <- vegan::adonis2(t(assay(RDS_015, 'clr')) ~ Genotype * Sex, by = 'margin', data = colData(RDS_015), method = 'euclidean')
mddperm3
```

```
## Permutation test for adonis under reduced model
## Marginal effects of terms
## Permutation: free
## Number of permutations: 999
##
## vegan::adonis2(formula = t(assay(RDS_015, "clr")) ~ Genotype * Sex, data = colData(RDS_015), permutations = 999,
method = "euclidean", by = "margin", na.action = na.exclude)
##           Df SumOfSqs      R2      F Pr(>F)
## Genotype:Sex  2    15366 0.03189 1.6198  0.001 ***
## Residual    91   431621 0.89584
## Total      96   481808 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
set.seed(123) # Ensures reproducibility
```

```
mddperm4 <- vegan::adonis2(t(assay(RDS_015, 'clr')) ~ Genotype * Sex, by = NULL, data = colData(RDS_015), method = 'euclidean')
mddperm4
```

```
## Permutation test for adonis under reduced model
## Permutation: free
## Number of permutations: 999
##
## vegan::adonis2(formula = t(assay(RDS_015, "clr")) ~ Genotype * Sex, data = colData(RDS_015), permutations = 999,
method = "euclidean", by = NULL, na.action = na.exclude)
##           Df SumOfSqs      R2      F Pr(>F)
## Model      5    50187 0.10416 2.1162  0.001 ***
## Residual  91   431621 0.89584
## Total    96   481808 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3.6 PAIRWISE.ADONIS

```
# Load packages
```

```
library(pairwiseAdonis)
```

```
set.seed(123) # Ensures reproducibility
```

```
# Calculate distance matrix
```

```
dist_matrix <- vegdist(t(assay(RDS_015, "clr")), method = "euclidean")
```

```
# Create an interaction term
```

```
colData(RDS_015)$Genotype_Sex <- interaction(colData(RDS_015)$Genotype, colData(RDS_015)$Sex)

# Pairwise comparisons
pairwise_results_interaction <- pairwise.adonis(dist_matrix, colData(RDS_015)$Genotype_Sex, p.adjust.m = "fdr")
print(pairwise_results_interaction)
```

```
##           pairs Df SumsOfSqs  F.Model      R2 p.value  p.adjusted sig
## 1  het.m vs hom.m  1 12364.008 2.468983 0.06960962  0.001 0.001875000  *
## 2  het.m vs wt.m  1  9074.425 1.848954 0.05628654  0.003 0.003750000  *
## 3  het.m vs het.f  1 10111.467 2.048723 0.05115576  0.001 0.001875000  *
## 4  het.m vs wt.f  1  9729.931 1.964175 0.05313727  0.003 0.003750000  *
## 5  het.m vs hom.f  1 12155.036 2.477466 0.07628262  0.001 0.001875000  *
## 6  hom.m vs wt.m  1  9487.959 2.048003 0.07301780  0.002 0.003000000  *
## 7  hom.m vs het.f  1 13601.012 2.879770 0.08026167  0.001 0.001875000  *
## 8  hom.m vs wt.f  1 10720.008 2.269779 0.07033762  0.001 0.001875000  *
## 9  hom.m vs hom.f  1 10851.706 2.348944 0.08588794  0.001 0.001875000  *
## 10 wt.m vs het.f  1  9986.115 2.168673 0.06538316  0.002 0.003000000  *
## 11 wt.m vs wt.f  1  8138.137 1.772234 0.05952640  0.006 0.006923077  *
## 12 wt.m vs hom.f  1 10604.591 2.382256 0.09385518  0.001 0.001875000  *
## 13 het.f vs wt.f  1  6474.707 1.381953 0.03798456  0.033 0.033000000  .
## 14 het.f vs hom.f  1  9685.986 2.108870 0.06567873  0.001 0.001875000  *
## 15 wt.f vs hom.f  1  7465.250 1.630498 0.05694968  0.014 0.015000000  .
```

4 Genera abundance and prevalence

4.1 MaAsLin3

```
# Load packages
library(maaslin3)
library(dplyr)

set.seed(123)

# Extract abundance matrix
feature_table_015_raw <- assay(RDS_015)

# Extract metadata
metadata_015 <- as.data.frame(colData(RDS_015))

# Extract taxonomy
taxonomy_015 <- as.data.frame(rowData(RDS_015))
taxonomy_015$genus <- as.character(taxonomy_015$genus)
keep_idx <- !is.na(taxonomy_015$genus) & taxonomy_015$genus != ""
taxonomy_015_filtered <- taxonomy_015[keep_idx, ]
feature_table_015_filtered <- feature_table_015_raw[keep_idx, ]

# Combine abundance with genus info
feature_table_015_df <- as.data.frame(feature_table_015_filtered)
feature_table_015_df$genus <- taxonomy_015_filtered$genus

# Aggregate by genus
feature_table_015_genus <- feature_table_015_df %>%
  group_by(genus) %>%
  summarise(across(everything(), sum), .groups = "drop")

# Set genus as row names
feature_table_015_genus <- as.data.frame(feature_table_015_genus)
```

```

rownames(feature_table_015_genus) <- feature_table_015_genus$genus
feature_table_015_genus$genus <- NULL

# Transpose
feature_table_015 <- t(feature_table_015_genus)

# Ensure categorical metadata
metadata_015$Genotype <- factor(metadata_015$Genotype, levels = c("wt", "het", "hom"))
metadata_015$Sex <- factor(metadata_015$Sex, levels = c("m", "f"))
metadata_015$Genotype_Sex <- factor(metadata_015$Genotype_Sex, levels = c("wt.m", "het.m", "hom.m", "wt.f", "het.f"),

# Set reference levels for interpretation
metadata_015$Genotype <- relevel(metadata_015$Genotype, ref = "wt")
metadata_015$Sex <- relevel(metadata_015$Sex, ref = "m")
metadata_015$Genotype_Sex <- relevel(metadata_015$Genotype_Sex, ref = "wt.m")

# Filter low-abundance genera
feature_table_015 <- feature_table_015[, colMeans(feature_table_015) > 0.10]

# Run MaAsLin3
fit_data_015_genus <- maaslin3(
  input_data = feature_table_015,
  input_metadata = metadata_015,
  output = "C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut microbiome analysis/Preprocessed GM data
  formula = (~ group(Genotype) + group(Sex) + group(Genotype_Sex)),
  normalization = "TSS",
  transform = "LOG",
  standardize = TRUE,
  max_significance = 0.05
)

```

```

## 2025-07-31 12:36:05.974023 INFO::Writing function arguments to log file
## 2025-07-31 12:36:06.018317 INFO::Verifying options selected are valid
## 2025-07-31 12:36:06.020199 WARNING::Input is a matrix,
##           passing through as.data.frame() .
## 2025-07-31 12:36:06.024683 INFO::Determining format of input files
## 2025-07-31 12:36:06.026558 INFO::Input format is data samples as rows and metadata samples as rows
## 2025-07-31 12:36:06.03218 INFO::Running selected normalization method: TSS
## 2025-07-31 12:36:06.038366 INFO::Writing normalized data to file C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive
Neuroscience/Gut microbiome analysis/Preprocessed GM data/Analysis WP2020-0022-015/Genus/features/data_norm.tsv
## 2025-07-31 12:36:06.07293 INFO::Filter data based on min abundance and min prevalence
## 2025-07-31 12:36:06.074219 INFO::Total samples in data: 97
## 2025-07-31 12:36:06.075329 INFO::Min samples required with min abundance for a feature not to be filtered: 0.000
000
## 2025-07-31 12:36:06.087779 INFO::Total filtered features: 0
## 2025-07-31 12:36:06.089199 INFO::Filtered feature names from abundance and prevalence filtering:
## 2025-07-31 12:36:06.092341 INFO::Total features filtered by non-zero variance filtering: 3
## 2025-07-31 12:36:06.093409 INFO::Filtered feature names from variance filtering: Frisingicoccus, Saccharimonad
ales, Sellimonas
## 2025-07-31 12:36:06.094403 INFO::Writing filtered data to file C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive N
euroscience/Gut microbiome analysis/Preprocessed GM data/Analysis WP2020-0022-015/Genus/features/filtered_data.tsv
## 2025-07-31 12:36:06.129274 INFO::Running selected transform method: LOG
## 2025-07-31 12:36:06.132736 INFO::Writing normalized, filtered, transformed data to file C:/Users/roger/OneDrive/
Desktop/Ph.D. Cognitive Neuroscience/Gut microbiome analysis/Preprocessed GM data/Analysis WP2020-0022-015/Genus/fe
atures/data_transformed.tsv
## 2025-07-31 12:36:06.171149 INFO::Applying z-score to standardize continuous metadata
## 2025-07-31 12:36:06.190365 INFO::Running the linear model component
## 2025-07-31 12:36:06.219618 INFO::Fitting model to feature number 1, A2

```

```
## 2025-07-31 12:36:06.232295 INFO::Fitting model to feature number 2, ASF356
## 2025-07-31 12:36:06.242361 INFO::Fitting model to feature number 3, Acetatifactor
## 2025-07-31 12:36:06.262141 INFO::Fitting model to feature number 4, Actinomyces
## 2025-07-31 12:36:06.274916 INFO::Fitting model to feature number 5, Adlercreutzia
## 2025-07-31 12:36:06.290479 INFO::Fitting model to feature number 6, Aerococcus
## 2025-07-31 12:36:06.305929 INFO::Fitting model to feature number 7, Alistipes
## 2025-07-31 12:36:06.321943 INFO::Fitting model to feature number 8, Allobaculum
## 2025-07-31 12:36:06.339913 INFO::Fitting model to feature number 9, Alloprevotella
## 2025-07-31 12:36:06.355462 INFO::Fitting model to feature number 10, Anaerobiospirillum
## 2025-07-31 12:36:06.366712 INFO::Fitting model to feature number 11, Anaerofilum
## 2025-07-31 12:36:06.373122 INFO::Fitting model to feature number 12, Anaerofustis
## 2025-07-31 12:36:06.380145 INFO::Fitting model to feature number 13, Anaeroplasma
## 2025-07-31 12:36:06.382541 INFO::Fitting model to feature number 14, Anaerotruncus
## 2025-07-31 12:36:06.389102 INFO::Fitting model to feature number 15, Anaerovibrio
## 2025-07-31 12:36:06.397842 INFO::Fitting model to feature number 16, Anaerovorax
## 2025-07-31 12:36:06.404352 INFO::Fitting model to feature number 17, Bacillus
## 2025-07-31 12:36:06.413709 INFO::Fitting model to feature number 18, BacteroidalesRF16group
## 2025-07-31 12:36:06.425718 INFO::Fitting model to feature number 19, Bacteroides
## 2025-07-31 12:36:06.440445 INFO::Fitting model to feature number 20, Barnesiella
## 2025-07-31 12:36:06.451822 INFO::Fitting model to feature number 21, Bifidobacterium
## 2025-07-31 12:36:06.458939 INFO::Fitting model to feature number 22, Blautia
## 2025-07-31 12:36:06.465255 INFO::Fitting model to feature number 23, Butyricicoccus
## 2025-07-31 12:36:06.475001 INFO::Fitting model to feature number 24, Butyricimonas
## 2025-07-31 12:36:06.482141 INFO::Fitting model to feature number 25, Caldicoprobacter
## 2025-07-31 12:36:06.489521 INFO::Fitting model to feature number 26, CandidatusArthromitus
## 2025-07-31 12:36:06.496771 INFO::Fitting model to feature number 27, CandidatusSaccharimonas
## 2025-07-31 12:36:06.504055 INFO::Fitting model to feature number 28, CandidatusSoleaferrea
## 2025-07-31 12:36:06.512262 INFO::Fitting model to feature number 29, CandidatusStoquefichus
## 2025-07-31 12:36:06.522018 INFO::Fitting model to feature number 30, Caproiciproducens
## 2025-07-31 12:36:06.524311 INFO::Fitting model to feature number 31, Catellicoccus
## 2025-07-31 12:36:06.530715 INFO::Fitting model to feature number 32, Chloroplast
## 2025-07-31 12:36:06.537113 INFO::Fitting model to feature number 33, Christensenellaceae
## 2025-07-31 12:36:06.544189 INFO::Fitting model to feature number 34, ChristensenellaceaeR7group
## 2025-07-31 12:36:06.55333 INFO::Fitting model to feature number 35, ClostridiaUCG014
## 2025-07-31 12:36:06.561221 INFO::Fitting model to feature number 36, ClostridiavadinBB60group
## 2025-07-31 12:36:06.56974 INFO::Fitting model to feature number 37, Clostridiumsensustricto1
## 2025-07-31 12:36:06.576613 INFO::Fitting model to feature number 38, Colidextribacter
## 2025-07-31 12:36:06.583557 INFO::Fitting model to feature number 39, CoriobacteriaceaeUCG002
## 2025-07-31 12:36:06.590419 INFO::Fitting model to feature number 40, Corynebacterium
## 2025-07-31 12:36:06.597454 INFO::Fitting model to feature number 41, DNF00809
## 2025-07-31 12:36:06.60436 INFO::Fitting model to feature number 42, DTU089
## 2025-07-31 12:36:06.606965 INFO::Fitting model to feature number 43, DefluviitaleaceaeUCG011
## 2025-07-31 12:36:06.614083 INFO::Fitting model to feature number 44, Desulfovibrio
## 2025-07-31 12:36:06.621616 INFO::Fitting model to feature number 45, Dielma
## 2025-07-31 12:36:06.627677 INFO::Fitting model to feature number 46, Dorea
## 2025-07-31 12:36:06.635101 INFO::Fitting model to feature number 47, Dubosiella
## 2025-07-31 12:36:06.642791 INFO::Fitting model to feature number 48, Eisenbergiella
## 2025-07-31 12:36:06.649457 INFO::Fitting model to feature number 49, Elusimicrobium
## 2025-07-31 12:36:06.656328 INFO::Fitting model to feature number 50, Enterococcus
## 2025-07-31 12:36:06.66353 INFO::Fitting model to feature number 51, Enterorhabdus
## 2025-07-31 12:36:06.671416 INFO::Fitting model to feature number 52, Erysipelatoclostridiaceae
## 2025-07-31 12:36:06.679277 INFO::Fitting model to feature number 53, Erysipelatoclostridium
## 2025-07-31 12:36:06.682083 INFO::Fitting model to feature number 54, ErysipelotrichaceaeUCG003
## 2025-07-31 12:36:06.685511 INFO::Fitting model to feature number 55, EscherichiaShigella
## 2025-07-31 12:36:06.693588 INFO::Fitting model to feature number 56, Eubacteriumbrachygroup
## 2025-07-31 12:36:06.70316 INFO::Fitting model to feature number 57, Eubacteriumcoprostanoligenesgroup
## 2025-07-31 12:36:06.713631 INFO::Fitting model to feature number 58, Eubacteriumnodatumgroup
## 2025-07-31 12:36:06.724247 INFO::Fitting model to feature number 59, Eubacteriumoxidoreducensgroup
## 2025-07-31 12:36:06.733993 INFO::Fitting model to feature number 60, Eubacteriumruminantiumgroup
## 2025-07-31 12:36:06.736692 INFO::Fitting model to feature number 61, Eubacteriumsiraumgroup
```

```
## 2025-07-31 12:36:06.746099 INFO::Fitting model to feature number 62, Eubacteriumventriosumgroup
## 2025-07-31 12:36:06.754716 INFO::Fitting model to feature number 63, Eubacteriumxylanophilumgroup
## 2025-07-31 12:36:06.762848 INFO::Fitting model to feature number 64, F082
## 2025-07-31 12:36:06.771098 INFO::Fitting model to feature number 65, Faecalibaculum
## 2025-07-31 12:36:06.779715 INFO::Fitting model to feature number 66, FamilyXIIIAD3011group
## 2025-07-31 12:36:06.788518 INFO::Fitting model to feature number 67, FamilyXIIIUCG001
## 2025-07-31 12:36:06.797184 INFO::Fitting model to feature number 68, Fournierella
## 2025-07-31 12:36:06.800565 INFO::Fitting model to feature number 69, Fusobacterium
## 2025-07-31 12:36:06.810423 INFO::Fitting model to feature number 70, GCA900066575
## 2025-07-31 12:36:06.819529 INFO::Fitting model to feature number 71, Gastranaerophilales
## 2025-07-31 12:36:06.829037 INFO::Fitting model to feature number 72, Gemella
## 2025-07-31 12:36:06.837421 INFO::Fitting model to feature number 73, Globicatella
## 2025-07-31 12:36:06.850713 INFO::Fitting model to feature number 74, Harryflintia
## 2025-07-31 12:36:06.861179 INFO::Fitting model to feature number 75, Helicobacter
## 2025-07-31 12:36:06.868957 INFO::Fitting model to feature number 76, Hydrogenoanaerobacterium
## 2025-07-31 12:36:06.876783 INFO::Fitting model to feature number 77, Ileibacterium
## 2025-07-31 12:36:06.88561 INFO::Fitting model to feature number 78, IncertaeSedis
## 2025-07-31 12:36:06.893211 INFO::Fitting model to feature number 79, Intestinimonas
## 2025-07-31 12:36:06.900537 INFO::Fitting model to feature number 80, Jeotgalicoccus
## 2025-07-31 12:36:06.903962 INFO::Fitting model to feature number 81, Lachnoclostridium
## 2025-07-31 12:36:06.912059 INFO::Fitting model to feature number 82, LachnospiraceaeFCS020group
## 2025-07-31 12:36:06.923233 INFO::Fitting model to feature number 83, LachnospiraceaeNC2004group
## 2025-07-31 12:36:06.93407 INFO::Fitting model to feature number 84, LachnospiraceaeND3007group
## 2025-07-31 12:36:06.937883 INFO::Fitting model to feature number 85, LachnospiraceaeNK4A136group
## 2025-07-31 12:36:06.945619 INFO::Fitting model to feature number 86, LachnospiraceaeUCG001
## 2025-07-31 12:36:06.954 INFO::Fitting model to feature number 87, LachnospiraceaeUCG004
## 2025-07-31 12:36:06.961817 INFO::Fitting model to feature number 88, LachnospiraceaeUCG006
## 2025-07-31 12:36:06.970303 INFO::Fitting model to feature number 89, LachnospiraceaeUCG008
## 2025-07-31 12:36:06.977961 INFO::Fitting model to feature number 90, LachnospiraceaeUCG009
## 2025-07-31 12:36:06.981048 INFO::Fitting model to feature number 91, LachnospiraceaeUCG010
## 2025-07-31 12:36:06.989504 INFO::Fitting model to feature number 92, Lactobacillus
## 2025-07-31 12:36:06.999166 INFO::Fitting model to feature number 93, Leptotrichia
## 2025-07-31 12:36:07.007644 INFO::Fitting model to feature number 94, Marvinbryantia
## 2025-07-31 12:36:07.015271 INFO::Fitting model to feature number 95, Methanosphaera
## 2025-07-31 12:36:07.017848 INFO::Fitting model to feature number 96, Mitochondria
## 2025-07-31 12:36:07.02462 INFO::Fitting model to feature number 97, Monoglobus
## 2025-07-31 12:36:07.06898 INFO::Fitting model to feature number 98, Mucispirillum
## 2025-07-31 12:36:07.076674 INFO::Fitting model to feature number 99, Muribaculaceae
## 2025-07-31 12:36:07.083525 INFO::Fitting model to feature number 100, Muribaculum
## 2025-07-31 12:36:07.090129 INFO::Fitting model to feature number 101, NK4A214group
## 2025-07-31 12:36:07.096456 INFO::Fitting model to feature number 102, Negativibacillus
## 2025-07-31 12:36:07.103465 INFO::Fitting model to feature number 103, Neisseria
## 2025-07-31 12:36:07.11257 INFO::Fitting model to feature number 104, Odoribacter
## 2025-07-31 12:36:07.119444 INFO::Fitting model to feature number 105, Oscillibacter
## 2025-07-31 12:36:07.126816 INFO::Fitting model to feature number 106, Oscillospira
## 2025-07-31 12:36:07.133465 INFO::Fitting model to feature number 107, Paludicola
## 2025-07-31 12:36:07.142743 INFO::Fitting model to feature number 108, Papillibacter
## 2025-07-31 12:36:07.148971 INFO::Fitting model to feature number 109, Parabacteroides
## 2025-07-31 12:36:07.15692 INFO::Fitting model to feature number 110, Paraprevotella
## 2025-07-31 12:36:07.159788 INFO::Fitting model to feature number 111, Parasutterella
## 2025-07-31 12:36:07.167471 INFO::Fitting model to feature number 112, Parvibacter
## 2025-07-31 12:36:07.173846 INFO::Fitting model to feature number 113, Pediococcus
## 2025-07-31 12:36:07.176076 INFO::Fitting model to feature number 114, Peptococcus
## 2025-07-31 12:36:07.18279 INFO::Fitting model to feature number 115, Peptostreptococcus
## 2025-07-31 12:36:07.185063 INFO::Fitting model to feature number 116, Prevotella
## 2025-07-31 12:36:07.191608 INFO::Fitting model to feature number 117, PrevotellaceaeGa6A1group
## 2025-07-31 12:36:07.19841 INFO::Fitting model to feature number 118, PrevotellaceaeNK3B31group
## 2025-07-31 12:36:07.204941 INFO::Fitting model to feature number 119, PrevotellaceaeUCG001
## 2025-07-31 12:36:07.212692 INFO::Fitting model to feature number 120, PrevotellaceaeUCG003
## 2025-07-31 12:36:07.223988 INFO::Fitting model to feature number 121, Proteus
```

```
## 2025-07-31 12:36:07.233721 INFO::Fitting model to feature number 122, Pygmaibacter
## 2025-07-31 12:36:07.240358 INFO::Fitting model to feature number 123, Quinella
## 2025-07-31 12:36:07.247078 INFO::Fitting model to feature number 124, RF39
## 2025-07-31 12:36:07.2537 INFO::Fitting model to feature number 125, Rikenella
## 2025-07-31 12:36:07.260341 INFO::Fitting model to feature number 126, RikenellaceaeRC9gutgroup
## 2025-07-31 12:36:07.266761 INFO::Fitting model to feature number 127, Rodentibacter
## 2025-07-31 12:36:07.27395 INFO::Fitting model to feature number 128, Romboutsia
## 2025-07-31 12:36:07.280488 INFO::Fitting model to feature number 129, Roseburia
## 2025-07-31 12:36:07.286947 INFO::Fitting model to feature number 130, Rothia
## 2025-07-31 12:36:07.296036 INFO::Fitting model to feature number 131, Ruminiclostridium
## 2025-07-31 12:36:07.30282 INFO::Fitting model to feature number 132, Ruminococcaceae
## 2025-07-31 12:36:07.310285 INFO::Fitting model to feature number 133, Ruminococcus
## 2025-07-31 12:36:07.316888 INFO::Fitting model to feature number 134, Ruminococcusgavreuiiigroup
## 2025-07-31 12:36:07.320067 INFO::Fitting model to feature number 135, Ruminococcustorquesgroup
## 2025-07-31 12:36:07.328906 INFO::Fitting model to feature number 136, Shuttleworthia
## 2025-07-31 12:36:07.33609 INFO::Fitting model to feature number 137, Sphingomonas
## 2025-07-31 12:36:07.34721 INFO::Fitting model to feature number 138, Sporobacter
## 2025-07-31 12:36:07.359832 INFO::Fitting model to feature number 139, Staphylococcus
## 2025-07-31 12:36:07.372907 INFO::Fitting model to feature number 140, Streptococcus
## 2025-07-31 12:36:07.383097 INFO::Fitting model to feature number 141, Subdoligranulum
## 2025-07-31 12:36:07.392504 INFO::Fitting model to feature number 142, Treponema
## 2025-07-31 12:36:07.402583 INFO::Fitting model to feature number 143, Turicibacter
## 2025-07-31 12:36:07.411074 INFO::Fitting model to feature number 144, Tuzzerella
## 2025-07-31 12:36:07.431383 INFO::Fitting model to feature number 145, Tyzzerella
## 2025-07-31 12:36:07.451525 INFO::Fitting model to feature number 146, UBA1819
## 2025-07-31 12:36:07.472324 INFO::Fitting model to feature number 147, UCG002
## 2025-07-31 12:36:07.485379 INFO::Fitting model to feature number 148, UCG004
## 2025-07-31 12:36:07.492262 INFO::Fitting model to feature number 149, UCG005
## 2025-07-31 12:36:07.499862 INFO::Fitting model to feature number 150, UCG007
## 2025-07-31 12:36:07.507354 INFO::Fitting model to feature number 151, UCG009
## 2025-07-31 12:36:07.518438 INFO::Fitting model to feature number 152, UCG010
## 2025-07-31 12:36:07.525326 INFO::Fitting model to feature number 153, UCG011
## 2025-07-31 12:36:07.531545 INFO::Fitting model to feature number 154, Veillonella
## 2025-07-31 12:36:07.539149 INFO::Fitting model to feature number 155, Victivallis
## 2025-07-31 12:36:07.541795 INFO::Fitting model to feature number 156, uncultured
## 2025-07-31 12:36:07.557122 INFO::Performing tests against medians
## 2025-07-31 12:36:07.582475 INFO::Counting total values for each feature
## 2025-07-31 12:36:07.596505 INFO::Running the logistic model component
## 2025-07-31 12:36:07.621798 INFO::Fitting model to feature number 1, A2
## 2025-07-31 12:36:07.65149 INFO::Fitting model to feature number 2, ASF356
## 2025-07-31 12:36:07.67384 INFO::Fitting model to feature number 3, Acetatifactor
## 2025-07-31 12:36:07.69195 INFO::Fitting model to feature number 4, Actinomyces
## 2025-07-31 12:36:07.710884 INFO::Fitting model to feature number 5, Adlercreutzia
## 2025-07-31 12:36:07.712642 INFO::Fitting model to feature number 6, Aerococcus
## 2025-07-31 12:36:07.731803 INFO::Fitting model to feature number 7, Alistipes
## 2025-07-31 12:36:07.734018 INFO::Fitting model to feature number 8, Allobaculum
## 2025-07-31 12:36:07.763542 INFO::Fitting model to feature number 9, Alloprevotella
## 2025-07-31 12:36:07.765527 INFO::Fitting model to feature number 10, Anaerobiospirillum
## 2025-07-31 12:36:07.783655 INFO::Fitting model to feature number 11, Anaerofilum
## 2025-07-31 12:36:07.803691 INFO::Fitting model to feature number 12, Anaerofustis
## 2025-07-31 12:36:07.826061 INFO::Fitting model to feature number 13, Anaeroplasma
## 2025-07-31 12:36:07.846708 INFO::Fitting model to feature number 14, Anaerotruncus
## 2025-07-31 12:36:07.866922 INFO::Fitting model to feature number 15, Anaerovibrio
## 2025-07-31 12:36:07.885979 INFO::Fitting model to feature number 16, Anaerovorax
## 2025-07-31 12:36:07.906594 INFO::Fitting model to feature number 17, Bacillus
## 2025-07-31 12:36:07.928705 INFO::Fitting model to feature number 18, BacteroidalesRF16group
## 2025-07-31 12:36:07.953394 INFO::Fitting model to feature number 19, Bacteroides
## 2025-07-31 12:36:07.955363 INFO::Fitting model to feature number 20, Barnesiella
## 2025-07-31 12:36:07.979852 INFO::Fitting model to feature number 21, Bifidobacterium
## 2025-07-31 12:36:07.982403 INFO::Fitting model to feature number 22, Blautia
```

```

## 2025-07-31 12:36:08.010317 INFO::Fitting model to feature number 23, Butyricicoccus
## 2025-07-31 12:36:08.037947 INFO::Fitting model to feature number 24, Butyricimonas
## 2025-07-31 12:36:08.066163 INFO::Fitting model to feature number 25, Caldicoprobacter
## 2025-07-31 12:36:08.086794 INFO::Fitting model to feature number 26, CandidatusArthromitus
## 2025-07-31 12:36:08.113309 INFO::Fitting model to feature number 27, CandidatusSaccharimonas
## 2025-07-31 12:36:08.157087 INFO::Fitting model to feature number 28, CandidatusSoleaferrea
## 2025-07-31 12:36:08.175089 INFO::Fitting model to feature number 29, CandidatusStoquefichus
## 2025-07-31 12:36:08.193842 INFO::Fitting model to feature number 30, Caproiciproducens
## 2025-07-31 12:36:08.223166 INFO::Fitting model to feature number 31, Catellicoccus
## 2025-07-31 12:36:08.248603 INFO::Fitting model to feature number 32, Chloroplast
## 2025-07-31 12:36:08.267131 INFO::Fitting model to feature number 33, Christensenellaceae
## 2025-07-31 12:36:08.284653 INFO::Fitting model to feature number 34, ChristensenellaceaeR7group
## 2025-07-31 12:36:08.306184 INFO::Fitting model to feature number 35, ClostridiaUCG014
## 2025-07-31 12:36:08.323352 INFO::Fitting model to feature number 36, ClostridiavadinBB60group
## 2025-07-31 12:36:08.32535 INFO::Fitting model to feature number 37, Clostridiumsensustricto1
## 2025-07-31 12:36:08.345458 INFO::Fitting model to feature number 38, Colidextribacter
## 2025-07-31 12:36:08.347227 INFO::Fitting model to feature number 39, CoriobacteriaceaeUCG002
## 2025-07-31 12:36:08.368771 INFO::Fitting model to feature number 40, Corynebacterium
## 2025-07-31 12:36:08.393642 INFO::Fitting model to feature number 41, DNF00809
## 2025-07-31 12:36:08.413845 INFO::Fitting model to feature number 42, DTU089
## 2025-07-31 12:36:08.438027 INFO::Fitting model to feature number 43, DefluviitaleaceaeUCG011
## 2025-07-31 12:36:08.455829 INFO::Fitting model to feature number 44, Desulfovibrio
## 2025-07-31 12:36:08.457606 INFO::Fitting model to feature number 45, Dielma
## 2025-07-31 12:36:08.476767 INFO::Fitting model to feature number 46, Dorea
## 2025-07-31 12:36:08.494834 INFO::Fitting model to feature number 47, Dubosiella
## 2025-07-31 12:36:08.497567 INFO::Fitting model to feature number 48, Eisenbergiella
## 2025-07-31 12:36:08.514399 INFO::Fitting model to feature number 49, Elusimicrobium
## 2025-07-31 12:36:08.530226 INFO::Fitting model to feature number 50, Enterococcus
## 2025-07-31 12:36:08.546372 INFO::Fitting model to feature number 51, Enterorhabdus
## 2025-07-31 12:36:08.548095 INFO::Fitting model to feature number 52, Erysipelatoclostridiaceae
## 2025-07-31 12:36:08.565529 INFO::Fitting model to feature number 53, Erysipelatoclostridium
## 2025-07-31 12:36:08.591036 INFO::Fitting model to feature number 54, ErysipelotrichaceaeUCG003
## 2025-07-31 12:36:08.618261 INFO::Fitting model to feature number 55, EscherichiaShigella
## 2025-07-31 12:36:08.639537 INFO::Fitting model to feature number 56, Eubacteriumbrachygroup
## 2025-07-31 12:36:08.686315 INFO::Fitting model to feature number 57, Eubacteriumcoprostanoligenesgroup
## 2025-07-31 12:36:08.688041 INFO::Fitting model to feature number 58, Eubacteriumnodatumgroup
## 2025-07-31 12:36:08.704529 INFO::Fitting model to feature number 59, Eubacteriumoxidoreducensgroup
## 2025-07-31 12:36:08.721866 INFO::Fitting model to feature number 60, Eubacteriumruminantiumgroup
## 2025-07-31 12:36:08.739962 INFO::Fitting model to feature number 61, Eubacteriumsiraeumgroup
## 2025-07-31 12:36:08.755988 INFO::Fitting model to feature number 62, Eubacteriumventriosumgroup
## 2025-07-31 12:36:08.771851 INFO::Fitting model to feature number 63, Eubacteriumxylanophilumgroup
## 2025-07-31 12:36:08.794868 INFO::Fitting model to feature number 64, F082
## 2025-07-31 12:36:08.814498 INFO::Fitting model to feature number 65, Faecalibaculum
## 2025-07-31 12:36:08.816197 INFO::Fitting model to feature number 66, FamilyXIIIAD3011group
## 2025-07-31 12:36:08.83525 INFO::Fitting model to feature number 67, FamilyXIIIUCG001
## 2025-07-31 12:36:08.855603 INFO::Fitting model to feature number 68, Fournierella
## 2025-07-31 12:36:08.874948 INFO::Fitting model to feature number 69, Fusobacterium
## 2025-07-31 12:36:08.892727 INFO::Fitting model to feature number 70, GCA900066575
## 2025-07-31 12:36:08.911392 INFO::Fitting model to feature number 71, Gastranaerophilales
## 2025-07-31 12:36:08.93083 INFO::Fitting model to feature number 72, Gemella
## 2025-07-31 12:36:08.947594 INFO::Fitting model to feature number 73, Globicatella
## 2025-07-31 12:36:08.967629 INFO::Fitting model to feature number 74, Harryflintia
## 2025-07-31 12:36:08.987838 INFO::Fitting model to feature number 75, Helicobacter
## 2025-07-31 12:36:08.989741 INFO::Fitting model to feature number 76, Hydrogenoanaerobacterium
## 2025-07-31 12:36:09.006941 INFO::Fitting model to feature number 77, Ileibacterium
## 2025-07-31 12:36:09.025903 INFO::Fitting model to feature number 78, IncertaeSedis
## 2025-07-31 12:36:09.027628 INFO::Fitting model to feature number 79, Intestinimonas
## 2025-07-31 12:36:09.046118 INFO::Fitting model to feature number 80, Jeotgalicoccus
## 2025-07-31 12:36:09.068812 INFO::Fitting model to feature number 81, Lachnoclostridium
## 2025-07-31 12:36:09.071754 INFO::Fitting model to feature number 82, LachnospiraceaeFCS020group

```

```

## 2025-07-31 12:36:09.097915 INFO::Fitting model to feature number 83, LachnospiraceaeNC2004group
## 2025-07-31 12:36:09.121997 INFO::Fitting model to feature number 84, LachnospiraceaeND3007group
## 2025-07-31 12:36:09.162217 INFO::Fitting model to feature number 85, LachnospiraceaeNK4A136group
## 2025-07-31 12:36:09.181045 INFO::Fitting model to feature number 86, LachnospiraceaeUCG001
## 2025-07-31 12:36:09.20091 INFO::Fitting model to feature number 87, LachnospiraceaeUCG004
## 2025-07-31 12:36:09.228686 INFO::Fitting model to feature number 88, LachnospiraceaeUCG006
## 2025-07-31 12:36:09.231716 INFO::Fitting model to feature number 89, LachnospiraceaeUCG008
## 2025-07-31 12:36:09.256937 INFO::Fitting model to feature number 90, LachnospiraceaeUCG009
## 2025-07-31 12:36:09.277456 INFO::Fitting model to feature number 91, LachnospiraceaeUCG010
## 2025-07-31 12:36:09.300086 INFO::Fitting model to feature number 92, Lactobacillus
## 2025-07-31 12:36:09.302573 INFO::Fitting model to feature number 93, Leptotrichia
## 2025-07-31 12:36:09.329609 INFO::Fitting model to feature number 94, Marvinbryantia
## 2025-07-31 12:36:09.361529 INFO::Fitting model to feature number 95, Methanosphaera
## 2025-07-31 12:36:09.382836 INFO::Fitting model to feature number 96, Mitochondria
## 2025-07-31 12:36:09.399856 INFO::Fitting model to feature number 97, Monoglobus
## 2025-07-31 12:36:09.401692 INFO::Fitting model to feature number 98, Mucispirillum
## 2025-07-31 12:36:09.419873 INFO::Fitting model to feature number 99, Muribaculaceae
## 2025-07-31 12:36:09.421637 INFO::Fitting model to feature number 100, Muribaculum
## 2025-07-31 12:36:09.423035 INFO::Fitting model to feature number 101, NK4A214group
## 2025-07-31 12:36:09.424442 INFO::Fitting model to feature number 102, Negativibacillus
## 2025-07-31 12:36:09.440283 INFO::Fitting model to feature number 103, Neisseria
## 2025-07-31 12:36:09.457935 INFO::Fitting model to feature number 104, Odoribacter
## 2025-07-31 12:36:09.477333 INFO::Fitting model to feature number 105, Oscillibacter
## 2025-07-31 12:36:09.479421 INFO::Fitting model to feature number 106, Oscillospira
## 2025-07-31 12:36:09.500863 INFO::Fitting model to feature number 107, Paludicola
## 2025-07-31 12:36:09.518867 INFO::Fitting model to feature number 108, Papillibacter
## 2025-07-31 12:36:09.534535 INFO::Fitting model to feature number 109, Parabacteroides
## 2025-07-31 12:36:09.536218 INFO::Fitting model to feature number 110, Paraprevotella
## 2025-07-31 12:36:09.555296 INFO::Fitting model to feature number 111, Parasutterella
## 2025-07-31 12:36:09.557029 INFO::Fitting model to feature number 112, Parvibacter
## 2025-07-31 12:36:09.573645 INFO::Fitting model to feature number 113, Pediococcus
## 2025-07-31 12:36:09.591539 INFO::Fitting model to feature number 114, Peptococcus
## 2025-07-31 12:36:09.608967 INFO::Fitting model to feature number 115, Peptostreptococcus
## 2025-07-31 12:36:09.629542 INFO::Fitting model to feature number 116, Prevotella
## 2025-07-31 12:36:09.631258 INFO::Fitting model to feature number 117, PrevotellaceaeGa6A1group
## 2025-07-31 12:36:09.665732 INFO::Fitting model to feature number 118, PrevotellaceaeNK3B31group
## 2025-07-31 12:36:09.668111 INFO::Fitting model to feature number 119, PrevotellaceaeUCG001
## 2025-07-31 12:36:09.672382 INFO::Fitting model to feature number 120, PrevotellaceaeUCG003
## 2025-07-31 12:36:09.674214 INFO::Fitting model to feature number 121, Proteus
## 2025-07-31 12:36:09.692015 INFO::Fitting model to feature number 122, Pygmaibacter
## 2025-07-31 12:36:09.720558 INFO::Fitting model to feature number 123, Quinella
## 2025-07-31 12:36:09.722928 INFO::Fitting model to feature number 124, RF39
## 2025-07-31 12:36:09.748632 INFO::Fitting model to feature number 125, Rikenella
## 2025-07-31 12:36:09.773925 INFO::Fitting model to feature number 126, RikenellaceaeRC9gutgroup
## 2025-07-31 12:36:09.776937 INFO::Fitting model to feature number 127, Rodentibacter
## 2025-07-31 12:36:09.801282 INFO::Fitting model to feature number 128, Romboutsia
## 2025-07-31 12:36:09.803022 INFO::Fitting model to feature number 129, Roseburia
## 2025-07-31 12:36:09.804524 INFO::Fitting model to feature number 130, Rothia
## 2025-07-31 12:36:09.832945 INFO::Fitting model to feature number 131, Ruminiclostridium
## 2025-07-31 12:36:09.860855 INFO::Fitting model to feature number 132, Ruminococcaceae
## 2025-07-31 12:36:09.887625 INFO::Fitting model to feature number 133, Ruminococcus
## 2025-07-31 12:36:09.88962 INFO::Fitting model to feature number 134, Ruminococcusgavreauigroup
## 2025-07-31 12:36:09.915898 INFO::Fitting model to feature number 135, Ruminococcusstorquesgroup
## 2025-07-31 12:36:09.942856 INFO::Fitting model to feature number 136, Shuttleworthia
## 2025-07-31 12:36:09.963242 INFO::Fitting model to feature number 137, Sphingomonas
## 2025-07-31 12:36:09.981324 INFO::Fitting model to feature number 138, Sporobacter
## 2025-07-31 12:36:10.000061 INFO::Fitting model to feature number 139, Staphylococcus
## 2025-07-31 12:36:10.017167 INFO::Fitting model to feature number 140, Streptococcus
## 2025-07-31 12:36:10.036095 INFO::Fitting model to feature number 141, Subdoligranulum
## 2025-07-31 12:36:10.054676 INFO::Fitting model to feature number 142, Treponema

```

```
## 2025-07-31 12:36:10.08233 INFO::Fitting model to feature number 143, Turicibacter
## 2025-07-31 12:36:10.085233 INFO::Fitting model to feature number 144, Tuzzerella
## 2025-07-31 12:36:10.105379 INFO::Fitting model to feature number 145, Tyzzerella
## 2025-07-31 12:36:10.126873 INFO::Fitting model to feature number 146, UBA1819
## 2025-07-31 12:36:10.162593 INFO::Fitting model to feature number 147, UCG002
## 2025-07-31 12:36:10.181073 INFO::Fitting model to feature number 148, UCG004
## 2025-07-31 12:36:10.218741 INFO::Fitting model to feature number 149, UCG005
## 2025-07-31 12:36:10.220519 INFO::Fitting model to feature number 150, UCG007
## 2025-07-31 12:36:10.236805 INFO::Fitting model to feature number 151, UCG009
## 2025-07-31 12:36:10.255312 INFO::Fitting model to feature number 152, UCG010
## 2025-07-31 12:36:10.276866 INFO::Fitting model to feature number 153, UCG011
## 2025-07-31 12:36:10.297224 INFO::Fitting model to feature number 154, Veillonella
## 2025-07-31 12:36:10.320238 INFO::Fitting model to feature number 155, Victivallis
## 2025-07-31 12:36:10.346294 INFO::Fitting model to feature number 156, uncultured
## 2025-07-31 12:36:10.402408 INFO::Counting total values for each feature
## 2025-07-31 12:36:10.419687 INFO::Re-running abundances for warn_prevalence
## 2025-07-31 12:36:10.420797 INFO::Running selected normalization method: TSS
## 2025-07-31 12:36:10.430952 INFO::Running selected transform method: LOG
## 2025-07-31 12:36:10.462701 INFO::Fitting model to feature number 1, A2
## 2025-07-31 12:36:10.473594 INFO::Fitting model to feature number 2, ASF356
## 2025-07-31 12:36:10.480992 INFO::Fitting model to feature number 3, Acetatifactor
## 2025-07-31 12:36:10.487181 INFO::Fitting model to feature number 4, Actinomyces
## 2025-07-31 12:36:10.49514 INFO::Fitting model to feature number 5, Adlercreutzia
## 2025-07-31 12:36:10.502651 INFO::Fitting model to feature number 6, Aerococcus
## 2025-07-31 12:36:10.510736 INFO::Fitting model to feature number 7, Alistipes
## 2025-07-31 12:36:10.517091 INFO::Fitting model to feature number 8, Allobaculum
## 2025-07-31 12:36:10.525264 INFO::Fitting model to feature number 9, Alloprevotella
## 2025-07-31 12:36:10.532112 INFO::Fitting model to feature number 10, Anaerobiospirillum
## 2025-07-31 12:36:10.539287 INFO::Fitting model to feature number 11, Anaerofilum
## 2025-07-31 12:36:10.545627 INFO::Fitting model to feature number 12, Anaerofustis
## 2025-07-31 12:36:10.552019 INFO::Fitting model to feature number 13, Anaeroplasma
## 2025-07-31 12:36:10.556215 INFO::Fitting model to feature number 14, Anaerotruncus
## 2025-07-31 12:36:10.562591 INFO::Fitting model to feature number 15, Anaerovibrio
## 2025-07-31 12:36:10.570583 INFO::Fitting model to feature number 16, Anaerovorax
## 2025-07-31 12:36:10.577911 INFO::Fitting model to feature number 17, Bacillus
## 2025-07-31 12:36:10.586915 INFO::Fitting model to feature number 18, BacteroidalesRF16group
## 2025-07-31 12:36:10.59553 INFO::Fitting model to feature number 19, Bacteroides
## 2025-07-31 12:36:10.608286 INFO::Fitting model to feature number 20, Barnesiella
## 2025-07-31 12:36:10.615705 INFO::Fitting model to feature number 21, Bifidobacterium
## 2025-07-31 12:36:10.624392 INFO::Fitting model to feature number 22, Blautia
## 2025-07-31 12:36:10.632542 INFO::Fitting model to feature number 23, Butyricicoccus
## 2025-07-31 12:36:10.640255 INFO::Fitting model to feature number 24, Butyricimonas
## 2025-07-31 12:36:10.648436 INFO::Fitting model to feature number 25, Caldicoprobacter
## 2025-07-31 12:36:10.655912 INFO::Fitting model to feature number 26, CandidatusArthromitus
## 2025-07-31 12:36:10.663337 INFO::Fitting model to feature number 27, CandidatusSaccharimonas
## 2025-07-31 12:36:10.672211 INFO::Fitting model to feature number 28, CandidatusSoleaferrea
## 2025-07-31 12:36:10.67884 INFO::Fitting model to feature number 29, CandidatusStoquefichus
## 2025-07-31 12:36:10.685361 INFO::Fitting model to feature number 30, Caproiciproducens
## 2025-07-31 12:36:10.687561 INFO::Fitting model to feature number 31, Catellicoccus
## 2025-07-31 12:36:10.69395 INFO::Fitting model to feature number 32, Chloroplast
## 2025-07-31 12:36:10.70135 INFO::Fitting model to feature number 33, Christensenellaceae
## 2025-07-31 12:36:10.709932 INFO::Fitting model to feature number 34, ChristensenellaceaeR7group
## 2025-07-31 12:36:10.719582 INFO::Fitting model to feature number 35, ClostridiaUCG014
## 2025-07-31 12:36:10.731175 INFO::Fitting model to feature number 36, ClostridiavadinBB60group
## 2025-07-31 12:36:10.740141 INFO::Fitting model to feature number 37, Clostridiumsensustricto1
## 2025-07-31 12:36:10.749729 INFO::Fitting model to feature number 38, Colidextribacter
## 2025-07-31 12:36:10.756365 INFO::Fitting model to feature number 39, CoriobacteriaceaeUCG002
## 2025-07-31 12:36:10.765619 INFO::Fitting model to feature number 40, Corynebacterium
## 2025-07-31 12:36:10.774282 INFO::Fitting model to feature number 41, DNF00809
## 2025-07-31 12:36:10.788155 INFO::Fitting model to feature number 42, DTU089
```

```
## 2025-07-31 12:36:10.791018 INFO::Fitting model to feature number 43, DefluviitaleaceaeUCG011
## 2025-07-31 12:36:10.798229 INFO::Fitting model to feature number 44, Desulfovibrio
## 2025-07-31 12:36:10.804747 INFO::Fitting model to feature number 45, Dielma
## 2025-07-31 12:36:10.807933 INFO::Fitting model to feature number 46, Dorea
## 2025-07-31 12:36:10.81605 INFO::Fitting model to feature number 47, Dubosiella
## 2025-07-31 12:36:10.826187 INFO::Fitting model to feature number 48, Eisenbergiella
## 2025-07-31 12:36:10.835496 INFO::Fitting model to feature number 49, Elusimicrobium
## 2025-07-31 12:36:10.842885 INFO::Fitting model to feature number 50, Enterococcus
## 2025-07-31 12:36:10.849531 INFO::Fitting model to feature number 51, Enterorhabdus
## 2025-07-31 12:36:10.85783 INFO::Fitting model to feature number 52, Erysipelatoclostridiaceae
## 2025-07-31 12:36:10.864253 INFO::Fitting model to feature number 53, Erysipelatoclostridium
## 2025-07-31 12:36:10.866639 INFO::Fitting model to feature number 54, ErysipelotrichaceaeUCG003
## 2025-07-31 12:36:10.869803 INFO::Fitting model to feature number 55, EscherichiaShigella
## 2025-07-31 12:36:10.879547 INFO::Fitting model to feature number 56, Eubacteriumbrachygroup
## 2025-07-31 12:36:10.891625 INFO::Fitting model to feature number 57, Eubacteriumcoprostanoligenesgroup
## 2025-07-31 12:36:10.903317 INFO::Fitting model to feature number 58, Eubacteriumnodatumgroup
## 2025-07-31 12:36:10.914847 INFO::Fitting model to feature number 59, Eubacteriumoxidoreducensgroup
## 2025-07-31 12:36:10.925083 INFO::Fitting model to feature number 60, Eubacteriumruminantiumgroup
## 2025-07-31 12:36:10.928793 INFO::Fitting model to feature number 61, Eubacteriumsiraeumgroup
## 2025-07-31 12:36:10.939787 INFO::Fitting model to feature number 62, Eubacteriumventriosumgroup
## 2025-07-31 12:36:10.95045 INFO::Fitting model to feature number 63, Eubacteriumxylophilumgroup
## 2025-07-31 12:36:10.961093 INFO::Fitting model to feature number 64, F082
## 2025-07-31 12:36:10.971736 INFO::Fitting model to feature number 65, Faecalibaculum
## 2025-07-31 12:36:10.980639 INFO::Fitting model to feature number 66, FamilyXIIIAD3011group
## 2025-07-31 12:36:10.99323 INFO::Fitting model to feature number 67, FamilyXIIIUCG001
## 2025-07-31 12:36:11.001623 INFO::Fitting model to feature number 68, Fournierella
## 2025-07-31 12:36:11.004642 INFO::Fitting model to feature number 69, Fusobacterium
## 2025-07-31 12:36:11.012396 INFO::Fitting model to feature number 70, GCA900066575
## 2025-07-31 12:36:11.019663 INFO::Fitting model to feature number 71, Gastranaerophilales
## 2025-07-31 12:36:11.027025 INFO::Fitting model to feature number 72, Gemella
## 2025-07-31 12:36:11.033999 INFO::Fitting model to feature number 73, Globicatella
## 2025-07-31 12:36:11.042776 INFO::Fitting model to feature number 74, Harryflintia
## 2025-07-31 12:36:11.050855 INFO::Fitting model to feature number 75, Helicobacter
## 2025-07-31 12:36:11.060896 INFO::Fitting model to feature number 76, Hydrogenoanaerobacterium
## 2025-07-31 12:36:11.067946 INFO::Fitting model to feature number 77, Ileibacterium
## 2025-07-31 12:36:11.074631 INFO::Fitting model to feature number 78, IncertaeSedis
## 2025-07-31 12:36:11.081217 INFO::Fitting model to feature number 79, Intestinimonas
## 2025-07-31 12:36:11.090699 INFO::Fitting model to feature number 80, Jeotgalicoccus
## 2025-07-31 12:36:11.093904 INFO::Fitting model to feature number 81, Lachnoclostridium
## 2025-07-31 12:36:11.101339 INFO::Fitting model to feature number 82, LachnospiraceaeFCS020group
## 2025-07-31 12:36:11.109168 INFO::Fitting model to feature number 83, LachnospiraceaeNC2004group
## 2025-07-31 12:36:11.118472 INFO::Fitting model to feature number 84, LachnospiraceaeND3007group
## 2025-07-31 12:36:11.121305 INFO::Fitting model to feature number 85, LachnospiraceaeNK4A136group
## 2025-07-31 12:36:11.12856 INFO::Fitting model to feature number 86, LachnospiraceaeUCG001
## 2025-07-31 12:36:11.137046 INFO::Fitting model to feature number 87, LachnospiraceaeUCG004
## 2025-07-31 12:36:11.146353 INFO::Fitting model to feature number 88, LachnospiraceaeUCG006
## 2025-07-31 12:36:11.160629 INFO::Fitting model to feature number 89, LachnospiraceaeUCG008
## 2025-07-31 12:36:11.173729 INFO::Fitting model to feature number 90, LachnospiraceaeUCG009
## 2025-07-31 12:36:11.178404 INFO::Fitting model to feature number 91, LachnospiraceaeUCG010
## 2025-07-31 12:36:11.18762 INFO::Fitting model to feature number 92, Lactobacillus
## 2025-07-31 12:36:11.199129 INFO::Fitting model to feature number 93, Leptotrichia
## 2025-07-31 12:36:11.208476 INFO::Fitting model to feature number 94, Marvinbryantia
## 2025-07-31 12:36:12.040958 INFO::Fitting model to feature number 95, Methanosphaera
## 2025-07-31 12:36:12.044058 INFO::Fitting model to feature number 96, Mitochondria
## 2025-07-31 12:36:12.050935 INFO::Fitting model to feature number 97, Monoglobus
## 2025-07-31 12:36:12.057351 INFO::Fitting model to feature number 98, Mucispirillum
## 2025-07-31 12:36:12.06447 INFO::Fitting model to feature number 99, Muribaculaceae
## 2025-07-31 12:36:12.072166 INFO::Fitting model to feature number 100, Muribaculum
## 2025-07-31 12:36:12.079086 INFO::Fitting model to feature number 101, NK4A214group
## 2025-07-31 12:36:12.0857 INFO::Fitting model to feature number 102, Negativibacillus
```

```
## 2025-07-31 12:36:12.092288 INFO::Fitting model to feature number 103, Neisseria
## 2025-07-31 12:36:12.101494 INFO::Fitting model to feature number 104, Odoribacter
## 2025-07-31 12:36:12.108167 INFO::Fitting model to feature number 105, Oscillibacter
## 2025-07-31 12:36:12.115221 INFO::Fitting model to feature number 106, Oscillospira
## 2025-07-31 12:36:12.121847 INFO::Fitting model to feature number 107, Paludicola
## 2025-07-31 12:36:12.130295 INFO::Fitting model to feature number 108, Papillibacter
## 2025-07-31 12:36:12.137124 INFO::Fitting model to feature number 109, Parabacteroides
## 2025-07-31 12:36:12.146073 INFO::Fitting model to feature number 110, Paraprevotella
## 2025-07-31 12:36:12.148607 INFO::Fitting model to feature number 111, Parasutterella
## 2025-07-31 12:36:12.154829 INFO::Fitting model to feature number 112, Parvibacter
## 2025-07-31 12:36:12.161392 INFO::Fitting model to feature number 113, Pediococcus
## 2025-07-31 12:36:12.16355 INFO::Fitting model to feature number 114, Peptococcus
## 2025-07-31 12:36:12.169619 INFO::Fitting model to feature number 115, Peptostreptococcus
## 2025-07-31 12:36:12.171714 INFO::Fitting model to feature number 116, Prevotella
## 2025-07-31 12:36:12.180845 INFO::Fitting model to feature number 117, PrevotellaceaeGa6A1group
## 2025-07-31 12:36:12.187403 INFO::Fitting model to feature number 118, PrevotellaceaeNK3B31group
## 2025-07-31 12:36:12.194666 INFO::Fitting model to feature number 119, PrevotellaceaeUCG001
## 2025-07-31 12:36:12.200824 INFO::Fitting model to feature number 120, PrevotellaceaeUCG003
## 2025-07-31 12:36:12.208508 INFO::Fitting model to feature number 121, Proteus
## 2025-07-31 12:36:12.21677 INFO::Fitting model to feature number 122, Pygmaibacter
## 2025-07-31 12:36:12.224937 INFO::Fitting model to feature number 123, Quinella
## 2025-07-31 12:36:12.2318 INFO::Fitting model to feature number 124, RF39
## 2025-07-31 12:36:12.238769 INFO::Fitting model to feature number 125, Rikenella
## 2025-07-31 12:36:12.245417 INFO::Fitting model to feature number 126, RikenellaceaeRC9gutgroup
## 2025-07-31 12:36:12.25193 INFO::Fitting model to feature number 127, Rodentibacter
## 2025-07-31 12:36:12.26029 INFO::Fitting model to feature number 128, Romboutsia
## 2025-07-31 12:36:12.26702 INFO::Fitting model to feature number 129, Roseburia
## 2025-07-31 12:36:12.274012 INFO::Fitting model to feature number 130, Rothia
## 2025-07-31 12:36:12.280445 INFO::Fitting model to feature number 131, Ruminiclostridium
## 2025-07-31 12:36:12.288254 INFO::Fitting model to feature number 132, Ruminococcaceae
## 2025-07-31 12:36:12.296774 INFO::Fitting model to feature number 133, Ruminococcus
## 2025-07-31 12:36:12.305872 INFO::Fitting model to feature number 134, Ruminococcusgavreuiiigroup
## 2025-07-31 12:36:12.308368 INFO::Fitting model to feature number 135, Ruminococcusstorquesgroup
## 2025-07-31 12:36:12.317867 INFO::Fitting model to feature number 136, Shuttleworthia
## 2025-07-31 12:36:12.326115 INFO::Fitting model to feature number 137, Sphingomonas
## 2025-07-31 12:36:12.33381 INFO::Fitting model to feature number 138, Sporobacter
## 2025-07-31 12:36:12.34045 INFO::Fitting model to feature number 139, Staphylococcus
## 2025-07-31 12:36:12.347519 INFO::Fitting model to feature number 140, Streptococcus
## 2025-07-31 12:36:12.354076 INFO::Fitting model to feature number 141, Subdoligranulum
## 2025-07-31 12:36:12.360809 INFO::Fitting model to feature number 142, Treponema
## 2025-07-31 12:36:12.367357 INFO::Fitting model to feature number 143, Turicibacter
## 2025-07-31 12:36:12.373786 INFO::Fitting model to feature number 144, Tuzzerella
## 2025-07-31 12:36:12.381202 INFO::Fitting model to feature number 145, Tyzzerella
## 2025-07-31 12:36:12.387651 INFO::Fitting model to feature number 146, UBA1819
## 2025-07-31 12:36:12.394863 INFO::Fitting model to feature number 147, UCG002
## 2025-07-31 12:36:12.401429 INFO::Fitting model to feature number 148, UCG004
## 2025-07-31 12:36:12.409244 INFO::Fitting model to feature number 149, UCG005
## 2025-07-31 12:36:12.416662 INFO::Fitting model to feature number 150, UCG007
## 2025-07-31 12:36:12.424289 INFO::Fitting model to feature number 151, UCG009
## 2025-07-31 12:36:12.430813 INFO::Fitting model to feature number 152, UCG010
## 2025-07-31 12:36:12.439253 INFO::Fitting model to feature number 153, UCG011
## 2025-07-31 12:36:12.441475 INFO::Fitting model to feature number 154, Veillonella
## 2025-07-31 12:36:12.448067 INFO::Fitting model to feature number 155, Victivallis
## 2025-07-31 12:36:12.451186 INFO::Fitting model to feature number 156, uncultured
## 2025-07-31 12:36:12.520266 WARNING::Deleting existing residuals file: C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut microbiome analysis/Preprocessed GM data/Analysis WP2020-0022-015/Genus/fits/residuals_linear.rds
## 2025-07-31 12:36:12.522159 INFO::Writing residuals to file C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut microbiome analysis/Preprocessed GM data/Analysis WP2020-0022-015/Genus/fits/residuals_linear.rds
## 2025-07-31 12:36:12.527479 WARNING::Deleting existing fitted file: C:/Users/roger/OneDrive/Desktop/Ph.D. Cogniti
```

```

ve Neuroscience/Gut microbiome analysis/Preprocessed GM data/Analysis WP2020-0022-015/Genus/fits/fitted_linear.rds
## 2025-07-31 12:36:12.529949 INFO::Writing fitted values to file C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive N
euroscience/Gut microbiome analysis/Preprocessed GM data/Analysis WP2020-0022-015/Genus/fits/fitted_linear.rds
## 2025-07-31 12:36:12.534102 WARNING::Deleting existing residuals file: C:/Users/roger/OneDrive/Desktop/Ph.D. Cogni
tive Neuroscience/Gut microbiome analysis/Preprocessed GM data/Analysis WP2020-0022-015/Genus/fits/residuals_logis
tic.rds
## 2025-07-31 12:36:12.536135 INFO::Writing residuals to file C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuro
science/Gut microbiome analysis/Preprocessed GM data/Analysis WP2020-0022-015/Genus/fits/residuals_logistic.rds
## 2025-07-31 12:36:12.543262 WARNING::Deleting existing fitted file: C:/Users/roger/OneDrive/Desktop/Ph.D. Cogniti
ve Neuroscience/Gut microbiome analysis/Preprocessed GM data/Analysis WP2020-0022-015/Genus/fits/fitted_logistic.rd
s
## 2025-07-31 12:36:12.545467 INFO::Writing fitted values to file C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive N
euroscience/Gut microbiome analysis/Preprocessed GM data/Analysis WP2020-0022-015/Genus/fits/fitted_logistic.rds
## 2025-07-31 12:36:12.554914 INFO::Writing all the results to file (ordered
##           by increasing individual q-values): C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut
microbiome analysis/Preprocessed GM data/Analysis WP2020-0022-015/Genus/all_results.tsv
## 2025-07-31 12:36:12.574754 INFO::Writing the significant results without errors (those which have joint q-values
less than or equal to the threshold of 0.050000 ) to file (ordered by increasing individual q-values): C:/Users/rog
er/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut microbiome analysis/Preprocessed GM data/Analysis WP2020-0022-
015/Genus/significant_results.tsv
## 2025-07-31 12:36:12.578424 INFO::Writing summary plot of significant
##           results to file: C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut microb
iome analysis/Preprocessed GM data/Analysis WP2020-0022-015/Genus/figures/summary_plot.pdf
## 2025-07-31 12:36:12.580609 INFO::No associations were without errors.
##           No summary plot generated.
## 2025-07-31 12:36:12.581571 INFO::Writing association plots (one for each significant association) to output fold
er: C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut microbiome analysis/Preprocessed GM data/Analy
sis WP2020-0022-015/Genus/figures
## 2025-07-31 12:36:12.58342 INFO::Plotting associations from most to least significant, grouped by metadata
## 2025-07-31 12:36:12.587448 INFO::Creating box plot for categorical data (linear),
##           Sex vs Faecalibaculum
## 2025-07-31 12:36:12.607399 INFO::Creating box plot for categorical data (linear),
##           Sex vs Ileibacterium
## 2025-07-31 12:36:12.627292 INFO::Creating box plot for categorical data (linear),
##           Genotype vs Corynebacterium
## 2025-07-31 12:36:12.643892 INFO::Creating tile plot for categorical data (logistic), Genotype vs Ruminococcusga
uvreaiigroup
## 2025-07-31 12:36:12.667702 INFO::Creating box plot for categorical data (linear),
##           Sex vs Enterorhabdus
## 2025-07-31 12:36:12.683647 INFO::Creating box plot for categorical data (linear),
##           Sex vs Muribaculum
## 2025-07-31 12:36:12.698923 INFO::Creating box plot for categorical data (linear),
##           Sex vs Peptococcus
## 2025-07-31 12:36:12.716603 INFO::Creating box plot for categorical data (linear),
##           Genotype_Sex vs A2
## 2025-07-31 12:36:12.73263 INFO::Creating tile plot for categorical data (logistic), Sex vs Staphylococcus
## 2025-07-31 12:36:12.756553 INFO::Creating box plot for categorical data (linear),
##           Genotype vs Allobaculum
## 2025-07-31 12:36:12.773282 INFO::Creating box plot for categorical data (linear),
##           Sex vs Bifidobacterium
## 2025-07-31 12:36:12.789213 INFO::Creating box plot for categorical data (linear),
##           Sex vs Clostridiumsensustricto1
## 2025-07-31 12:36:12.805695 INFO::Creating box plot for categorical data (linear),
##           Sex vs CoriobacteriaceaeUCG002
## 2025-07-31 12:36:12.838689 INFO::Creating box plot for categorical data (linear),
##           Sex vs Corynebacterium
## 2025-07-31 12:36:12.863459 INFO::Creating box plot for categorical data (linear),
##           Genotype vs ClostridiavadinBB60group
## 2025-07-31 12:36:12.885713 INFO::Creating box plot for categorical data (linear),
##           Sex vs Anaerobiospirillum

```

```
## 2025-07-31 12:36:12.91158 INFO::Creating box plot for categorical data (linear),
##                               Genotype vs Rothia
## 2025-07-31 12:36:12.931089 INFO::Creating tile plot for categorical data (logistic), Genotype vs CandidatusSole
aferrea
## 2025-07-31 12:36:12.958561 INFO::Creating box plot for categorical data (linear),
##                               Genotype vs Caldicoprobacter
```

Load and view significant results

```
signif_results_015_genus <- read.delim("C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut microbiome a
head(signif_results_015_genus)
```

```
##                               feature metadata    value    name coef null_hypothesis
## 1                               Faecalibaculum    Sex      Sex      Sex      NA      NA
## 2                               Ileibacterium    Sex      Sex      Sex      NA      NA
## 3                               Corynebacterium Genotype Genotype Genotype  NA      NA
## 4 Ruminococcusgavreuiigroup Genotype Genotype Genotype  NA      0
## 5                               Enterorhabdus    Sex      Sex      Sex      NA      NA
## 6                               Muribaculum      Sex      Sex      Sex      NA      NA
##      stderr pval_individual qval_individual  pval_joint  qval_joint error
## 1      NA      6.891677e-09    5.292808e-06 6.891677e-09 3.225305e-06 <NA>
## 2      NA      2.524382e-08    9.693626e-06 5.048763e-08 1.181411e-05 <NA>
## 3      NA      3.385841e-06    6.500814e-04 6.771670e-06 7.922854e-04 <NA>
## 4      NA      3.011524e-06    6.500814e-04 3.011524e-06 4.697978e-04 <NA>
## 5      NA      1.271359e-05    1.952808e-03 1.271359e-05 1.189992e-03 <NA>
## 6      NA      3.417908e-05    4.374923e-03 3.417908e-05 2.665968e-03 <NA>
##      model  N N_not_zero
## 1 abundance 97      97
## 2 abundance 97      82
## 3 abundance 97      80
## 4 prevalence 97     14
## 5 abundance 97      97
## 6 abundance 97      97
```

```
all_results_015_genus <- read.delim("C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut microbiome anal
```

4.2 Significant genera

Extract significant feature (taxa) names

```
sig_taxa <- as.character(signif_results_015_genus$feature)
sig_taxa <- trimws(sig_taxa) # Remove any accidental whitespace
```

Load normalized and transformed abundance data

```
final_data <- read.delim("C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut microbiome analysis/Prep
```

Clean up both sets just

```
sig_taxa <- trimws(sig_taxa)
final_data_cols <- trimws(colnames(final_data))
```

Find overlap

```
matching_cols <- final_data_cols[final_data_cols %in% sig_taxa]
```

Subset normalized data to only the significant taxa

```
sig_norm_data <- final_data[, matching_cols, drop = FALSE]
```

```
# Check if anything was matched and export data
if (length(matching_cols) == 0) {
  message("No matching taxa found between normalized data and significant features.")
} else {
  write.csv(sig_norm_data, file = "C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut microbiome analys
}
```

5 Visualization

5.1 Group averages

```
# Load necessary package for data manipulation and plotting
library(tidyr)

# Get Significant Genera from MaAsLin3
sig_genera <- unique(signif_results_015_genus$feature)

# Compute Relative Abundance per Sample
rel_abund <- feature_table_015 / rowSums(feature_table_015)
rel_abund <- as.data.frame(rel_abund)
rel_abund$SampleID <- rownames(rel_abund) # Add SampleID to the data frame

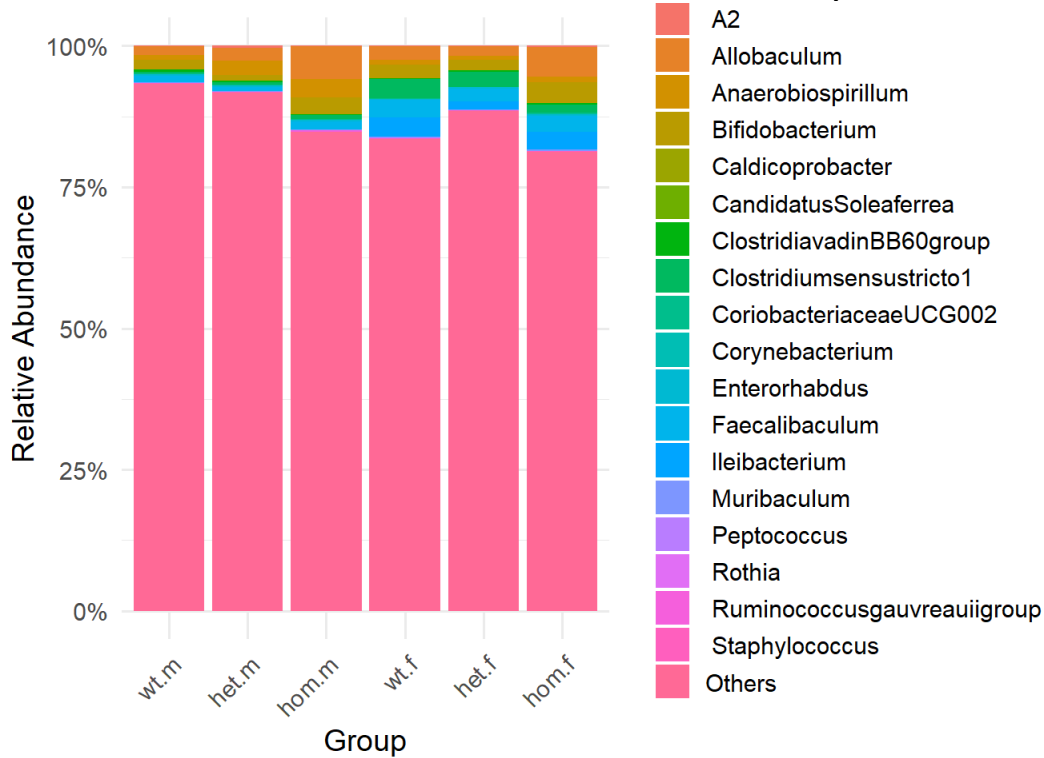
# Reshape Data and Label Non-Significant Genera as "Others"
rel_abund_long <- rel_abund %>%
  pivot_longer(cols = -SampleID, names_to = "Genus", values_to = "RelAbundance") %>%
  mutate(Genus = ifelse(Genus %in% sig_genera, Genus, "Others"))

# Aggregate the data by SampleID and Genus
rel_abund_collapsed <- rel_abund_long %>%
  group_by(SampleID, Genus) %>%
  summarise(RelAbundance = sum(RelAbundance), .groups = "drop")

# Merge with Metadata and Assign Group Labels
rel_abund_collapsed <- rel_abund_collapsed %>%
  left_join(metadata_015 %>% mutate(SampleID = rownames(metadata_015)), by = "SampleID") %>%
  mutate(Group = Genotype_Sex)

# Calculate Average Abundance per Group
plot_data <- rel_abund_collapsed %>%
  group_by(Group, Genus) %>%
  summarise(MeanRelAbund = mean(RelAbundance), .groups = "drop")

# Plot Group Average Data
ggplot(plot_data, aes(x = Group, y = MeanRelAbund, fill = Genus)) +
  geom_bar(stat = "identity", position = "stack") +
  scale_y_continuous(labels = scales::percent_format(accuracy = 1)) + # Format y-axis as percentage
  labs(x = "Group", y = "Relative Abundance", fill = "Genus") + # Labels and Legend
  theme_minimal(base_size = 14) + # Minimal theme for the plot
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) # Rotate x-axis labels for clarity
```



Export to CSV

```
write.csv(plot_data, file = "C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut microbiome analysis/Pre
```

5.2 Individual samples

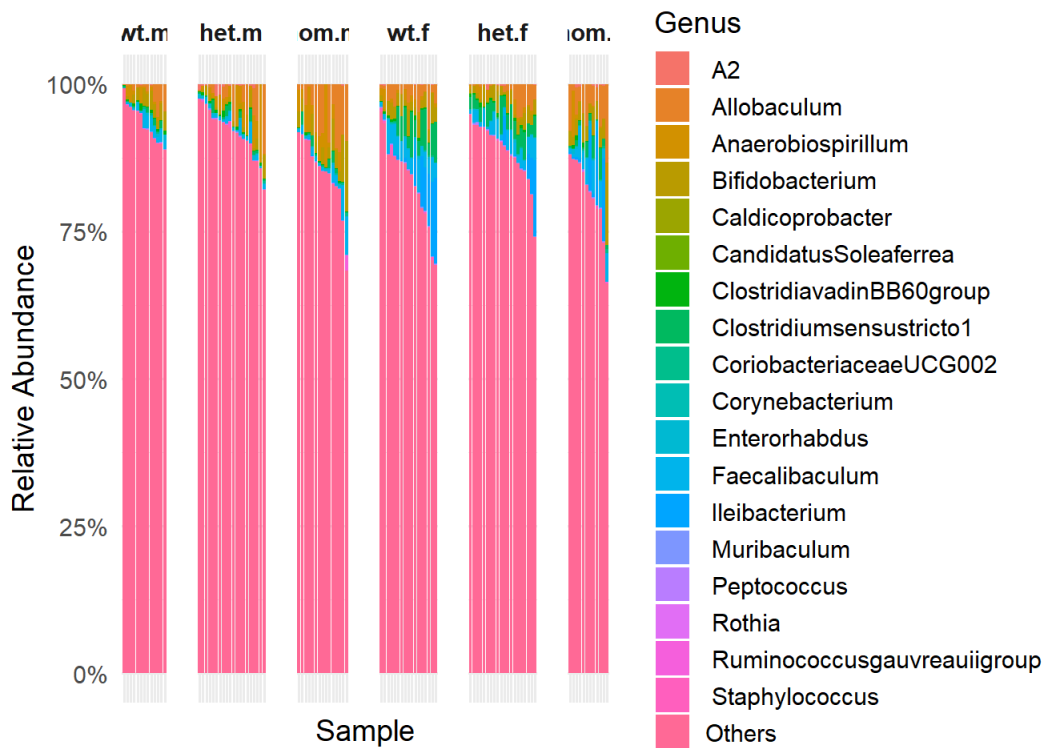
Determine Sample Order by "Others" within Each Group

```
sample_order <- rel_abund_collapsed %>%
  filter(Genus == "Others") %>%
  arrange(Group, desc(RelAbundance)) %>%
  pull(SampleID)
```

```
rel_abund_collapsed$SampleID <- factor(rel_abund_collapsed$SampleID, levels = sample_order)
```

Plot Individual Sample Data

```
ggplot(rel_abund_collapsed, aes(x = SampleID, y = RelAbundance, fill = Genus)) +
  geom_bar(stat = "identity", position = "stack") +
  facet_grid(~Group, scales = "free_x", space = "free_x") + # Create a separate facet for each group
  scale_y_continuous(labels = scales::percent_format(accuracy = 1)) + # Format y-axis as percentage
  labs(x = "Sample", y = "Relative Abundance", fill = "Genus") + # Labels and Legend
  theme_minimal(base_size = 14) + # Minimal theme for the plot
  theme(
    axis.text.x = element_blank(), # Remove x-axis text for clarity
    axis.ticks.x = element_blank(), # Remove x-axis ticks
    strip.text = element_text(face = "bold"), # Bold facet titles
    panel.spacing = unit(1, "lines") # Increase spacing between facets
  )
```



Export to CSV

```
write.csv(rel_abund_collapsed[, 1:3], "C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut microbiome an
```

5.3 Density plots

Load required libraries

```
library(ggplot2)
```

```
library(tidyr)
```

```
library(dplyr)
```

```
library(scales)
```

```
library(writexl)
```

=== 1. Read significant genera from MaAsLin3 ===

```
signif_results_015_genus <- read.delim("C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut microbiome a
```

```
significant_genera <- unique(signif_results_015_genus$feature)
```

=== 2. Filter abundance matrix to significant genera ===

```
feature_table_genus_filtered <- t(feature_table_015) # genera x samples
```

```
feature_table_sig <- feature_table_genus_filtered[rownames(feature_table_genus_filtered) %in% significant_genera, ]
```

```
feature_table_sig_t <- t(feature_table_sig) # samples x genera
```

=== 3. Normalize to relative abundance ===

```
feature_table_relabund <- apply(feature_table_sig_t, 1, function(x) x / sum(x))
```

```
feature_table_relabund <- t(feature_table_relabund)
```

Create dataframe and merge with metadata

```
abund_df <- as.data.frame(feature_table_relabund)
```

```
abund_df$sample_id <- rownames(abund_df)
```

```
metadata_015$sample_id <- rownames(metadata_015)
```

```
abund_df <- merge(abund_df, metadata_015, by = "sample_id")
```

Pivot to Long format

```

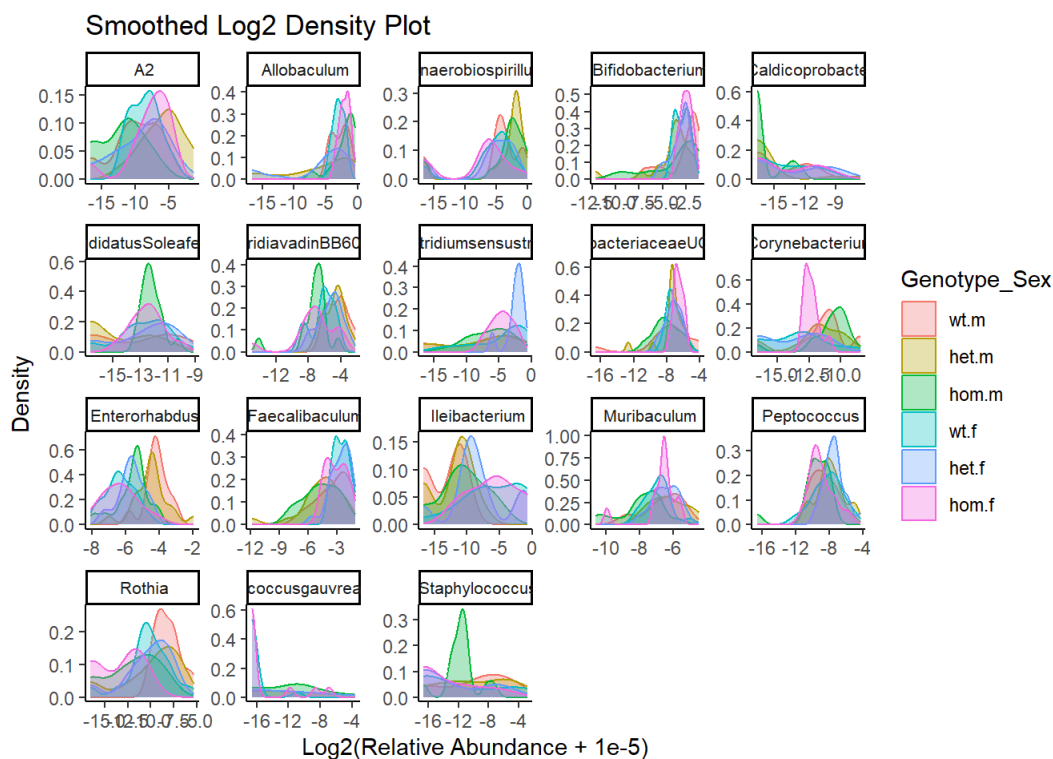
abund_long <- pivot_longer(
  abund_df,
  cols = all_of(significant_genera),
  names_to = "genus",
  values_to = "abund"
)

# === 4. Log2 transformation ===
abund_long$log2_abund <- log2(abund_long$abund + 1e-5) # Avoid Log(0)

# === 5. Log2 density plot ===
p_log2 <- ggplot(abund_long, aes(x = log2_abund, color = Genotype_Sex, fill = Genotype_Sex)) +
  geom_density(alpha = 0.3, adjust = 1) +
  facet_wrap(~ genus, scales = "free") +
  labs(
    x = "Log2(Relative Abundance + 1e-5)",
    y = "Density",
    title = "Smoothed Log2 Density Plot"
  ) +
  theme_classic() +
  theme(strip.text = element_text(size = 8))

# Show plot
print(p_log2)

```



```

# Create a list to store data frames per genus
density_list <- list()

# Loop over each genus
for (genus_name in unique(abund_long$genus)) {

  # Subset data for one genus
  genus_data <- abund_long %>% filter(genus == genus_name)
}

```

```
# List to hold each Genotype_Sex group's density data
group_density <- list()

# Loop over Genotype_Sex groups
for (group in unique(genus_data$Genotype_Sex)) {
  group_data <- genus_data %>% filter(Genotype_Sex == group)

  # Compute density on log2_abund
  dens <- density(group_data$log2_abund, adjust = 1)

  # Create a data frame with x and y values
  dens_df <- data.frame(
    log2_abund = dens$x,
    density = dens$y,
    Genotype_Sex = group
  )

  # Store in list
  group_density[[group]] <- dens_df
}

# Combine all groups for this genus
genus_density_df <- bind_rows(group_density)

# Add to final list
density_list[[genus_name]] <- genus_density_df
}

# Write all genera to an Excel file (each as a separate sheet)
write_xlsx(density_list, path = "C:/Users/roger/OneDrive/Desktop/Ph.D. Cognitive Neuroscience/Gut microbiome analysis")
```