

OUTPOST

OUTPOST REPORT

1. Configuration

Methods

This file is filled by user, which records the information passed to OUTPOST.

Results

The result is in /OUTPOST/OUTPOST_config.tsv.

Visualization example1

Path: OUTPOST/OUTPOST_config.tsv

*samples: identifier for each sample.

*fq_dir: directory location of associated fastq files.

*bam_dir: directory location of associated bam files.

*assembly: name of the assembly used.

*assembly_dir: directory location of assembly files.

*group: grouping category or classification of the samples.

*batch: batch of the samples.

samples	fq_dir	bam_dir	assembly	assembly_dir	group	batch
111030_normal	/home/yihang/data/111030_normal_onlyPE_noncatvirus.fastq	/home/yihang/bam Sam/111030_normal_contigs_sorted.cat.bam	cat	/home/yihang/refgenomes/cat_microbiome/GCA_022675345.1_ASM2267534v1_genomic.fna	normal	1
111041_normal	/home/yihang/data/111041_normal_onlyPE_noncatvirus.fastq	/home/yihang/bam Sam/111041_normal_contigs_sorted.cat.bam	cat	/home/yihang/refgenomes/cat_microbiome/GCA_022675345.1_ASM2267534v1_genomic.fna	normal	1
111468_normal	/home/yihang/data/111468_normal_onlyPE_noncatvirus.fastq	/home/yihang/bam Sam/111468_normal_contigs_sorted.cat.bam	cat	/home/yihang/refgenomes/cat_microbiome/GCA_022675345.1_ASM2267534v1_genomic.fna	normal	1

2. Assembly analysis

Methods

For each group pair, OUTPOST calculates the P-value and q-value of all contigs based on counts, allowing users to manually choose contigs for further MAG (metagenome-assembled genome) assembly based on statistical results and counts. OUTPOST offers Bonferroni, Bonferroni-Holm, and Benjamini-Hochberg corrections.

Results

The result is in the directory `cat/assembly_analysis/`. This module is used to describe the profiling of factors and microbiota at various taxonomic levels, as well as to calculate statistical information for taxa.

Visualization example1

Path: cat/assembly_analysis/cat.normal_vs_obese.contig_table.tsv

*contigID: the name of each contig.

*taxaID: NCBI taxonomy ID.

*(superkingdom-species) columns: various levels, including superkingdom, phylum, class, order, family, genus, species.

*sample_count: the mapping count of each sample.

*pvalue: the pvalue between two group.

*qvalue: the qvalue between two group.

contigID	taxaID	superkingdom	phylum	class	order	family
JAIZPE010000 001.1	1262830	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae
JAIZPE010000 002.1	1262792	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae
JAIZPE010000 003.1	1735	Bacteria	Firmicutes	Erysipelotrichi a	Erysipelotricha les	Erysipelotricha ceae

Visualization example2

cat/assembly_analysis/metagenemark/cat.nucl.fa: predicted genes from used microbiome assembly.

cat/assembly_analysis/metagenemark/cat.nucl.nonrd.fa: non-redundant predicted genes from used microbiome assembly.

cat/assembly_analysis/metagenemark/cat.prot.fa: predicted proteins from used microbiome assembly.

cat.gtf : the annotation of predicted open reading frame.

3. Taxonomy Analysis

Methods

OUTPOST uses the idxstats function in SAMtools to summarize the counts for every microbial contig across samples. OUTPOST then employs Kaiju to annotate the taxonomy of contigs at various levels, including superkingdom, phylum, class, order, family, genus, species, and NCBI taxonomy ID. Users can optionally remove the batch effects by CPM (counts per million reads) normalization given the batch information. For group pairs, the taxonomy analysis was further conducted to assess the differences between them. Here, the criterion for defining significance with a P-value < 0.05.

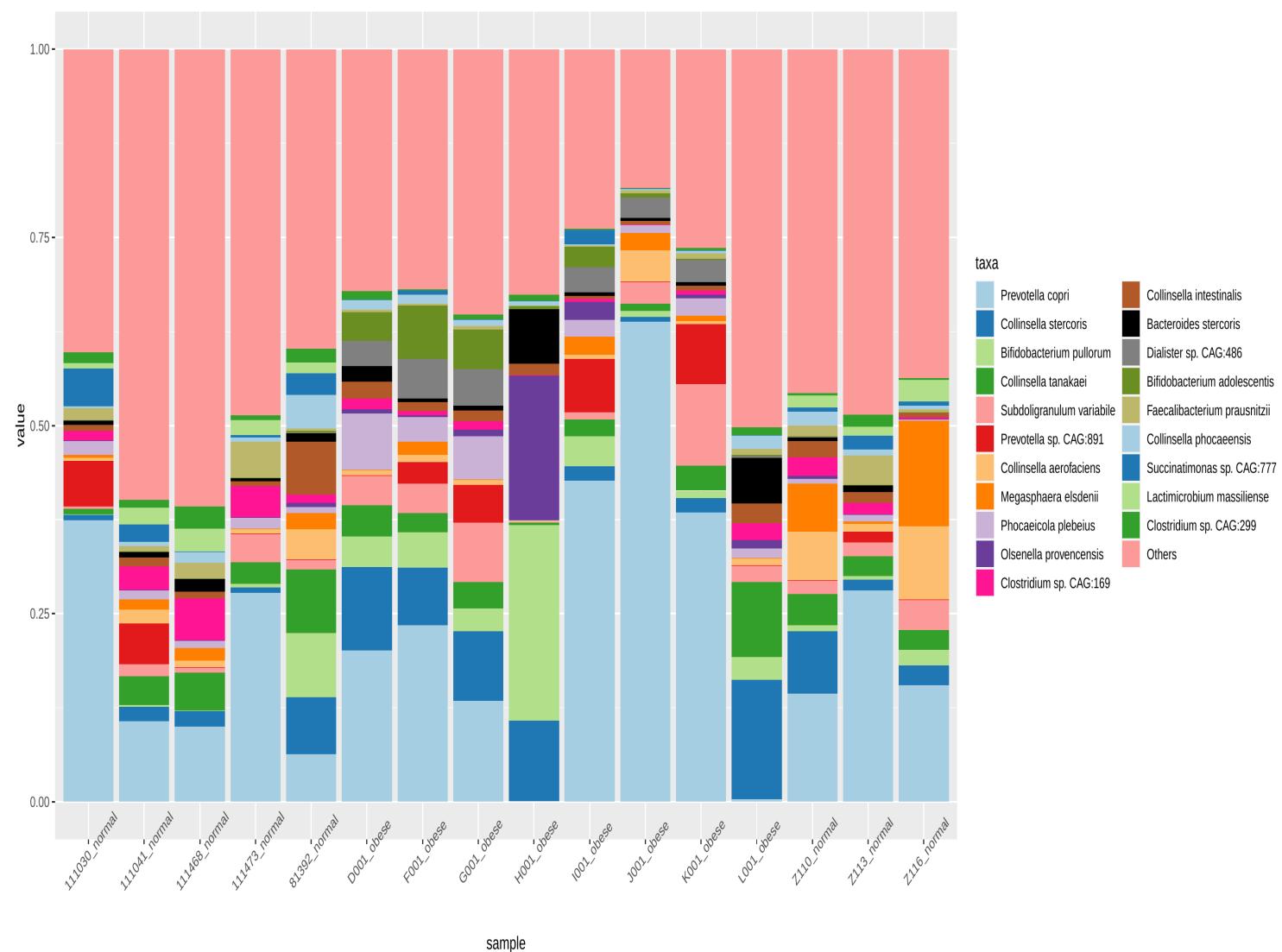
Results

The results are in the directory cat/taxonomy_analysis. This analytical module helps researchers understand the composition of samples at different taxonomic levels. If the samples are derived from different groups, this analysis can also provide insights into the differences between these groups. The data is presented in the form of a bar plot or a heatmap plot.

Visualization example1

Path: cat/taxonomy_analysis/figs/cat.taxa_counts.rel_abun.species.rmU.top20.barplot.pdf

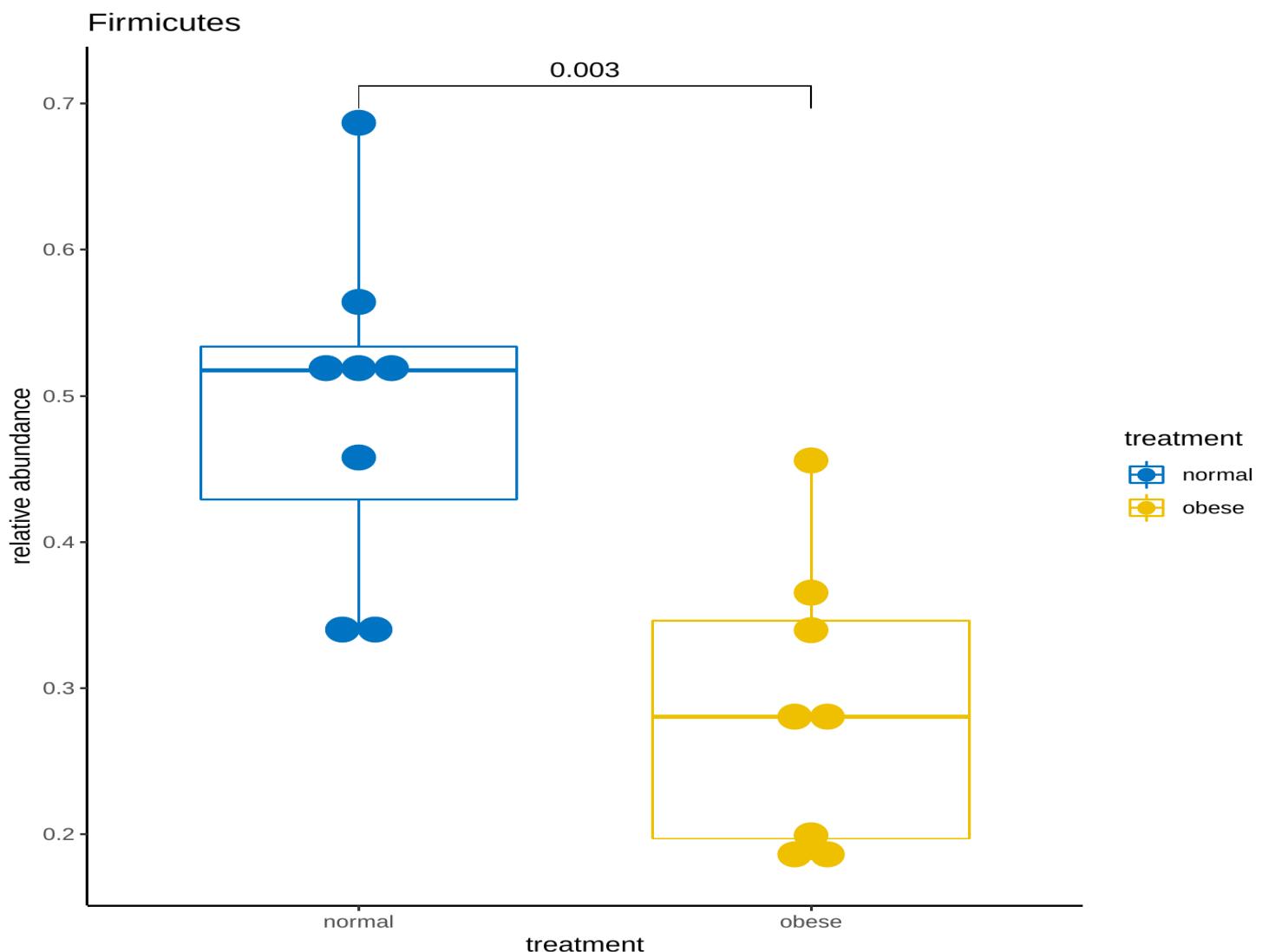
The following bar plot presents the relative frequency at the species level. The x-axis represents different samples, and the y-axis represents relative abundance. In the graph, species outside the top 20 are combined as "others" and displayed. Different colors mean different species.



Visualization example2

Path: cat/taxonomy_analysis/boxplot_normal_vs_obese/phylum/cat.rel_abun.normal_vs_obese.at_phylum.rel_abun.unequal.Firmicutes.boxplot.pdf

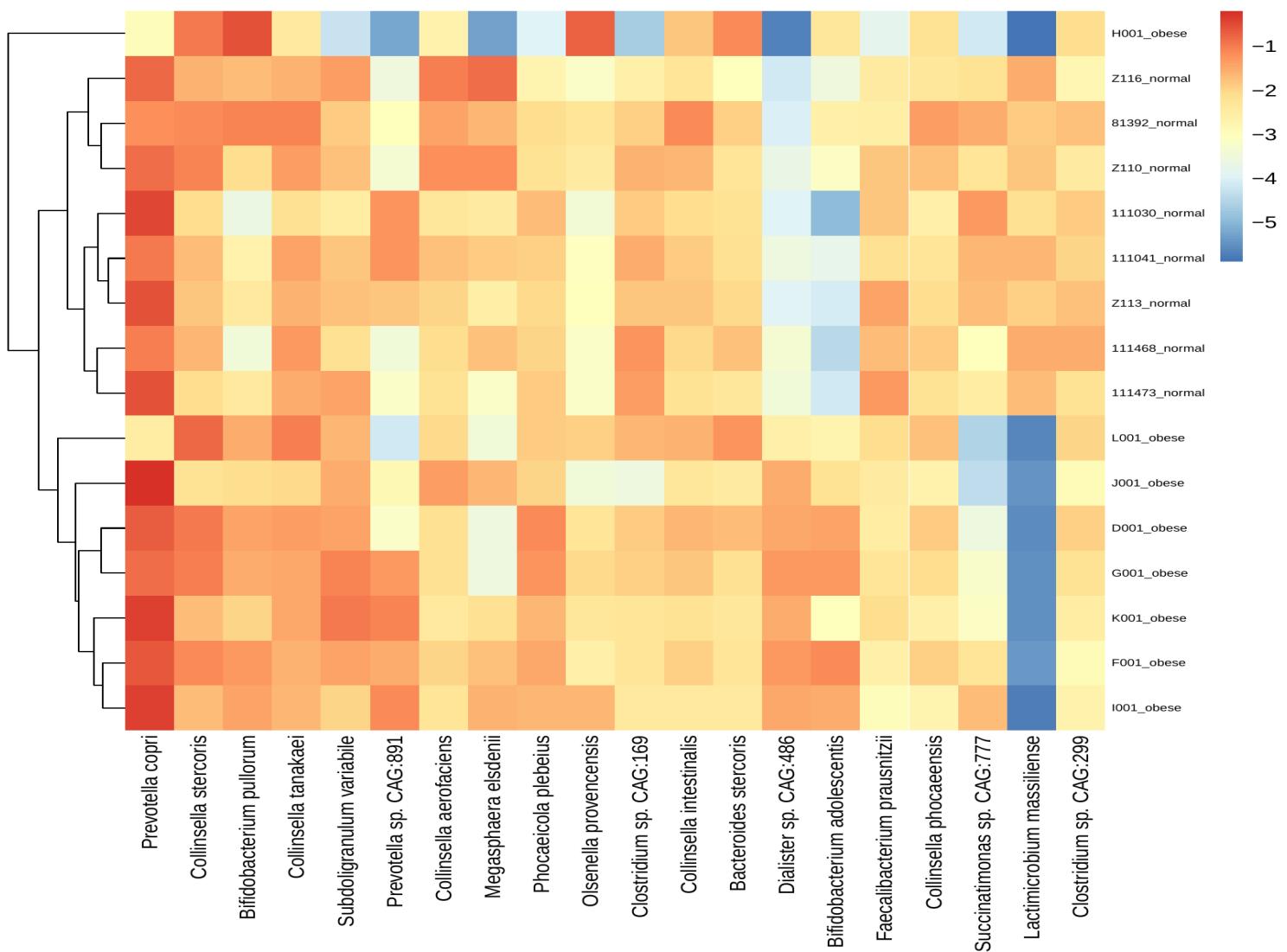
The following box plot is used to present the variations in abundance of each microbial species in different group at the species level. The x-axis represents different samples, and the y-axis represents the relative abundance. Statistical significance was calculated using two-tailed Mann-Whitney U test.



Visualization example3

Path: cat/taxonomy_analysis/figs/cat.taxa_counts.rel_abun.species.rmU.top20.fillmin.scaled.heatmap.pdf

We can also examine the top 20 differentially abundant microbial species between the groups at the genus level in the following heatmap plot. The x-axis represents top20 differential genera, and the y-axis represents different samples. The color scale represents the relative abundance of species.



4. Diversity Analysis

Methods

OUTPOST calculates the diversity based on taxonomy relative abundance at species and genus levels using vegan and self-scripts. The alpha diversities include the Simpson index, Shannon index, Pielou index, and their derivates. The beta diversities include the Bray-Curtis index and the Jaccard index. OUTPOST generates box plots for alpha diversity and principal coordinate analysis (PCoA) plots for beta diversity with P-value.

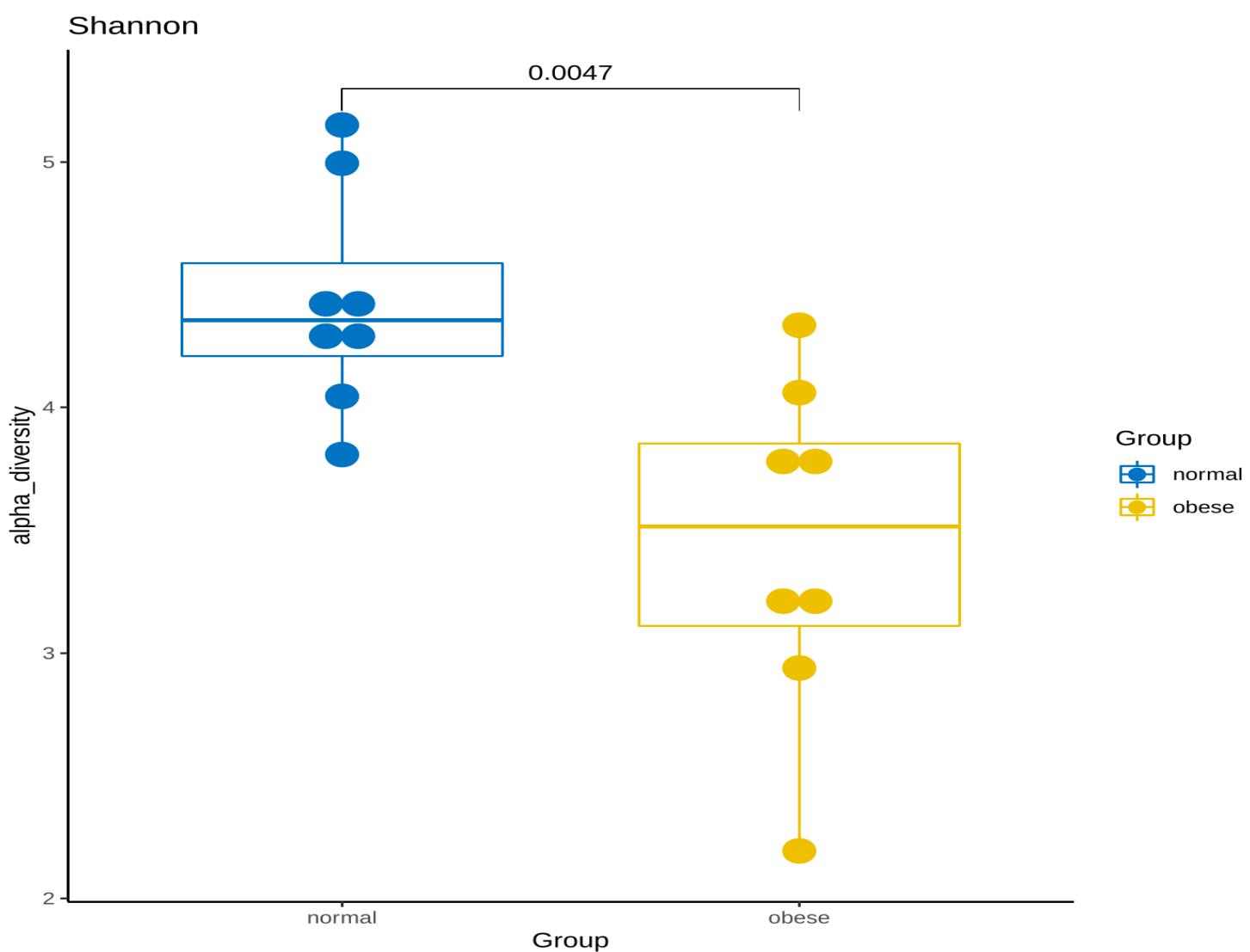
Results

The results are in the directory cat/diversity_analysis. This analytical module helps researchers to understand the complexity of the microbial communities at different taxonomic levels.

Visualization example1

Path: cat/diversity_analysis/alpha_beta_normal_vs_obese/Shannon.alpha_diversity.at_species.pdf

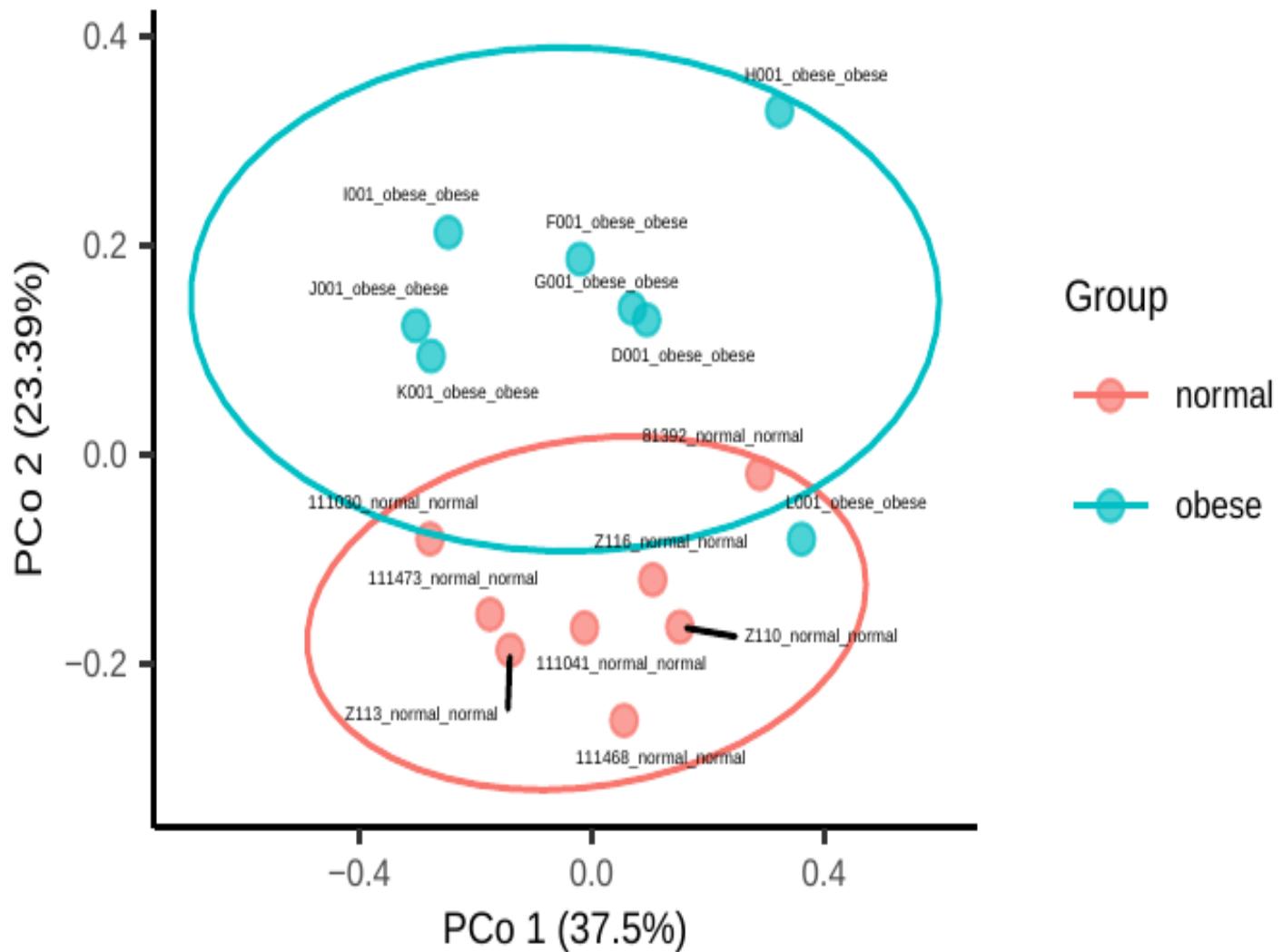
The following box plot shows alpha diversities using the Shannon index at the species levels. The x-axis represents different samples, and the y-axis represents the alpha diversities. Statistical significance was calculated using two-tailed Mann-Whitney U test.



Visualization example2

Path: cat/diversity_analysis/alpha_beta_normal_vs_obese/PCoA12.bray.at_genus.pdf

The PCoA plot presents beta diversities of samples at the genus level



5. Virulence factors, antibiotic genes and plasmids analysis

Methods

For virulence factors analysis, OUTPOST aligns contigs to the Escherichia coli virulence factors (Ecoli_VF) and virulence factor database (VFDB) using ABRicate. The resulting output tables contain information on hit virulence factors and corresponding contigs. OUTPOST further joins the taxonomy annotation and abundance of contigs with the hit virulence factors. OUTPOST normalizes counts before visualizing the distribution of virulence factors among taxa and between groups using cluster heatmaps and histograms separately. For plasmids analysis, OUTPOST aligns contigs to PlasmidFinder. For antibiotic genes analysis, the contigs are aligned to Resfinder, the comprehensive antibiotic resistance database (CARD), ARG-ANNOT, MEGARes, and AMRFinderPlus database. The analytic and visualization procedures of plasmids and antibiotic genes analysis are similar to virulence factors analysis

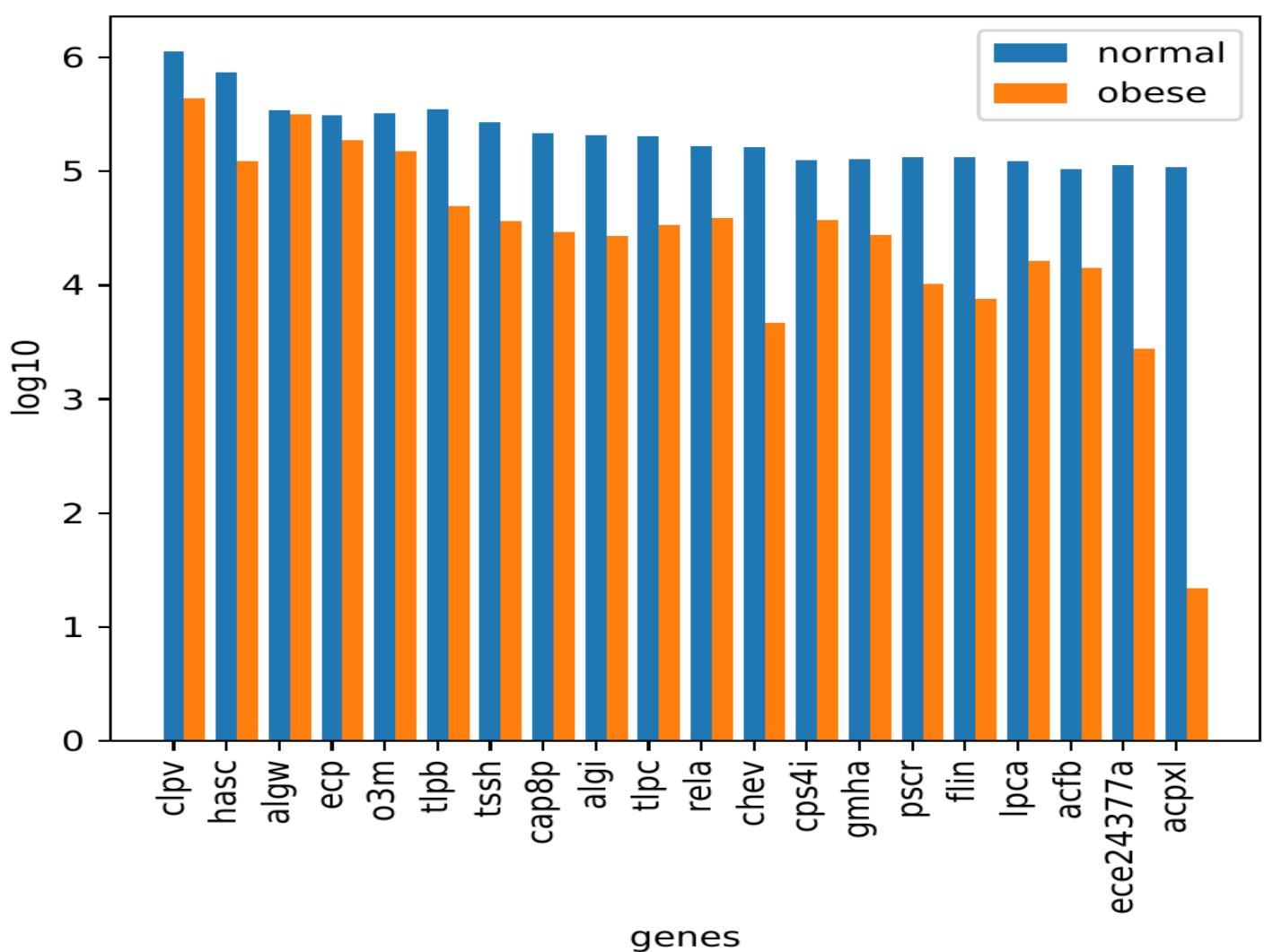
Results

These results are in the directories cat/virulence_factors_analysis, cat/plasmids_analysis, and cat/antibiotic_genes_analysis. These analytical modules help researchers to link taxa with specific genes and profile the distribution of genes along with samples.

Visualization example1

Path: cat/virulence_factors_analysis/genes_normal_vs_obese_distrplot.pdf

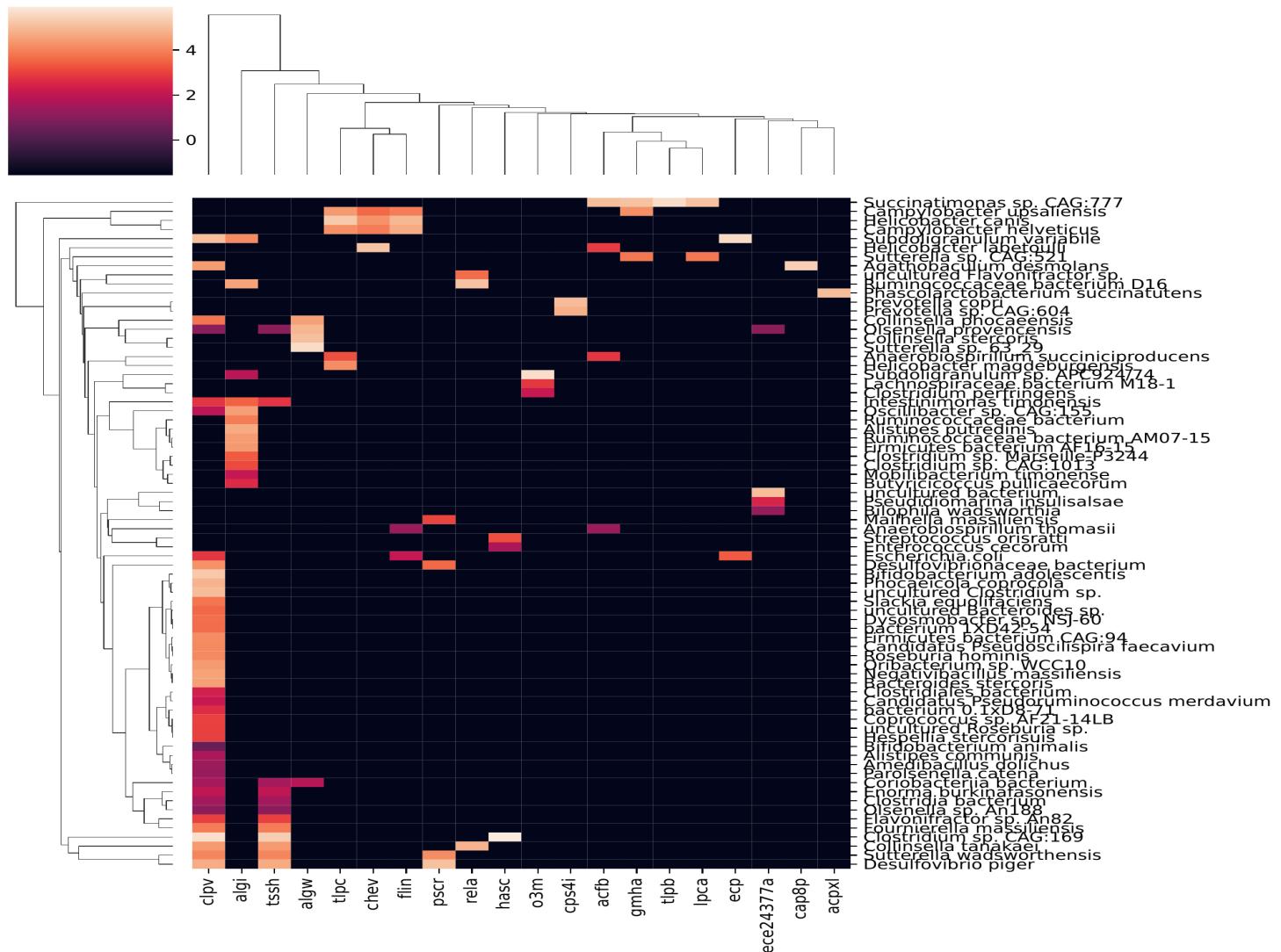
The following distribution plot presents the top genes and their counts of different groups from virulence analysis. The x-axis represents different virulence factors, and the y-axis represents log10(counts). Different color means different group.



Visualization example2

Path: cat/virulence_factors_analysis/genes_species_normal_vs_obese_heatmap.pdf

We also showcase the results of the cluster heatmap of virulence factor genes with species which carried the genes at the species level. The x-axis represents different virulence factors, and the y-axis represents each species. The color means the counts.



6. Function Analysis

Methods

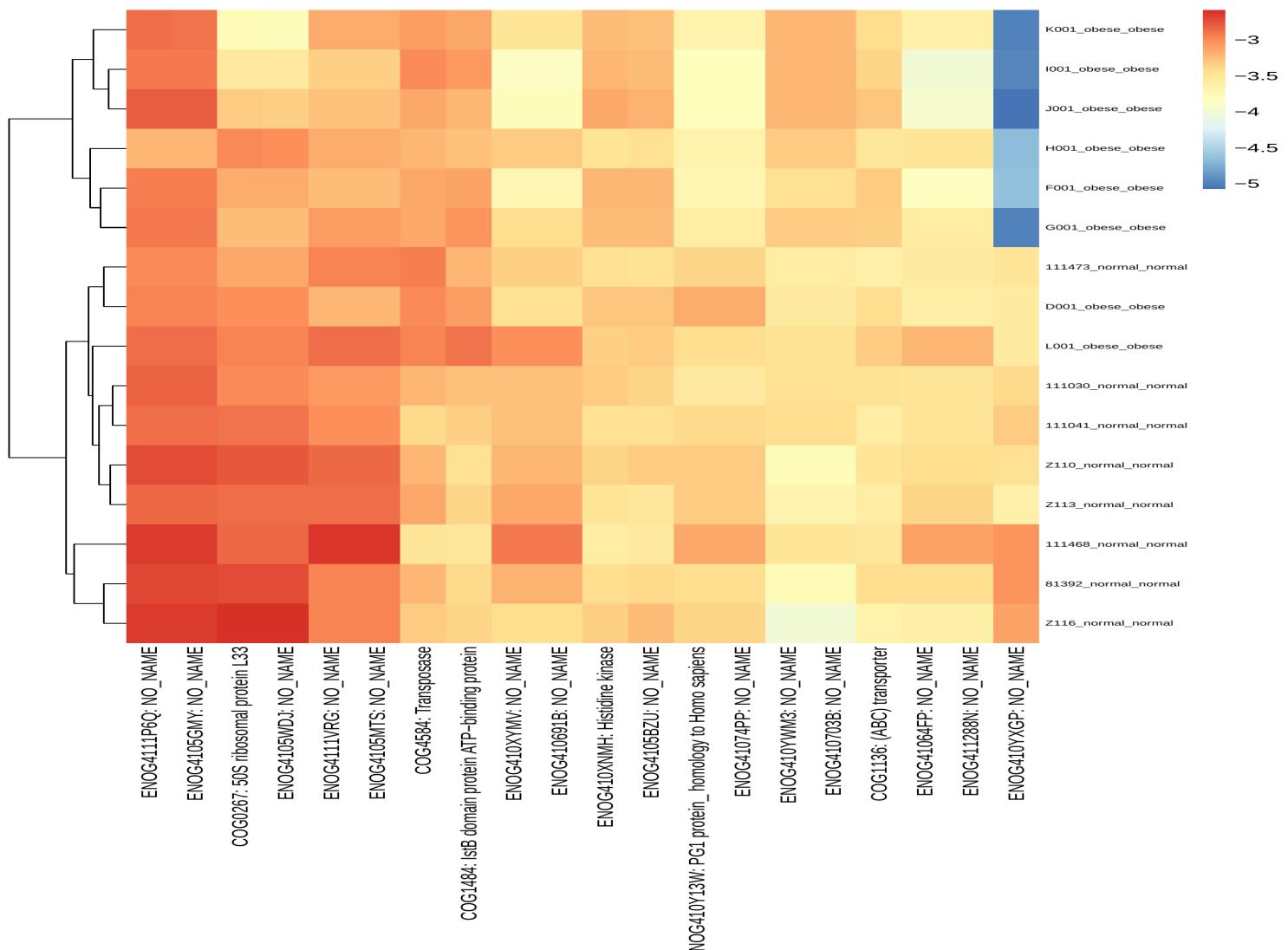
The cleaned reads are passed to HUMAnN 3 to generate function tables for each sample. The function tables are counts of features across 5 independent databases, which are MetaCyc, KEGG Orthology, Pfam, ENZYME, and EggNOG by default. OUTPOST utilizes the interior CPM function in HUMAnN 3 to normalize the feature counts for each sample. To be consistent with taxonomy analysis, the CPM values are further transferred to relative frequency. OUTPOST visualizes the data using cluster heatmaps based on the statistical test results and relative frequency

Results

The results are in the directory cat/function_analysis. This analytic module characterizes all samples, starting from multiple independent functional databases.

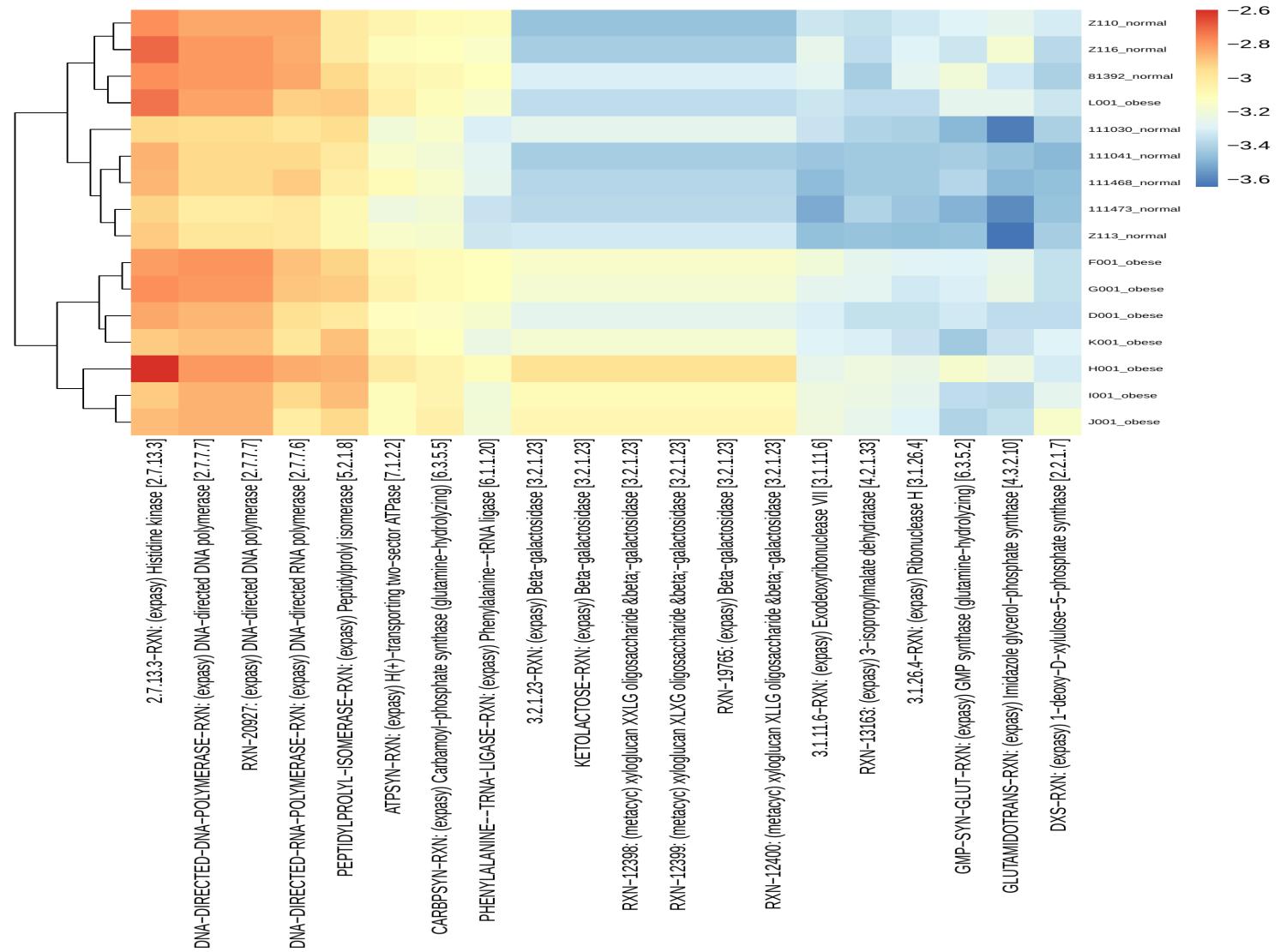
Visualization example1

Path: cat/function_analysis/figs/allSamples_genefamilies_uniref90names_relab_eggnog_unstratified.named.rel_abun_format.normal_vs_obese.rel_abun.unequal.top20.fillmin.scaled.heatmap.pdf



Visualization example2

Path: cat/function_analysis/figs/allSamples_genefamilies_uniref90names_relab_rxn_unstratified.named.rel_abun_format.top20.fillmin.scaled.heatmap.pdf



7. Linear Discriminant Analysis

Methods

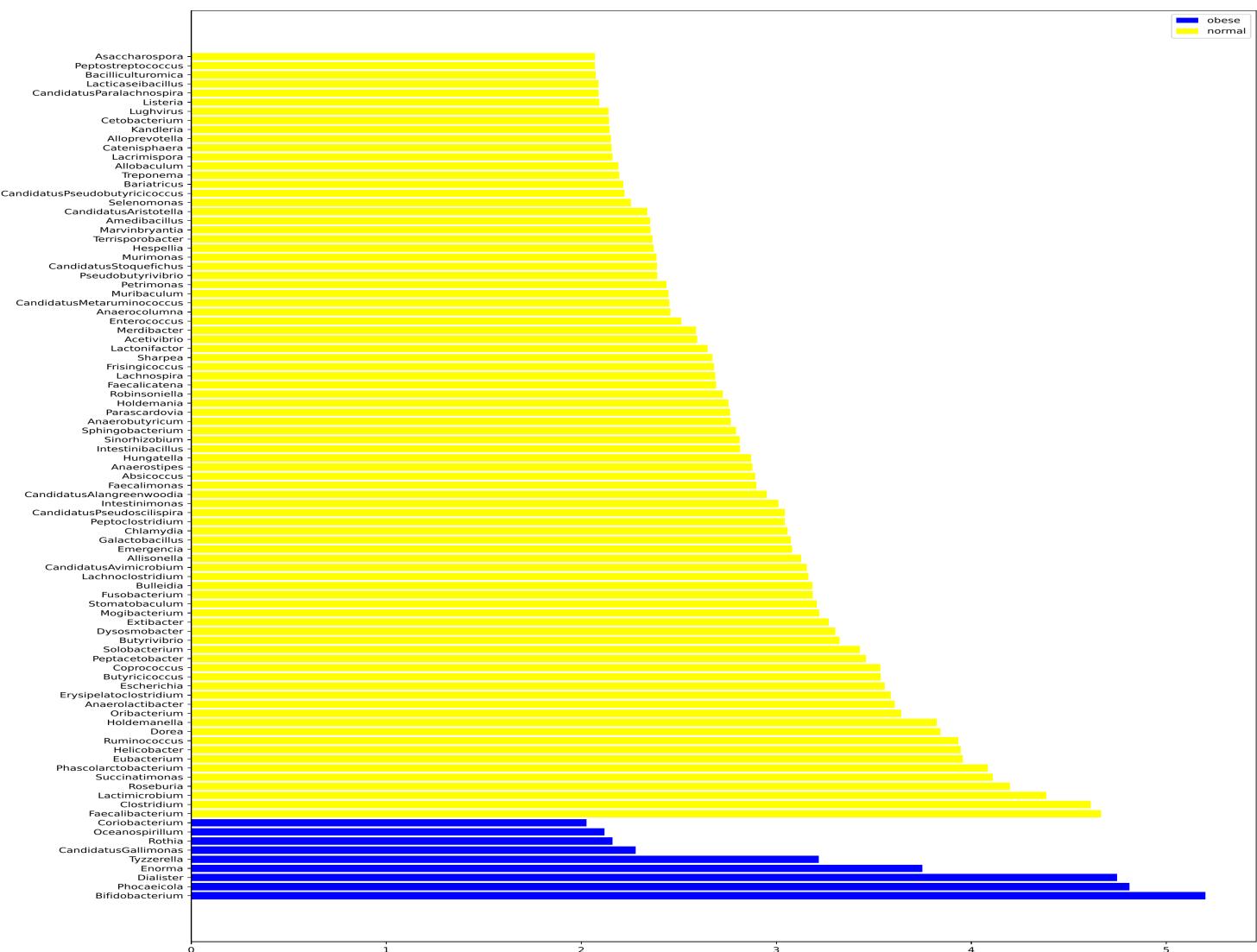
With the relative frequency tables from the taxonomy and function analysis, OUTPOST applies linear discriminant analysis across all taxonomy levels and metabolism databases. It calculates the LDA scores with LEfSe and generates bar plots for features with scores exceeding the threshold (default by 2).

Results

The results are in the directory cat/LDA_analysis. This analytic module help researchers to identify characteristic taxa or functional items between any two groups.

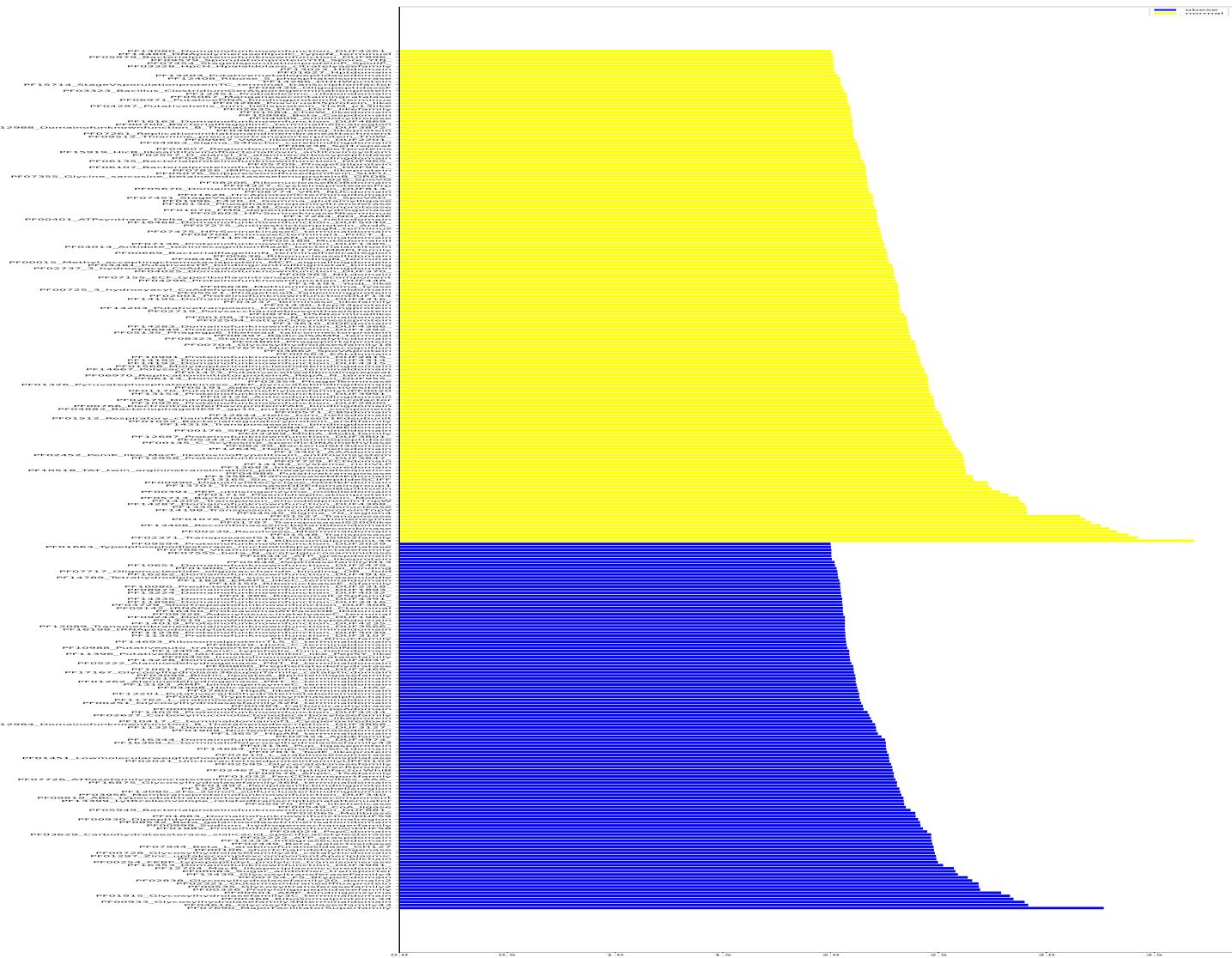
Visualization example1

Path: cat/LDA_analysis/figs_taxa/cat.rel_abun.normal_vs_obese.at_genus.rel_abun.unequal.lefse.pdf



Visualization example2

Path: cat/LDA_analysis/figs_humann/allSamples_genefamilies_uniref90names_relab_pfam_unstratified.named.rel_abun_format.normal_vs_obese.rel_abun.unequal.lefse.pdf



8. Biomarkers Analysis

Methods

Drawing from the results of the previous eight analysis modules, OUTPOST sets seven evaluation criteria. These include the ranking of relative abundance, the comparison of inter-group relative abundance, the verification of a sufficiently high LDA score, the ranking of virulence factors' carriage, the ranking of plasmids' carriage, the ranking of antibiotic genes' carriage, and the application of ANCOM (Analysis of Composition of Microbiomes).

Results

The results are in the directory cat/biomarkers_analysis. This analytic module helps researchers explore the biomarkers using the results from the previous eight analysis modules.

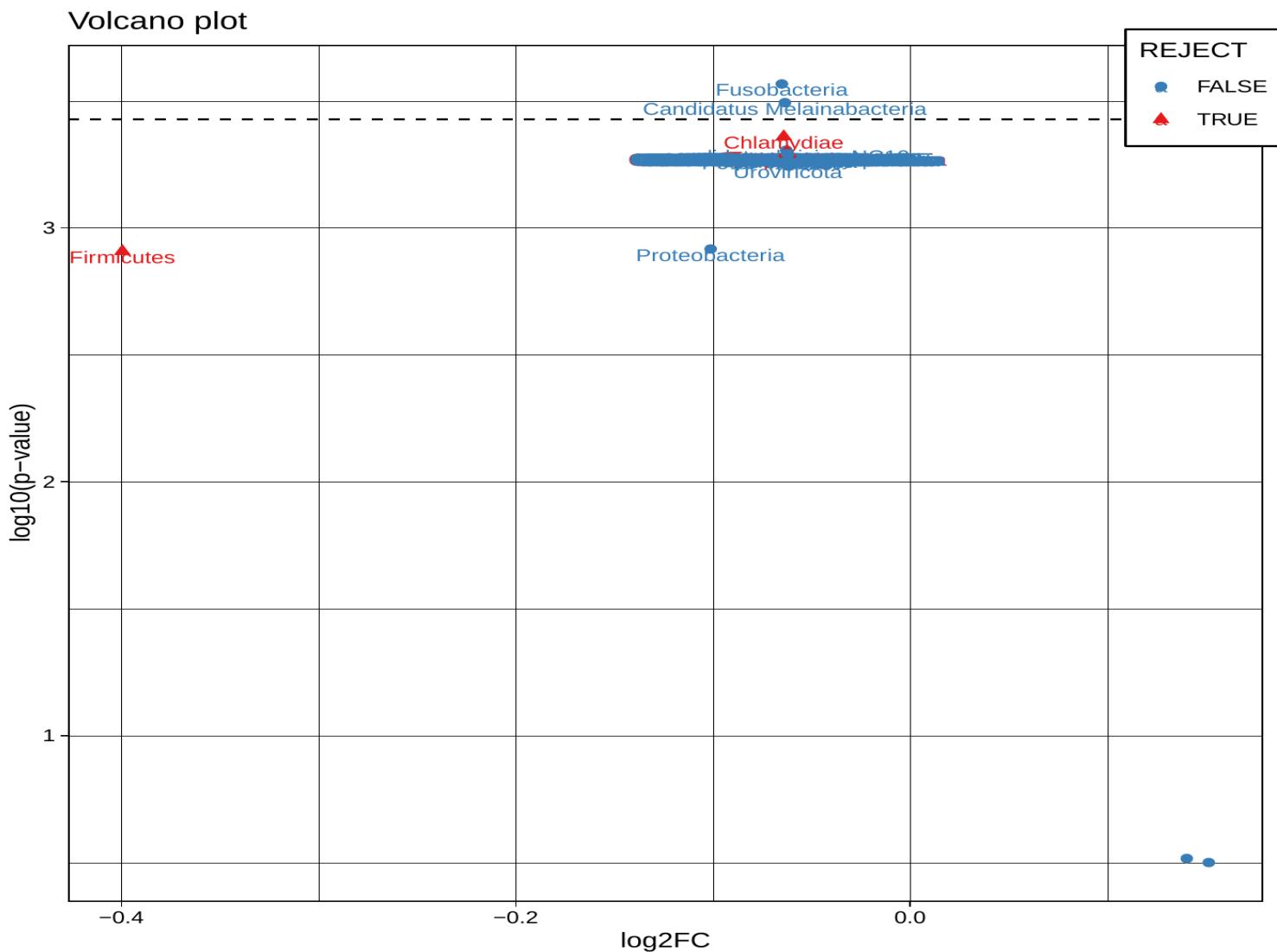
Visualization example1

Path: cat/biomarkers_analysis/OUTPOST_biomarker_scores.normal_vs_obese.species.tsv

taxa_abun_top_100	taxa_abun_significant	LDA_larger_than_2	virulence_top_100	plasmid_top_100	antibiotic_top_100	ANCOM_idenitified
Escherichia coli	1.0	1.0	1.0	1.0	1.0	1.0
Holdemanella biformis	1.0	1.0	1.0	1.0	0.0	1.0
Succinatimona s sp. CAG:777	1.0	1.0	1.0	1.0	0.0	1.0

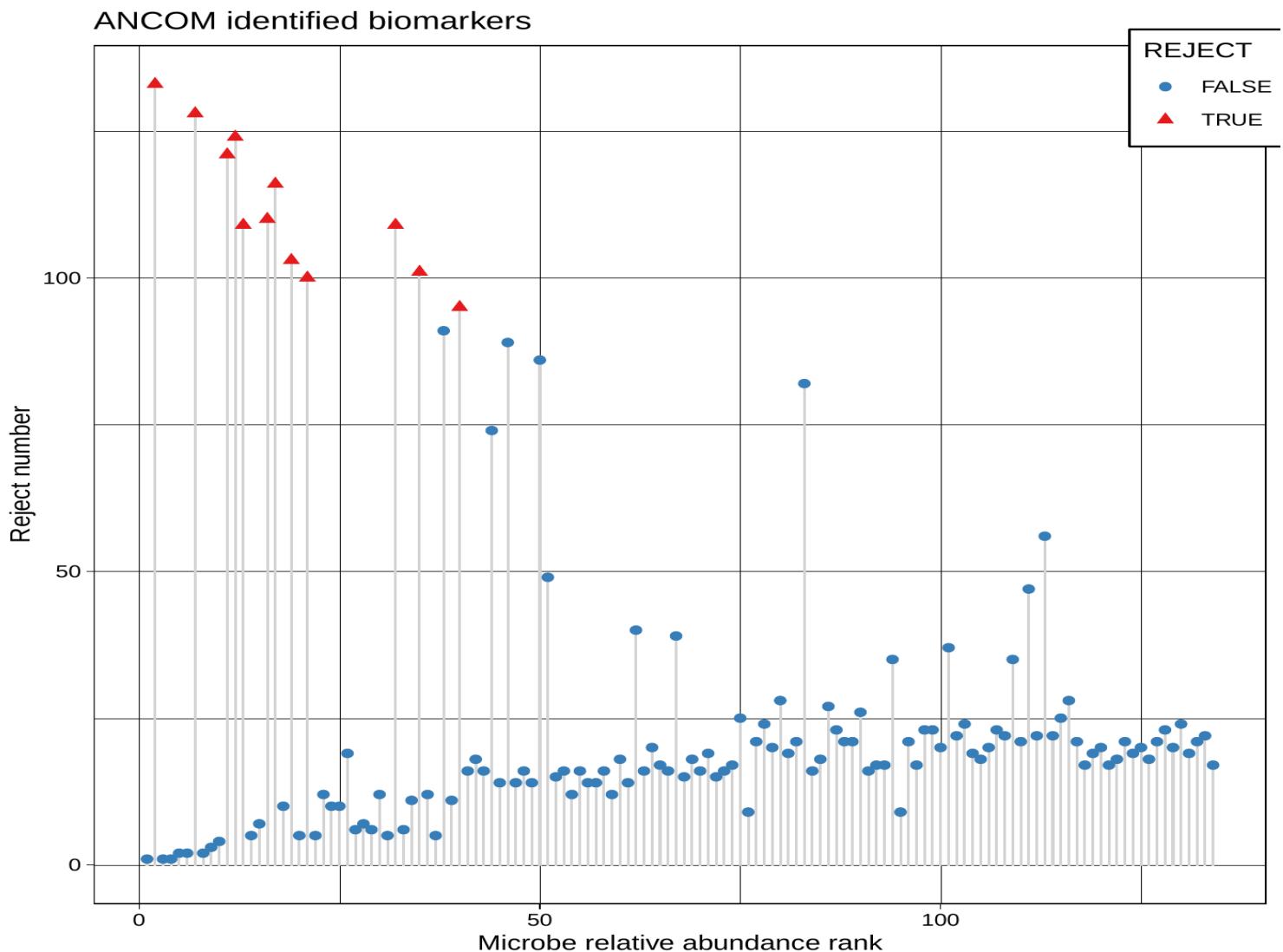
Visualization example2

Path: cat/biomarkers_analysis/ANCOM_identification/ancom_biomarkers.volcano.normal_vs_obese.at_phylum.pdf



Visualization example3

Path: cat/biomarkers_analysis/ANCOM_identification/ancom_biomarkers.dotplot.normal_vs_obese.at_phylum.pdf



9. Notes

The log files and benchmark files for each process step are in directories /log and /benchmark.

For multi-batch data, OUTPOST can erase the batch effect and store the results in the directory /batch effect.

For multi-group projects, OUTPOST generates results for each pairwise comparison between groups. This means that OUTPOST usually produces tens of thousands of plots and tables, making it an incredibly powerful tool for microbiome analysis.