Package ‘mbX’

May 8, 2026

Type Package

Title A Comprehensive Microbiome Data Processing Pipeline

Version 0.2.0

Depends R (>= 4.1.0)

Description Provides tools for cleaning, processing, and preparing microbiome sequencing data (e.g., 16S rRNA) for downstream analysis. Supports CSV, TXT, and Excel file formats. The main function, ezclean(), automates microbiome data transformation, including format validation, transposition, numeric conversion, and metadata integration. It also handles taxonomic levels efficiently, resolves duplicated taxa entries, and outputs a well-structured, analysis-ready dataset. The companion functions ezstat() run statistical tests and summarize results, while ezviz() produces publication-ready visualizations.

License MIT + file LICENSE

Encoding UTF-8

Imports tools, readxl, openxlsx, dplyr, tidyr, ggplot2, rstatix, tibble, FSA, multcompView

RoxygenNote 7.3.2

VignetteBuilder knitr

Suggests knitr, rmarkdown

NeedsCompilation no

Maintainer Utsav Lamichhane <`utsav.lamichhane@gmail.com`>

Author Utsav Lamichhane [aut, cre]

Repository CRAN

Date/Publication 2025-07-23 23:00:02 UTC

Contents

ezclean 2
ezstat 3
ezviz 4

Index **6**

`ezclean` *Clean and Process Microbiome Data*

Description

Processes microbiome and metadata files (e.g., 16S rRNA sequencing data) to produce an analysis-ready dataset. Supports CSV, TXT, and 'Excel' file formats. This function validates file formats, reads the data, and merges the datasets by the common column 'SampleID'. If a 'Taxonomy' column exists, the data are filtered to include only rows matching the provided taxonomic level.

Usage

```
ezclean(microbiome_data, metadata, level = "d")
```

Arguments

`microbiome_data`	A string specifying the path to the microbiome data file.
`metadata`	A string specifying the path to the metadata file.
`level`	A string indicating the taxonomic level for filtering the data (e.g., "genus").

Value

A data frame containing the cleaned and merged dataset.

Examples

```
## Not run:
  mb  <- system.file("extdata", "microbiome.csv", package = "mbX")
  md  <- system.file("extdata", "metadata.csv",   package = "mbX")
  if (nzchar(mb) && nzchar(md)) {
    cleaned_data <- ezclean(mb, md, "g")
    head(cleaned_data)
  } else {
    message("Sample data files not found.")
  }

## End(Not run)
```

`ezstat` *Statistical Analysis and Visualization of Microbiome Data*

Description

Performs Kruskal_Wallis tests, post_hoc Dunn comparisons, Compact Letter Display (CLD) summaries, and generates boxplots annotated with CLD letters for taxa abundances grouped by a chosen metadata variable.

Usage

```
ezstat(microbiome_data, metadata, level, selected_metadata)
```

Arguments

`microbiome_data`	Character; path to the microbiome abundance table (CSV, TSV, XLS, or XLSX).
`metadata`	Character; path to the sample metadata file (CSV, TXT, XLS, or XLSX).
`level`	Character; taxonomic rank to aggregate at (e.g. "genus", "g").
`selected_metadata`	Character; name of the categorical metadata column to group by.

Details

This function first calls ezclean to produce a cleaned, merged table of sample IDs, metadata, and taxa abundances at the requested taxonomic level. It then:

1. Runs Kruskal_Wallis tests on each taxon and writes results with FDR_correction.
2. Performs Dunns pairwise post_hoc tests (BH_adjusted) for taxa with KW p less than or equal to 0.05.
3. Computes CLD letters for significantly different groups and writes a summary Excel.
4. Generates high-resolution (900 dpi) boxplots annotated with CLD letters.

Value

Invisibly returns the `data.frame` of cleaned sample_taxa abundances used for all analyses.

Examples

```
## Not run:
  mb  <- system.file("extdata", "microbiome.csv", package = "mbX")
  md  <- system.file("extdata", "metadata.csv",   package = "mbX")
  if (nzchar(mb) && nzchar(md)) {
    ezstat(mb, md, "genus", "Group")
  }

## End(Not run)
```

`ezviz` *Visualize Microbiome Data*

Description

Generates publication-ready visualizations for microbiome data. This function first processes the microbiome and metadata files using ezclean(), then creates a bar plot using ggplot2. Supported file formats are CSV, TXT, and 'Excel'. Note: Only one of the parameters top_taxa or threshold should be provided.

Usage

```
ezviz(
  microbiome_data,
  metadata,
  level,
  selected_metadata,
  top_taxa = NULL,
  threshold = NULL,
  flip = FALSE
)
```

Arguments

`microbiome_data`	A string specifying the path to the microbiome data file.
`metadata`	A string specifying the path to the metadata file.
`level`	A string indicating the taxonomic level for filtering the data (e.g., "genus").
`selected_metadata`	A string specifying the metadata column used for grouping.
`top_taxa`	An optional numeric value indicating the number of top taxa to keep. Use this OR threshold, but not both.
`threshold`	An optional numeric value indicating the minimum threshold value; taxa below this threshold will be grouped into an "Other" category.
`flip`	Logical. If 'TRUE', the order of the stacks is reversed.

Value

A ggplot object containing the visualization.

Examples

```
mb  <- system.file("extdata", "microbiome.csv", package = "mbX")
md  <- system.file("extdata", "metadata.csv",   package = "mbX")
plot_obj <- ezviz(
  microbiome_data = mb,
```

```
    metadata        = md,
    level           = "genus",
    selected_metadata    = "sample_type",
    top_taxa        = 20,
    flip            = FALSE
  )
  print(plot_obj)
```

Index

`ezclean`, 2
`ezstat`, 3
`ezviz`, 4