

Package ‘lefser’

September 26, 2023

Type Package

Title R implementation of the LEfSE method for microbiome biomarker discovery

Description lefser is an implementation in R of the popular “LDA Effect Size (LEfSe)” method for microbiome biomarker discovery. It uses the Kruskal-Wallis test, Wilcoxon-Rank Sum test, and Linear Discriminant Analysis to find biomarkers of groups and sub-groups.

Version 1.10.1

License Artistic-2.0

LazyData true

Depends SummarizedExperiment, R (>= 4.0.0)

Imports coin, MASS, ggplot2, stats, S4Vectors, methods

Suggests knitr, rmarkdown, curatedMetagenomicData, BiocStyle, testthat, pkgdown, covr, withr

Encoding UTF-8

BugReports <https://github.com/waldronlab/lefser/issues>

URL <https://github.com/waldronlab/lefser>

VignetteBuilder knitr

biocViews Software, Sequencing, DifferentialExpression, Microbiome, StatisticalMethod, Classification

RoxygenNote 7.2.3

Roxygen list(markdown = TRUE)

git_url <https://git.bioconductor.org/packages/lefser>

git_branch RELEASE_3_17

git_last_commit 24fb5ff

git_last_commit_date 2023-08-17

Date/Publication 2023-09-26

Author Asya Khleborodova [cre, aut],
 Ludwig Geistlinger [ctb],
 Marcel Ramos [ctb] (<<https://orcid.org/0000-0002-3242-0582>>),
 Levi Waldron [ctb]

Maintainer Asya Khleborodova <asya.bioconductor@gmail.com>

R topics documented:

lefser	2
lefserPlot	4
relativeAb	4
zeller14	5

Index	6
--------------	----------

lefser	<i>R implementation of the LEfSe method</i>
--------	---

Description

Perform a LEfSe analysis: the function carries out differential analysis between two sample groups for multiple microorganisms and uses linear discriminant analysis to establish their effect sizes. Subclass information for each class can be incorporated into the analysis (see examples). Microorganisms with large differences between two sample groups are identified as biomarkers.

Usage

```
lefser(
  expr,
  kruskal.threshold = 0.05,
  wilcox.threshold = 0.05,
  lda.threshold = 2,
  groupCol = "GROUP",
  blockCol = NULL,
  assay = 1L,
  trim.names = FALSE,
  checkAbundances = TRUE
)
```

Arguments

expr	A SummarizedExperiment with relative abundances in the assay
kruskal.threshold	numeric(1) The p-value for the Kruskal-Wallis Rank Sum Test (default 0.05).
wilcox.threshold	numeric(1) The p-value for the Wilcoxon Rank-Sum Test when 'blockCol' is present (default 0.05).

lda.threshold	numeric(1) The effect size threshold (default 2.0).
groupCol	character(1) Column name in 'colData(expr)' indicating groups, usually a factor with two levels (e.g., 'c("cases", "controls"); default "GROUP").
blockCol	character(1) Optional column name in 'colData(expr)' indicating the blocks, usually a factor with two levels (e.g., 'c("adult", "senior"); default NULL).
assay	The i-th assay matrix in the 'SummarizedExperiment' ('expr'; default 1).
trim.names	If 'TRUE' extracts the most specific taxonomic rank of organism.
checkAbundances	logical(1) Whether to check if the assay data in the expr input are relative abundances or counts. If counts are found, a warning will be emitted (default TRUE).

Details

The LEfSe method expects relative abundances in the expr input. A warning will be emitted if the column sums do not result in 1. Use the relativeAb helper function to convert the data in the SummarizedExperiment to relative abundances. The checkAbundances argument enables checking the data for presence of relative abundances and can be turned off by setting the argument to FALSE.

Value

The function returns a dataframe with two columns, which are names of microorganisms and their LDA scores.

Examples

```
# (1) Using classes only
data(zeller14)
# exclude 'adenoma'
zeller14 <- zeller14[, zeller14$study_condition != "adenoma"]
res_group <- lefser(zeller14, groupCol = "study_condition")
head(res_group)

# (2) Using classes and subclasses
data(zeller14)
# exclude 'adenoma'
zeller14 <- zeller14[, zeller14$study_condition != "adenoma"]
res_block <- lefser(
  zeller14, groupCol = "study_condition", blockCol = "age_category"
)
head(res_block)
```

lefserPlot	<i>Plots results from 'lefser' function</i>
------------	---

Description

'lefserPlot' function displays effect sizes for differentially expressed microorganisms and whether they are more abundant in '0' or '1' sample group.

Usage

```
lefserPlot(df, colors = c("red", "forestgreen"), trim.names = TRUE)
```

Arguments

df	Data frame produced by 'lefser'.
colors	character(2) The two colors corresponding to class 0 and 1, respectively. Defaults to 'c("red", "forestgreen)".
trim.names	If 'TRUE' extracts the most specific taxonomic rank of organism.

Value

Function returns plot of effect size scores produced by 'lefser'. Positive scores represent microorganisms with that are more abundant in class '1'. Negative scores represent microorganisms with that are more abundant in class '0'.

Examples

```
example("lefser")
lefserPlot(res_group)
```

relativeAb	<i>Utility function to calculate relative abundances</i>
------------	--

Description

The function calculates the column totals and divides each value within the column by the respective column total.

Usage

```
relativeAb(se, assay = 1L)
```

Arguments

se	A SummarizedExperiment object with counts
assay	The i-th assay matrix in the 'SummarizedExperiment' ('expr'; default 1).

Examples

```
se <- SummarizedExperiment(  
  assays = list(  
    counts = matrix(  
      rep(1, 4), ncol = 1, dimnames = list(LETTERS[1:4], "SAMP")  
    )  
  )  
)  
assay(se)  
assay(relativeAb(se))
```

zeller14

Example dataset for lefser

Description

The ZellerG_2014 dataset contains microbiome count data for CRC patients and controls. It was for curatedMetagenomicData using the script in the package directory "data-raw".

Usage

```
zeller14
```

Format

A SummarizedExperiment with 1585 features, 199 samples

study_condition adenoma, control, CRC

age_category adult, senior

Source

<https://pubmed.ncbi.nlm.nih.gov/25432777/>

Index

* **datasets**

zeller14, [5](#)

lefser, [2](#)

lefserPlot, [4](#)

relativeAb, [4](#)

SummarizedExperiment, [2](#)

zeller14, [5](#)