Package 'ADAPT'

November 13, 2024

Title Analysis of Microbiome Differential Abundance by Pooling Tobit Models

Version 1.0.0

Description ADAPT carries out differential abundance analysis for microbiome metagenomics data in phyloseq format.
It has two innovations. One is to treat zero counts as left censored and use Tobit models for log count ratios.
The other is an innovative way to find non-differentially abundant taxa as reference, then use the reference taxa to find the differentially abundant ones.

License MIT + file LICENSE

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.2

Imports Rcpp (>= 1.0.8), RcppArmadillo (>= 0.10.8), RcppParallel (>= 5.1.5), phyloseq (>= 1.39.0), methods, stats, ggplot2 (>= 3.4.1), ggrepel (>= 0.9.1)

Suggests rmarkdown (>= 2.11), knitr (>= 1.37), testthat (>= 3.0.0)

Config/testthat/edition 3

LinkingTo Rcpp, RcppArmadillo, RcppParallel

biocViews DifferentialExpression, Microbiome, Normalization, Sequencing, Metagenomics, Software, MultipleComparison

Depends R (>= 4.1.0)

LazyData false

VignetteBuilder knitr

git_url https://git.bioconductor.org/packages/ADAPT

git_branch RELEASE_3_20

git_last_commit d0f8248

git_last_commit_date 2024-10-29

Repository Bioconductor 3.20

Date/Publication 2024-11-13

Author Mukai Wang [aut, cre] (<https://orcid.org/0000-0002-1413-1904>),
Simon Fontaine [ctb],
Hui Jiang [ctb],
Gen Li [aut, ctb]

Maintainer Mukai Wang <wangmk@umich.edu>

Contents

`adapt` *ADAPT*

Description

Analysis of microbiome differential abundance by pooling tobit models

Usage

```
adapt(
  input_data,
  cond.var,
  base.cond = NULL,
  adj.var = NULL,
  censor = 1,
  prev.filter = 0.05,
  depth.filter = 1000,
  alpha = 0.05
)
```

Arguments

`input_data`	a phyloseq object
`cond.var`	the variable representing the conditions to compare, a character string
`base.cond`	the condition chosen as baseline. This is only used when the condition is categorical.
`adj.var`	the names of the variables to be adjusted, a vector of character strings
`censor`	the value to censor at for zero counts, default 1
`prev.filter`	taxa whose prevalences are smaller than the cutoff will be excluded from analysis, default 0.05
`depth.filter`	a sample would be discarded if its library size is smaller than the threshold
`alpha`	the cutoff of the adjusted p values

Details

ADAPT takes in a metagenomics count table as a phyloseq object. The phyloseq object needs to have metadata containing at least one variable `cond.var` representing the conditions that the user is testing on. The condition variable `cond.var` can be numeric (as a continuous variable) or character (representing categorical variable). ADAPT does not support multigroup comparison yet. If there are multiple conditions, the user can specify the condition to single out through `base.cond`. ADAPT then carry out DAA between the selected `base.cond` and all the others. ADAPT allows adjusting for other covariates. The user can specify all the covariates to adjust for by specifying `adj.var` with a vector of variable names.

Differential abundance analysis may be too challenging for rare taxa and samples with too low sequencing depth. The users can filter out taxa whose prevalences are lower than `prev.filter` (default 0.05). The users can also filter out samples whose sequencing depths (library sizes) are smaller than `depth.filter` (default 1000).

One major feature of ADAPT is treating zero counts as left censored observations and use Tobit models for log count ratios. The zero counts by default are left censored at one. The users can change the value to censor at through `censor`. Change the cutoff of BH-adjusted p-values with `alpha` (default 0.05) for calling DA taxa.

The returned value of `adapt` is a customized S4 type called `DAresult`. We have developed two helper functions `summary` and `plot` for this special data type.

Value

a `DAresult` type object contains the input and the output. Use summary and plot to explore the output

Examples

```
data(ecc_plaque)
plaque_results <- adapt(input_data=ecc_plaque, cond.var="CaseStatus",
       base.cond="case")
data(ecc_saliva)
saliva_results <- adapt(input_data=ecc_saliva, cond.var="CaseStatus",
       base.cond="Control", adj.var="Site")
```

DAresult-class *Differential abundance analysis result*

Description

An S4 class to represent ADAPT analysis results

Details

The analysis result object contains the analysis name, reference taxa, DA taxa, detailed analysis results as a dataframe and the input phyloseq object. The analysis name contains the condition variable. The reference taxa `reference` must be nonempty. DA taxa `signal` may be an empty string if no taxa are differentially abundant. The `details` dataframe contains the taxa names, the prevalence of taxa, the estimated log10 absolute abundance fold changes, the raw hypothesis test p-values and BH-adjusted p-values.

Slots

`DAAname` The name of differential abundance analysis

`reference` A vector of taxa names corresponding to all the reference taxa

`signal` A vector of taxa names corresponding to all the DA taxa

`details` A dataframe with the analysis results for all taxa

`input` Input phyloseq object

---

ecc_plaque *Plaque samples from early childhood dental caries studies*

Description

A phyloseq object with 30 samples and 610 taxa for whole genome sequencing of plaque samples. The samples were collected from children at 36, 48 or 60 months old. 15 samples were from teeth with dental lesions and 15 samples were controls. The samples of cases were collected at the onset visit.

Usage

```
data(ecc_plaque)
```

Format

The metadata of `ecc_plaque` has two columns

CaseStatus Whether the kids had dental caries

Site The location of sample collection (Site 1 or Site 2).

Source

The original publication of this data is "Evaluating the ecological hypothesis: early life salivary microbiome assembly predicts dental caries in a longitudinal case-control study" (Blostein etal, 2022). The sequence data are available under Project number PRJNA752888.

---

ecc_saliva *Saliva samples from early childhood dental caries studies*

Description

A phyloseq object with 161 samples and 280 ASVs for 16S sequencing of saliva samples. The samples were collected from 12-month-old infants. 84 out of 161 children developed dental caries after 36 months old. All samples have been de-identified.

Usage

```
data(ecc_saliva)
```

Format

The metadata of ecc_saliva has two columns

CaseStatus Whether the child developed dental caries after 36 months old

Site The location of sample collection (Site 1 or Site 2).

Source

The original publication of this data is "Evaluating the ecological hypothesis: early life salivary microbiome assembly predicts dental caries in a longitudinal case-control study" (Blostein etal, 2022). The sequence data are available under Project number PRJNA752888.

---

plot,DAresult,ANY-method

Plotting differential abundance analysis results

Description

Volcano plot of ADAPT results

Usage

```
## S4 method for signature 'DAresult,ANY'
plot(x, n.label = 5)
```

Arguments

x	analysis result in DAresult type
n.label	Number of taxa to label on the plot. Note that no taxa will be labeled if no DA taxa.

Details

The customized plot function for DAresult type object generates a volcano plot with the differentially abundant taxa highlighted. The users can decide how many taxa with the smallest p-values are labeled on the plot.

Value

A ggplot object of the volcano plot

Examples

```
data(ecc_saliva)
saliva_results <- adapt(input_data=ecc_saliva, cond.var="CaseStatus",
       base.cond="Control", adj.var="Site")
plot(saliva_results, n.label=10)
```

summary,DAresult-method

Summary of differential abundance analysis

Description

Summary function for DAresult type object

Usage

```
## S4 method for signature 'DAresult'
summary(object, select = c("all", "da", "ref"))
```

Arguments

object	analysis result in DAresult type
select	Taxa whose results to be returned, can be all the taxa ("all"), only the differentially abundant taxa ("da") or reference taxa ("ref").

Details

This customized summary function reports the dimension of input count table, number of reference taxa and number of differentially abundant taxa. It also returns a data frame with the detailed analysis result and taxonomy of all the taxa. The user can choose to only get the detailed analysis result of DA taxa or the reference taxa through the select parameter.

Value

A dataframe with detailed analysis results

Examples

```
data(ecc_saliva)
saliva_results <- adapt(input_data=ecc_saliva, cond.var="CaseStatus",
       base.cond="Control", adj.var="Site")
summary(saliva_results, select="da")
```

Index