

Package ‘POSTm’

April 30, 2021

Type Package

Title Phylogeny-Guided OTU-Specific Association Test for Microbiome Data

Version 1.1

Date 2021-04-28

Author Caizhi Huang [aut],
Jung-Ying Tzeng [aut],
Shannon Holloway [aut, cre]

Maintainer Shannon Holloway <sthollow@ncsu.edu>

Description Implements the Phylogeny-Guided Microbiome OTU-Specific Association Test method, which boosts the testing power by adaptively borrowing information from phylogenetically close OTUs (operational taxonomic units) of the target OTU. This method is built on a kernel machine regression framework and allows for flexible modeling of complex microbiome effects, adjustments for covariates, and can accommodate both continuous and binary outcomes.

Depends methods

Imports CompQuadForm, ape

Suggests multtest, rmarkdown, knitr

biocViews

License GPL-2

Encoding UTF-8

LazyData true

VignetteBuilder knitr

RoxygenNote 7.1.1

Collate 'ACAT.R' 'CKAT.R' 'POSTmData.R' 'distanceMatrix.R'
'getKernel.R' 'glmFit.R' 'lmFit.R' 'p.adjust.POST.R'
'plot.POST.R' 'post.R' 'print.POST.R'

NeedsCompilation no

Repository CRAN

Date/Publication 2021-04-30 06:50:18 UTC

R topics documented:

p.adjust	2
plot	3
post	4
POSTmData	6
print	7

Index	9
--------------	----------

p.adjust	<i>Adjust P-values for Multiple Comparisons</i>
----------	---

Description

Given a set of p-values, returns p-values adjusted using one of several methods.

Given an object of class POST, returns the POST and SO p-values adjusted using the specified method(s).

Usage

```
p.adjust(p, ...)
```

```
## S3 method for class 'POST'
```

```
p.adjust(p, ..., method = p.adjust.methods, n = length(x = p), alpha = 0.05)
```

Arguments

p	An object of class POST. The value object returned by post().
...	Ignored.
method	A character vector object. One or more correction methods. must be one or more from "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none", "TSBH".
n	number of comparisons, must be at least 'length(p)'; Per ?stats::p.adjust – only set this (to non-default) when you know what you are doing!
alpha	A numeric object. A nominal type I error rate, or a vector of error rates, used for estimating the number of true null hypotheses in the two-stage Benjamini & Hochberg procedure ("TSBH"). Default is 0.05. NOTE: This definition does not conform to stats::p.adjust.

Details

Extends the p.adjust() method of stats and incorporates the TSBH method available through multtest::mt.raw2adjp.

Value

A list, the contents of which will depend on the type of analysis. If a tree object was provided for the analysis, the list will contain

`adjPOST` A matrix object. The adjusted p-values for the POST p-values. The rows correspond to each OTU; the columns to each specified adjustment method.

`adjSO` A matrix object. The adjusted p-values for the single OTU p-values. The rows correspond to each OTU; the columns to each specified adjustment method.

If no tree object was provided, only `adjSO` as described above will be returned.

Examples

```
data("POSTmData")

y <- as.integer(x = metadata[, "GC"] == "BV")
X <- metadata[, "mRace"]

result <- post(y = y,
              X = X,
              OTU = otu[, 1:20],
              tree = otutree,
              cValues = seq(0, 0.05, by = 0.01))

p.adjust(p = result, method = c("BH", "BY"))
```

plot

Plot Phylogenies with Significant OTU

Description

Plot phylogenetic tree on the current graphical device as defined by package `ape` with tip labels highlighted to show the significant OTU.

Usage

```
## S3 method for class 'POST'
plot(x, ..., siglevel = 0.05, method = "none", alpha = 0.05, subTree = TRUE)
```

Arguments

`x` A POST object. The value object returned by `post()`. Note that a tree of class "phylo" must have been provided to the original analysis.

`...` Additional arguments to be passed to `plot.phylo`.

`siglevel` A numeric object. The significance level; OTU's with (adjusted or raw) p-values below this level will be highlighted. Default is 0.05.

method	A character object. Must be one of {"holm", "hochberg", "hommel", "bonferoni", "BH", "BY", "fdr", "TSBH", "none"}, where 'none' indicates the raw POST p-values. "TSBH" is the two-stage Benjamini & Hochberg procedure. See ?stats::p.adjust for descriptions of all other possible values.
alpha	A numeric object. A nominal type I error rate, or a vector of error rates, used for estimating the number of true null hypotheses in the two-stage Benjamini & Hochberg procedure ("TSBH"). Default is 0.05. Ignored if method is not TSBH.
subTree	A logical object. If TRUE, only the OTU used in the analysis are included in the plot.

Value

No return value, called to produce graphical elements.

Examples

```
data("POSTmData")

y <- as.integer(x = metadata[, "GC"] == "BV")
X <- metadata[, "mRace"]

result <- post(y = y,
              X = X,
              OTU = otu[, 1:20],
              tree = otutree,
              cValues = seq(0, 0.05, by = 0.01))

plot(x = result)
```

post *Phylogeny-Guided OTU-Specific Association Test for Microbiome Data*

Description

The POST implements a phylogeny-guided OTU-specific association test for microbiome data. This method boosts the testing power by adaptively borrowing information from phylogenetically close OTUs of the target OTU. Whether or not borrowing information or the amount of information from the neighboring OTUs is data adaptive and supervised by phylogenetic distance and the outcome variable. POST is built on a kernel machine regression framework and inherited the advantages including flexibly model complex microbiome effects (e.g., effects from opposite direction), easily adjust for covariates, and accommodate both continuous and binary outcomes.

Usage

```

post(
  y,
  OTU,
  tree = NULL,
  X = NULL,
  cValues = seq(from = 0, to = 0.05, by = 0.01)
)

```

Arguments

y	A numerical vector. The outcome of interest. Data can be binary or continuous.
OTU	A matrix object. The operational taxonomic units (OTU). Data can be provided as counts or as proportions. Each row indicates a single sample; each column a single OTU. NA/0 values are allowed, but their presence will trigger a shift of all data by a small internally defined value. The matrix must include column headers providing unique identification for each OTU; these identifiers are expected to be included in the tip labels of the input tree object. Any identifiers that are not included as tip labels are removed from the analysis.
tree	An object of class "phylo", "hclust", "phylog", a matrix object or NULL. If NULL, only the single OTU test will be estimated. Objects of class "phylo", "hclust", and "phylog" are phylogenetic trees, the tip labels of which must include all of the identifiers used as column headers of OTU. If a matrix, a square symmetric matrix containing the pairwise distances between OTUs as defined by the branch lengths. Note that the full tree should be provided/used and should not be subset or truncated, even if OTU does not contain all tips. See details for further information.
X	A data.frame object, matrix object or NULL. The covariates data. If NULL, an intercept only model is assumed. Factor covariates are allowed.
cValues	A numeric vector. The c values at which p-values are to be estimated. The default is a vector of evenly spaced values between zero and the recommended maximum value for OTUs defined at 97% sequence similarity, c_max = 0.05. If no tree is provided, cValues will be set to 0.

Details

It is assumed that the OTU table is defined by a 97% sequence similarity. Though this threshold is not (cannot be) enforced in the implementation, the recommended maximum c-value may not be appropriate for other thresholds.

There are numerous packages available for generating phylogenetic trees. The sole purpose of the tree input is to enable the calculation of pairwise distances between the pairs of tips using branch lengths. Because it is not feasible to support this functionality for all packages that generate such trees, the package allows for the specification of the distance matrix as an alternative to providing a specifically formatted tree object. For objects of class "phylo", "hclust", and "phylog", the ape package provides tools to obtain the distance matrix, and these tools are used in this implementation. For all others, the distance matrix (defined by branch lengths) must be provided by the user through input tree.

Value

Returns a POST object. Analysis results are provided as a matrix, the contents of which will depend on the inputs selected for the analysis. Possible columns include

POST_pvalue A numeric object. The POST p-value.
SO_pvalue A numeric object. The single OTU test p-value.
BEST_C A numeric object. The c value corresponding to the minimum POST p-value.

Row names indicate the OTU to which the results pertain. Results are ordered according to the POST (if tree provided) or SO (if tree not provided) p-values.

References

Huang, C., Callahan, B., Wu, M. C., Holloway, S. T., Brochu, H., Lu, W., Peng, X., and Tzeng, J-Y. (2021). Phylogeny-guided microbiome OTU-specific association test (POST). *Bioinformatics*, under revision.

Examples

```
data("POSTmData")

y <- as.integer(x = metadata[, "GC"] == "BV")
X <- metadata[, "mRace"]

result <- post(y = y,
               X = X,
               OTU = otu[, 1:20],
               tree = otutree,
               cValues = seq(0, 0.05, by=0.01))
```

POSTmData

Example Dataset For Illustration

Description

This data set is provided for the purposes of illustrating the use of the software. It was adapted from the vaginal microbiome dataset from Subramaniam et al. (2016). The original dataset consists of 39 individuals with 19 bacterial vaginosis (BV) patients and 20 healthy controls. The sequencing data and metadata are publicly available at NCBI SRA database (PRJNA600021). Initial data processing, lead to 2711 OTUs formed at 97% sequence similarity. FastTree (Price et al, 2010) was used to construct the phylogenetic tree. The OTUs data have been filtered to exclude those with abundance < 0.005% and prevalence < 10%.

Usage

```
data(POSTmData)
```

Format

POSTmData contains 4 objects.

metadata a data.frame containing 30 participants. The data.frame contains 3 columns

SampleID Character. Unique participant identifiers.

mRace A dichotomous covariate describing the maternity race.

GC Gestational complication indicator. "BV" indicates that individual was diagnosed with bacterial vaginosis; "Normal" otherwise

otu a 39 x 189 matrix containing OTU count data

otutree a phylogenetic tree constructed using all 2711 OTU sequences

otuseq a character vector of OTU sequencing data used to construct the phylogenetic tree

print

Print the Primary Results of a post() Analysis

Description

Print the primary results of a post() analysis.

Usage

```
## S3 method for class 'POST'
print(x, ..., siglevel = 1)
```

Arguments

x A POST object. The value object returned by a call to post()

... Further arguments passed to or from other methods

siglevel A numeric object. If < 1, shows only the OTUs for which the p-value is below the specified siglevel.

Value

No return value, called to display key results.

Examples

```
data("POSTmData")

y <- as.integer(x = metadata[,"GC"] == "BV")
X <- metadata[,"mRace"]

result <- post(y = y,
              X = X,
              OTU = otu[,1:20],
```

```
tree = otutree,  
cValues = seq(0,0.05,by=0.01))  
print(x = result)
```


Index

* **datasets**

POSTmData, [6](#)

metadata (POSTmData), [6](#)

otu (POSTmData), [6](#)

otuseq (POSTmData), [6](#)

otutree (POSTmData), [6](#)

p.adjust, [2](#)

plot, [3](#)

post, [4](#)

POSTmData, [6](#)

print, [7](#)