Package: LDM (via r-universe)

September 1, 2024

Type Package

Title Testing Hypotheses About the Microbiome using the Linear Decomposition Model

Version 6.0.1

Date 2023-09-07

Description A single analysis path that includes distance-based ordination, global tests of any effect of the microbiome, and tests of the effects of individual taxa with false-discovery-rate (FDR) control. It accommodates both continuous and discrete covariates as well as interaction terms to be tested either singly or in combination, allows for adjustment of confounding covariates, and uses permutation-based p-values that can control for sample correlations. It can be applied to transformed data, and an omnibus test can combine results from analyses conducted on different transformation scales. It can also be used for testing presence-absence associations based on infinite number of rarefaction replicates, testing mediation effects of the microbiome, analyzing censored time-to-event outcomes, and for compositional analysis by fitting linear models to centered-log-ratio taxa count data.

License GPL (>=2)

RoxygenNote 7.2.3

Depends R (>= 3.5.0)

Imports GUniFrac, vegan, permute, parallel, BiocParallel, matrixStats, castor, phangorn, modeest

Suggests R.rsp, testthat, survival

VignetteBuilder R.rsp

URL https://github.com/yijuanhu/LDM

BugReports https://github.com/yijuanhu/LDM/issues

Encoding UTF-8

LazyData true

Repository https://yijuanhu.r-universe.dev

RemoteUrl https://github.com/yijuanhu/ldm

RemoteRef HEAD

RemoteSha 7751ed69bad1b7a4daa0bd04902d501285d88c03

Contents

---

adjust.data.by.covariates

Adjusting data (distance matrix and OTU table) by covariates

Description

This function produces adjusted distance matrix and OTU table (if provided) after removing the effects of covariates (e.g., confounders). Observations with any missing data are removed.

Usage

```
adjust.data.by.covariates(
  formula = NULL,
  data = .GlobalEnv,
  otu.table = NULL,
  tree = NULL,
  dist.method = "bray",
  binary = FALSE,
  dist = NULL,
  square.dist = TRUE,
  center.dist = TRUE,
  scale.otu.table = TRUE,
```

```
    center.otu.table = TRUE,
    freq.scale.only = FALSE
)
```

Arguments

formula
: a symbolic description of the covariate model in the form ~ model, where model is specified in the same way as for lm or glm. For example, ~ a + b specifies a model with the main effects of covariates a and b, and ~ a*b, equivalently ~ a + b + a:b, specifies a model with the main effects of a and b as well as their interaction.

data
: an optional data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the covariates. If not found in data, the covariates are taken from environment (formula), typically the environment from which adjust.data.by.covariates is called. The default is .GlobalEnv.

otu.table
: the n.obs by n.otu matrix of read counts. If provided, an adjusted (and column-centered) OTU table at the frequency (i.e., relative abundance) scale and an adjusted (and columnn-centered) OTU table at the arcsin-root-transformed frequency scale are output. If provided, it is also used for calculating the distance matrix unless the distance matrix is directly imported through dist. The default is NULL.

tree
: a phylogenetic tree. Only used for calculating a phylogenetic-tree-based distance matrix. Not needed if the calculation of requested distance does not require a phylogenetic tree, or if the distance matrix is directly imported through dist. The default is NULL.

dist.method
: method for calculating the distance measure, partial match to all methods supported by vegdist in the vegan package (i.e., "manhattan", "euclidean", "canberra", "bray", "kulczynski", "jaccard", "gower", "altGower", "morisita", "horn", "mountford", "raup" , "binomial", "chao", "cao", "mahalanobis") as well as "hellinger" and "wt-unifrac". The default is "bray". For more details, see the dist.method argument in the ldm function.

binary
: the "binary" parameter in vegdist. The default is FALSE.

dist
: a distance matrix. Can be either an object of class "dist" or "matrix". The elements of the distance matrix will be squared and then the matrix will be centered if the default choices square.dist=TRUE and center.dist=TRUE are used. If dist=NULL, the distance matrix is calculated from the otu.table, using the value of dist.method (and tree if required). The default is NULL.

square.dist
: a logical variable indicating whether to square the distance matrix. The default is TRUE.

center.dist
: a logical variable indicating whether to center the distance matrix as described by Gower (1966). The default is TRUE.

scale.otu.table
: a logical variable indicating whether to scale the rows of the OTU table for the frequency scale. For count data, this corresponds to dividing by the library size to give relative frequencies. The default is TRUE.

center.otu.table

>a logical variable indicating whether to center the columns of the OTU table. The OTU table should be centered if the distance matrix has been centered. Applied to both OTU tables at frequency and transformed scales. The default is TRUE.

freq.scale.only

>a logical variable indicating whether to provide adjusted frequency-scale OTU table only (not adjusted OTU table at the arcsin-root transformed frequency scale). The default is FALSE.

Value

a list consisting of

adj.dist   the (squared/centered) distance matrix after adjustment of covariates.

y.freq   the (column-centered) frequency-scale OTU table after adjustment of covariates.

y.tran   the (column-centered) arcsin-root-transformed OTU table after adjustment of covariates.

Author(s)

Yi-Juan Hu <yijuan.hu@emory.edu>, Glen A. Satten <gsatten@emory.edu>

Examples

```
adj.data <- adjust.data.by.covariates(formula= ~ Sex + AntibioticUse, data=throat.meta,
                                      otu.table=throat.otu.tab5, dist.method="bray")
```

---

avgdist.squared	*Averaging the squared distance matrices each calculated from a rarefied OTU table*

Description

This function computes a distance matrix for each rarefied OTU table, square the distance matrix (in an element-wise manner), and then average the squared distance matrices.

Usage

```
avgdist.squared(
  otu.table,
  dist.method = "jaccard",
  tree = NULL,
  scale.otu.table = FALSE,
  n.rarefy = 100,
  binary = TRUE,
  seed = 123
)
```

Arguments

otu.table	the n.obs by n.otu matrix of read counts.
dist.method	method for calculating the distance measure, partial match to all methods supported by vegdist in the vegan package. The default is "jaccard". For more details, see the dist.method argument in the ldm function.
tree	the phylogeneic tree. The default is NULL.
scale.otu.table	
	a logical variable indicating whether to scale the rows of the OTU table. For count data, this corresponds to dividing by the library size to give relative frequencies. The default is FALSE.
n.rarefy	number of rarefactions. The default is 100.
binary	the "binary" parameter in vegdist. The default is TRUE.
seed	a single-value integer seed for the random process of drawing rarefaction replicates. The seed is user supplied or internally generated. The default is 123.

Value

a single matrix object

D2.avg	The average of the squared distance matrices.

Author(s)

Yi-Juan Hu <yijuan.hu@emory.edu>, Glen A. Satten <gsatten@emory.edu>

Examples

```
dist.avg.D2 <- avgdist.squared(throat.otu.tab5, dist.method="jaccard", n.rarefy=100)
```

---

jaccard.mean	*Expected value of the Jaccard distance matrix*

Description

This function computes the expected value of the Jaccard distance matrix over rarefaction replicates.

Usage

```
jaccard.mean(
  otu.table,
  rarefy.depth = min(rowSums(otu.table)),
  first.order.approx.only = FALSE
)
```

Arguments

`otu.table`	the `n.obs` by `n.otu` matrix of read counts.
`rarefy.depth`	rarefaction depth. The default is the minimum library size observed in the OTU table.
`first.order.approx.only`	
	a logical value indicating whether to calculate the expected value using the first order approixmation by the delta method. The default is FALSE, using the second order approixmation.

Value

a list consisting of

`jac.mean.o1`	Expected Jaccard distance matrix by the first order approixmation.
`jac.mean.o2`	Expected Jaccard distance matrix by the second order approixmation.
`jac.mean.sq.o1`	Expected squared Jaccard distance matrix by the first order approixmation.
`jac.mean.sq.o2`	Expected squared Jaccard distance matrix by the second order approixmation.

Author(s)

Yi-Juan Hu <yijuan.hu@emory.edu>, Glen A. Satten <gsatten@emory.edu>

Examples

```
res.jaccard <- jaccard.mean( throat.otu.tab5 )
```

---

ldm	*Testing hypotheses about the microbiome using a linear decomposition model (LDM)*

Description

This function allows you to 1. simultaneously test the global association with the overall microbiome composition and individual OTU associations to give coherent results; 2. test hypotheses based on data at both the frequency (i.e., relative abundance) and arcsine-root-transformed frequency scales, and perform an "omnibus" test that combines results from analyses conducted on the two scales; 3. test presence-absence associations based on infinite number of rarefaction replicates; 4. handle complex design features such as confounders, interactions, and clustered data (with between- and within-cluster covariates); 5. test associations with a survival outcome (i.e., censored survival times); 6. perform mediation analysis of the microbiome; 7. perform the omnibus test LDM-omni3 that combines results from analyses conducted on the frequency, arcsine-root-transformed, and presence-absence scales.

Usage

```
ldm(
  formula,
  other.surv.resid = NULL,
  data = .GlobalEnv,
  tree = NULL,
  dist.method = "bray",
  dist = NULL,
  cluster.id = NULL,
  strata = NULL,
  how = NULL,
  perm.within.type = "free",
  perm.between.type = "none",
  perm.within.ncol = 0,
  perm.within.nrow = 0,
  n.perm.max = NULL,
  n.rej.stop = 100,
  seed = NULL,
  fdr.nominal = 0.1,
  square.dist = TRUE,
  scale.otu.table = TRUE,
  center.otu.table = TRUE,
  freq.scale.only = FALSE,
  binary = FALSE,
  n.rarefy = 0,
  test.mediation = FALSE,
  test.omni3 = FALSE,
  comp.anal = FALSE,
  comp.anal.adjust = "median",
  n.cores = 4,
  verbose = TRUE
)
```

Arguments

formula   a symbolic description of the model to be fitted. The details of model specification are given under "Details".

other.surv.resid

     a vector of data, usually the Martingale or deviance residuals from fitting the Cox model to the survival outcome (if it is the outcome of interest) and other covariates.

data   an optional data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the covariates of interest and confounding covariates. If not found in data, the covariates are taken from environment(formula), typically the environment from which ldm is called. The default is .GlobalEnv.

tree   a phylogenetic tree. Only used for calculating a phylogenetic-tree-based distance matrix. Not needed if the calculation of the requested distance does not

	involve a phylogenetic tree, or if the distance matrix is directly imported through `dist`.
dist.method	method for calculating the distance measure, partial match to all methods supported by `vegdist` in the `vegan` package (i.e., "manhattan", "euclidean", "canberra", "bray", "kulczynski", "jaccard", "gower", "altGower", "morisita", "horn", "mountford", "raup" , "binomial", "chao", "cao", "mahalanobis") as well as "hellinger" and "wt-unifrac". The Hellinger distance measure (dist.method="hellinger") takes the form $0.5*E$, where E is the Euclidean distance between the square-root-transformed frequency data. The weighted UniFrac distance (dist.method="wt-unifrac") is calculated by interally calling GUniFrac in the GUniFrac package. Not used when anything other than `dist=NULL` is specified for `dist`. The default is "bray".
dist	a distance matrix. Can be an object of class either "dist" or "matrix". The elements of the distance matrix will be squared and then the matrix will be centered if the default choices `square.dist=TRUE` and `center.otu.table=TRUE` are used. If `dist=NULL`, the distance matrix is calculated from the `otu.table`, using the value of `dist.method` (and `tree` if required). The default is NULL.
cluster.id	character or factor variable that identifies clusters. The default value cluster.id=NULL if the observations are not clustered (i.e., are independent).
strata	a character or factor variable that defines strata (groups), within which to constrain permutations. The default is NULL.
how	a permutation control list, for users who want to specify their own call to the how function from the `permute` package. The default is NULL.
perm.within.type	a character string that takes values "free", "none", "series", or "grid". The default is "free" (for random permutations).
perm.between.type	a character string that takes values "free", "none", or "series". The default is "none".
perm.within.ncol	a positive integer, only used if perm.within.type="grid". The default is 0. See documentation for permute package for additional details.
perm.within.nrow	a positive integer, only used if perm.within.type="grid". The default is 0. See documentation for permute package for additional details.
n.perm.max	the maximum number of permutations. The default is NULL, in which case a maximum of 5000 permutations are used for the global test and a maximum of n.otu * n.rej.stop * (1/fdr.nominal) are used for the OTU test, where `n.otu` is the number of OTUs. If a numeric value for `n.perm.max` is specified, this value is used for both global and OTU-level tests.
n.rej.stop	the minimum number of rejections (i.e., the permutation statistic exceeds the observed statistic) to obtain before stopping. The default is 100.
seed	a user-supplied integer seed for the random number generator in the permutation procedure. The default is NULL; with the default value, an integer seed will be generated internally and randomly. In either case, the integer seed will be stored in the output object in case the user wants to reproduce the permutation replicates.

fdr.nominal	the nominal FDR value. The default is 0.1.
square.dist	a logical variable indicating whether to square the distance matrix. The default is TRUE.
scale.otu.table	
	a logical variable indicating whether to scale the rows of the OTU table. For count data, this corresponds to dividing by the library size to give frequencies (i.e., relative abundances). Does not affect the tran scale. The default is TRUE.
center.otu.table	
	a logical variable indicating whether to center the columns of the OTU table. The OTU table should be centered if the distance matrix has been centered. Applied to both the frequency and transformed scales. The default is TRUE.
freq.scale.only	
	a logical variable indicating whether to perform analysis of the frequency-scale data only (not the arcsin-root transformed frequency data and the omnibus test). The default is FALSE.
binary	a logical value indicating whether to perform presence-absence analysis. The default is FALSE (analyzing relative abundance data).
n.rarefy	an integer-valued number of rarefactions. The value "all" is also allowed, and requests the LDM-A method that essentially aggregate information from all rarefactions. The default is 0 (no rarefaction).
test.mediation	a logical value indicating whether to perform the mediation analysis. The default is FALSE. If TRUE, the formula takes the specific form `otu.table ~ exposure + outcome` or most generally `otu.table	(set of confounders) ~ (set of exposures) + (set of outcomes)`.
test.omni3	a logical value indicating whether to perform the new omnibus test (LDM-omni3). The default is FALSE.
comp.anal	a logical value indicating whether the centered-log-ratio taxa count data are used (LDM-clr). The default is FALSE.
comp.anal.adjust	
	a character string that takes value "median" or "mode" to choose the estimator for the beta mean (Hu and Satten, 2023). The default is "median".
n.cores	The number of cores to use in parallel computing, i.e., at most how many child processes will be run simultaneously. The default is 4.
verbose	a logical value indicating whether to generate verbose output during the permutation process. Default is TRUE.

Details

The formula has the form

`otu.table ~ (first set of covariates) + (second set of covariates) ... + (last set of covariates)`

or

`otu.table | confounders ~ (first set of covariates) + (second set of covariates) ... + (last set of covariates)`

where `otu.table` is the OTU table with rows for samples and columns for OTUs and each set of covariates are enclosed in parentheses. The covariates in each submodel (set of covariates) are tested jointly, after projecting off terms in submodels that appear earlier in the model.

For example, given OTU table y and a data frame `metadata` that contains 4 covariates, a, b, c and d, some valid formulas would be:

y ~ a + b + c + d ### no confounders, 4 submodels (i.e., sets of covariates)

y ~ (a+b) + (c+d) ### no confounders, 2 submodels each having 2 covariates

y | b ~ (a+c) + d ### b is a confounder, submodel 1 is (a+c), and submodel 2 is d

y | b+c ~ a*d ### there are 2 confounders b and c; there is 1 submodel consisting of the three terms a, d, and a:d (interaction). This example is equivalent to y | b+c ~ (a+d+a:d)

y | as.factor(b) ~ (a+d) + a:d ### the confounder b will be treated as a factor variable, submodel 1 will have the main effects a and d, and submodel 2 will have only the interaction between a and d

y | as.factor(b) ~ (a) + (d) + (a:d) ### there are 3 submodels a, d, and a:d. Putting paratheses around a single variable is allowed but not necessary.

Submodels that combine character and numeric values are allowed; character-valued variables are coerced into factor variables. Confounders are distinguished from other covariates as test statistics are not calculated for confounders (which are included for scientific reasons, not by virtue of significance test results); consequently they also do not contribute to stopping criteria. If tests of confounders are desired, confounders should put on the right hand side of the formula as the first submodel.

For testing mediation effects of the microbiome that mediate the effect of the exposure(s) on the outcome(s), the formula takes the specific form:

otu.table ~ exposure + outcome

or most generally

otu.table | (set of confounders) ~ (set of exposures) + (set of outcomes)

in which there should be exactly two terms on the right hand side of the regression, corresponding to the exposure(s) and the outcome(s), the outcome(s) must appear after the exposure(s), and the covariates or confounders must appear after |.

LDM uses two sequential stopping criteria. For the global test, LDM uses the stopping rule of Besag and Clifford (1991), which stops permutation when a pre-specified minimum number (default=100) of rejections (i.e., the permutation statistic exceeded the observed test statistic) has been reached. For the OTU-specific tests, LDM uses the stopping rule of Sandve et al. (2011), which stops permutation when every OTU test has either reached the pre-specified number (default=100) of rejections or yielded a q-value that is below the nominal FDR level (default=0.1). As a convention, we call a test "stopped" if the corresponding stopping criterion has been satisfied. Although all tests are always terminated if a pre-specified maximum number (see description of `n.perm.max` in Arguments list) of permutations have been generated, some tests may not have "stopped". This typically occurs when the relevant p-value is small or near the cutoff for inclusion in a list of significant findings; for global tests meeting the stopping criterion is not critical, but caution is advised when interpreting OTU-level tests that have not stopped as additional OTUs may be found with a larger number of permutations.

Value

a list consisting of

x	the (orthonormal) design matrix X as defined in Hu and Satten (2020)
dist	the (squared/centered) distance matrix
mean.freq	the mean relative abundance of OTUs (the column means of the frequency-scale OTU table)
y.freq	the frequency-scale OTU table, scaled and centered if so specified
d.freq	a vector of the non-negative diagonal elements of D that satisfies x^T y.freq = D v^T
v.freq	the v matrix with unit columns that satisfies x^T y.freq = D v^T
y.tran	the (column-centered) arcsin-root-transformed OTU table
d.tran	a vector of the non-negative diagonal elements of D that satisfies x^T y.tran = D v^T
v.tran	the v matrix with unit columns that satisfies x^T y.tran = D v^T
low	a vector of lower indices for confounders (if there is any) and submodels
up	a vector of upper indices for confounders (if there is any) and submodels
beta	a matrix of effect sizes of every trait on every OTU
phi	a matrix of probabilities that the rarefied count of an OTU in a sample is non-zero
VE.global.freq.confounders	
	Variance explained (VE) by confounders, based on the frequency-scale data
VE.global.freq.submodels	
	VE by each submodel, based on the frequency-scale data
VE.global.freq.residuals	
	VE by each component in the residual distance, based on the frequency-scale data
VE.otu.freq.confounders	
	Contribution of each OTU to VE by confounders, based on the frequency-scale data
VE.otu.freq.submodel	
	Contribution of each OTU to VE by each submodel, based on the frequency-scale data
VE.global.tran.confounders	
	VE by confounders, based on the arcsin-root-transformed frequency data
VE.global.tran.submodels	
	VE by each submodel, based on the arcsin-root-transformed frequency data
VE.global.tran.residuals	
	VE by each component in the residual distance, based on the arcsin-root-transformed frequency data
VE.otu.tran.confounders	
	Contribution of each OTU to VE by confounders, based on the arcsin-root-transformed frequency data

VE.otu.tran.submodels

> Contribution of each OTU to VE by each submodel, based on the arcsin-root-transformed frequency data

VE.df.confounders

> Degree of freedom (i.e., number of components) associated with the VE for confounders

VE.df.submodels

> Degree of freedom (i.e., number of components) associated with the VE for each submodel

F.global.freq   F statistics for testing each submodel, based on the frequency-scale data

F.global.tran   F statistics for testing each submodel, based on the arcsin-root-transformed frequency data

F.otu.freq   F statistics for testing each OTU for each submodel, based on the frequency-scale data

F.otu.tran   F statistics for testing each OTU for each submodel, based on the arcsin-root-transformed data

p.global.freq   p-values for the global test of each set of covariates based on the frequency-scale data

p.global.tran   p-values for the global test of each set of covariates based on the arcsin-root-transformed frequency data

p.global.pa   p-values for the global test of each set of covariates based on the presence-absence data

p.global.omni   p-values for the global test of each set of covariates based on the omnibus statistics in LDM-omni, which are the minima of the p-values obtained from the frequency scale and the arcsin-root-transformed frequency data as the final test statistics, and use the corresponding minima from the permuted data to simulate the null distributions

p.global.harmonic

> p-values for the global test of each set of covariates based on the Harmonic-mean p-value combination method applied to the OTU-level omnibus p-values

p.global.fisher

> p-values for the global test of each set of covariates based on the Fisher p-value combination method applied to the OTU-level omnibus p-values

p.global.omni3   p-values for the global test of each set of covariates based on the omnibus test LDM-omni3

p.global.freq.OR, p.global.tran.OR, p.global.pa.OR, p.global.omni.OR,
p.global.harmonic.OR, p.global.fisher.OR, p.global.omni3.OR

> global p-values for testing other.surv.resid

p.global.freq.com,   p.global.tran.com,   p.global.pa.com,
p.global.omni.com,   p.global.harmonic.com,   p.global.fisher.com,
p.global.omni3.com

> global p-values from the combination test that combines the results from analyzing both the Martingale and deviance residuals from a Cox model (one of them is supplied by other.surv.resid)

p.otu.freq   p-values for the OTU-specific tests based on the frequency scale data

`p.otu.tran`	p-values for the OTU-specific tests based on the arcsin-root-transformed frequency data
`p.otu.pa`	p-values for the OTU-specific tests based on the presence-absence data
`p.otu.omni`	p-values for the OTU-specific tests based on the omnibus test LDM-omni
`p.otu.omni3`	p-values for the OTU-specific tests based on the omnibus test LDM-omni3
`q.otu.freq`	q-values (i.e., FDR-adjusted p-values) for the OTU-specific tests based on the frequency scale data
`q.otu.tran`	q-values for the OTU-specific tests based on the arcsin-root-transformed frequency data
`q.otu.pa`	q-values (i.e., FDR-adjusted p-values) for the OTU-specific tests based on the presence-absence data
`q.otu.omni`	q-values for the OTU-specific tests based on the omnibus test LDM-omni
`q.otu.omni3`	q-values for the OTU-specific tests based on the omnibus test LDM-omni3

`p.otu.freq.OR,` `p.otu.tran.OR,` `p.otu.pa.OR,` `p.otu.omni.OR,` `p.otu.omni3.OR,` `q.otu.freq.OR,` `q.otu.tran.OR,` `q.otu.pa.OR,` `q.otu.omni.OR, q.otu.omni3.OR`

> OTU-level p-values and q-values for testing `other.surv.resid`

`p.otu.freq.com,` `p.otu.tran.com,` `p.otu.pa.com,` `p.otu.omni.com,` `p.otu.omni3.com,` `q.otu.freq.com,` `q.otu.tran.com,` `q.otu.pa.com,` `q.otu.omni.com, q.otu.omni3.com`

> OTU-level p-values and q-values from the combination tests that combine the results from analyzing both the Martingale and deviance residuals from a Cox model (one of them is supplied by `other.surv.resid`)

`detected.otu.freq`

> detected OTUs (whose names are found in the column names of the OTU table) at the nominal FDR, based on the frequency scale data

`detected.otu.tran`

> detected OTUs based on the arcsin-root-transformed frequency data

`detected.otu.pa`

> detected OTUs based on the presence-absence data

`detected.otu.omni`

> detected OTU based on the omnibus test LDM-omni

`detected.otu.omni3`

> detected OTU based on the omnibus test LDM-omni3

`detected.otu.freq.OR,` `detected.otu.tran.OR,` `detected.otu.pa.OR,` `detected.otu.omni.OR, detected.otu.omni3.OR`

> detected OTUs for `other.surv.resid`

`detected.otu.freq.com, detected.otu.tran.com, detected.otu.pa.com,` `detected.otu.omni.com, detected.otu.omni3.com`

> detected OTUs by the combination tests that combines the Martingale and deviance residuals from a Cox model (one of them is supplied by `other.surv.resid`)

`med.p.global.freq,` `med.p.global.tran,` `med.p.global.omni,` `med.p.global.pa,` `med.p.global.harmonic,` `med.p.global.fisher,` `med.p.global.omni3`

> p-values for the global tests of the overall mediation effect by the microbiome

```
med.p.global.freq.OR,   med.p.global.tran.OR,   med.p.global.omni.OR,
med.p.global.pa.OR,                             med.p.global.harmonic.OR,
med.p.global.fisher.OR, med.p.global.omni3.OR
```
>   p-values for the global tests of the overall mediation effect by the microbiome,
>   when the outcome is `other.surv.resid`

```
med.p.global.freq.com, med.p.global.tran.com, med.p.global.omni.com,
med.p.global.pa.com,                           med.p.global.harmonic.com,
med.p.global.fisher.com, med.p.global.omni3.com
```
>   p-values for the global tests of the overall mediation effect by the microbiome,
>   combining the results from analyzing both the Martingale and deviance residuals
>   as outcomes

```
med.detected.otu.freq, med.detected.otu.tran, med.detected.otu.omni,
med.detected.otu.pa, med.detected.otu.omni3
```
>   detected mediating OTUs

```
med.detected.otu.freq.OR,                      med.detected.otu.tran.OR,
med.detected.otu.omni.OR,                      med.detected.otu.pa.OR,
med.detected.otu.omni3.OR
```
>   detected mediating OTUs for the outcome `other.surv.resid`

```
med.detected.otu.freq.com,                     med.detected.otu.tran.com,
med.detected.otu.omni.com,                     med.detected.otu.pa.com,
med.detected.otu.omni3.com
```
>   detected mediating OTUs, combining the results from analyzing both the Mar-
>   tingale and deviance residuals as outcomes

`n.perm.completed`
>   number of permutations completed

`global.tests.stopped`
>   a logical value indicating whether the stopping criterion has been met by all
>   global tests

`otu.tests.stopped`
>   a logical value indicating whether the stopping criterion has been met by all
>   OTU-specific tests

`seed`   the seed that is user supplied or internally generated, stored in case the user
>   wants to reproduce the permutation replicates

Author(s)

Yi-Juan Hu <yijuan.hu@emory.edu>, Glen A. Satten <gsatten@emory.edu>

References

Hu YJ, Satten GA (2020). Testing hypotheses about the microbiome using the linear decomposition model (LDM) Bioinformatics, 36(14), 4106-4115.

Hu YJ, Lane A, and Satten GA (2021). A rarefaction-based extension of the LDM for testing presence-absence associations in the microbiome. Bioinformatics, 37(12):1652-1657.

Zhu Z, Satten GA, Caroline M, and Hu YJ (2020). Analyzing matched sets of microbiome data using the LDM and PERMANOVA. Microbiome, 9(133), https://doi.org/10.1186/s40168-021-01034-9.

Zhu Z, Satten GA, and Hu YJ (2022). Integrative analysis of relative abundance data and presence-absence data of the microbiome using the LDM. Bioinformatics, doi.org/10.1093/bioinformatics/btac181.

Yue Y and Hu YJ (2021) A new approach to testing mediation of the microbiome using the LDM. bioRxiv, https://doi.org/10.1101/2021.11.12.468449.

Hu Y, Li Y, Satten GA, and Hu YJ (2022) Testing microbiome associations with censored survival outcomes at both the community and individual taxon levels. bioRxiv, doi.org/10.1101/2022.03.11.483858.

Examples

```
res.ldm <- ldm(formula=throat.otu.tab5 | (Sex+AntibioticUse) ~ SmokingStatus+PackYears,
               data=throat.meta, seed=67817, fdr.nominal=0.1, n.perm.max=1000, n.cores=1,
               verbose=FALSE)
```

---

permanovaFL	*PERMANOVA test of association based on the Freedman-Lane permutation scheme*

Description

This function performs the PERMANOVA test that can allow adjustment of confounders and control of clustered data. It can also be used for testing presence-absence associations based on infinite number of rarefaction replicates. As in `ldm`, `permanovaFL` allows multiple sets of covariates to be tested, in the way that the sets are entered sequentially and the variance explained by each set is that part that remains after the previous sets have been fit. It allows testing of a survival outcome, by using the Martingale or deviance residual (from fitting a Cox model to the survival outcome and other covariates) as a covariate in the regression. It allows multiple distance matrices and provides an omnibus test in such cases. It also allows testing of the mediation effect of the microbiome in the pathway between the exposure(s) and the outcome(s), where the exposure(s) and outcomes(s) are specified as the first and second (sets of) covariates.

Usage

```
permanovaFL(
  formula,
  other.surv.resid = NULL,
  data = .GlobalEnv,
  tree = NULL,
  dist.method = c("bray"),
  dist.list = NULL,
  cluster.id = NULL,
  strata = NULL,
  how = NULL,
  perm.within.type = "free",
  perm.between.type = "none",
  perm.within.ncol = 0,
  perm.within.nrow = 0,
  n.perm.max = 5000,
```

```
    n.rej.stop = 100,
    seed = NULL,
    square.dist = TRUE,
    center.dist = TRUE,
    scale.otu.table = c(TRUE),
    binary = c(FALSE),
    n.rarefy = 0,
    test.mediation = FALSE,
    n.cores = 4,
    verbose = TRUE
)
```

Arguments

formula
a symbolic description of the model to be fitted in the form of data.matrix ~ sets of covariates or data.matrix | confounders ~ sets of covariates. The details of model specification are given in "Details" of ldm. Additionally, in permanovaFL, the data.matrix can be either an OTU table or a distance matrix. If it is an OTU table, the distance matrix will be calculated internally using the OTU table, tree (if required), and dist.method. If data.matrix is a distance matrix (having class dist or matrix), it can be squared and//or centered by specifying square.dist and center.dist (described below). Distance matrices are distinguished from OTU tables by checking for symmetry of as.matrix(data.matrix).

other.surv.resid
a vector of data, usually the Martingale or deviance residuals from fitting the Cox model to the survival outcome (if it is the outcome of interest) and other covariates.

data
an optional data frame, list or environment (or object coercible to a dataframe) containing the covariates of interest and confounding covariates. If not found in data, the covariates are taken from environment(formula), typically the environment from which permanovaFL is called. The default is .GlobalEnv.

tree
a phylogenetic tree. Only used for calculating a phylogenetic-tree-based distance matrix. Not needed if the calculation of the requested distance does not involve a phylogenetic tree, or if a distance matrix is directly imported through formula.

dist.method
a vector of methods for calculating the distance measure, partial match to all methods supported by vegdist in the vegan package (i.e., "manhattan", "euclidean", "canberra", "bray", "kulczynski", "jaccard", "gower", "altGower", "morisita", "horn", "mountford", "raup" , "binomial", "chao", "cao", "mahalanobis") as well as "hellinger" and "wt-unifrac". Not used if a distance matrix is specified in formula or dist.list. The default is c("bray"). For more details, see the dist.method argument in the ldm function.

dist.list
a list of pre-calculated distance matrices.

cluster.id
cluster identifiers. The default is value of NULL should be used if the observations are not in clusters (i.e., independent).

strata	a factor variable (or, character variable converted into a factor) to define strata (groups), within which to constrain permutations. The default is NULL.
how	a permutation control list, for users who want to specify their permutation control list using the how function from the permute R package. The default is NULL.
perm.within.type	
	a character string that takes values "free", "none", "series", or "grid". The default is "free" (for random permutations).
perm.between.type	
	a character string that takes values "free", "none", or "series". The default is "none".
perm.within.ncol	
	a positive integer, only used if perm.within.type="grid". The default is 0. See the documentation for the R package permute for further details.
perm.within.nrow	
	a positive integer, only used if perm.within.type="grid". The default is 0. See the documentation for the R package permute for further details.
n.perm.max	the maximum number of permutations. The default is 5000.
n.rej.stop	the minimum number of rejections (i.e., the permutation statistic exceeds the observed statistic) to obtain before stopping. The default is 100.
seed	a user-supplied integer seed for the random number generator in the permutation procedure. The default is NULL; with the default value, an integer seed will be generated internally and randomly. In either case, the integer seed will be stored in the output object in case the user wants to reproduce the permutation replicates.
square.dist	a logical variable indicating whether to square the distance matrix. The default is TRUE.
center.dist	a logical variable indicating whether to center the distance matrix as described by Gower (1966). The default is TRUE.
scale.otu.table	
	a vector of logical variables indicating whether to scale the OTU table in calculating the distance matrices in dist.method. For count data, this corresponds to dividing by the library size to give relative abundances. The default is TRUE.
binary	a vector of logical values indicating whether to base the calculation of the distance matrices in dist.method on presence-absence (binary) data. The default is c(FALSE) (analyzing relative abundance data).
n.rarefy	number of rarefactions. The default is 0 (no rarefaction).
test.mediation	a logical value indicating whether to perform the mediation analysis. The default is FALSE. If TRUE, the formula takes the specific form otu.table ~ exposure + outcome or most generally otu.table or distance matrix	(set of confounders) ~ (set of exposures) + (set of outcomes).
n.cores	The number of cores to use in parallel computing, i.e., at most how many child processes will be run simultaneously. The default is 4.
verbose	a logical value indicating whether to generate verbose output during the permutation process. Default is TRUE.

Value

a list consisting of

F.statistics   F statistics for testing each set of covariates

R.squared   R-squared statistic for each set of covariates

F.statistics.OR, R.squared.OR

                   F statistics and R-squared statistic when the last covariate is other.surv.resid

p.permanova   p-values for testing each set of covariates

p.permanova.omni

                   the omnibus p-values (that combines information from multiple distance matrices) for testing each set of covariates

med.p.permanova

                   p-values for testing mediation

med.p.permanova.omni

                   the omnibus p-values for testing mediation

p.permanova.OR, p.permanova.omni.OR

                   when using other.surv.resid as the last covariate

med.p.permanova.OR, med.p.permanova.omni.OR

                   when using other.surv.resid as the outcome in the mediation analysis

p.permanova.com, p.permanova.omni.com

                   the combination test that combines the results from analyzing the Martingale residual and the Deviance residual (one specified in the formula and one specified in other.surv.resid)

med.p.permanova.com, med.p.permanova.omni.com

                   the combination test for the mediation effect

n.perm.completed

                   number of permutations completed

permanova.stopped

                   a logical value indicating whether the stopping criterion has been met by all tests of covariates

seed   the seed that is user supplied or internally generated, stored in case the user wants to reproduce the permutation replicates

Author(s)

Yi-Juan Hu <yijuan.hu@emory.edu>, Glen A. Satten <gsatten@emory.edu>

References

Hu YJ, Satten GA (2020). Testing hypotheses about the microbiome using the linear decomposition model (LDM) Bioinformatics, 36(14), 4106-4115.

Hu YJ and Satten GA (2021). A rarefaction-without-resampling extension of PERMANOVA for testing presence-absence associations in the microbiome. bioRxiv, https://doi.org/10.1101/2021.04.06.438671.

Zhu Z, Satten GA, Caroline M, and Hu YJ (2020). Analyzing matched sets of microbiome data using the LDM and PERMANOVA. Microbiome, 9(133), https://doi.org/10.1186/s40168-021-01034-9.

Hu Y, Li Y, Satten GA, and Hu YJ (2022) Testing microbiome associations with censored survival outcomes at both the community and individual taxon levels. bioRxiv, doi.org/10.1101/2022.03.11.483858.

Examples

```
res.perm <- permanovaFL(throat.otu.tab5 | (Sex+AntibioticUse) ~ SmokingStatus+PackYears,
                data=throat.meta, dist.method="bray", seed=82955, n.perm.max=1000, n.cores=1,
                    verbose=FALSE)
```

---

sim.meta   *Metadata of the simulated microbiome samples*

Description

This data set includes 100 simulated samples from 50 subjects, with two samples from each subject. The metadata contain subject IDs and data from four variables, two (X.between and Y.between) varying between subjects and two (X.within and Y.within) varying within subjects.

Usage

```
data("sim.meta")
```

Format

A data frame with 100 observations on 5 variables.

Examples

```
data(sim.meta)
```

---

sim.otu.tab   *OTU count table of the simulated microbiome samples*

Description

This is the OTU table corresponding to the 100 simulated samples in sim.meta. The samples in the rows of the OTU table match with the samples in the rows of sim.meta.

Usage

```
data("sim.otu.tab")
```

Format

A data frame with 100 observations on 813 variables.

Examples

```
data(sim.otu.tab)
```

---

sim.otu.tab5   *Filtered OTU count table of the simulated microbiome samples*

Description

This table was derived from sim.otu.tab, after filtering out OTUs that are present (having non-zero counts) in less than five samples. This filter reduced the number of OTUs from 813 to 593.

Usage

```
data("sim.otu.tab5")
```

Format

A data frame with 100 observations on 593 variables.

Examples

```
data(sim.otu.tab5)
```

---

throat.meta   *Metadata of the throat microbiome samples*

Description

This data set includes samples from the microbiome of the nasopharynx and oropharynx on each side of the body. It were generated to study the effect of smoking on the microbiota of the upper respiratory tract in 60 individuals, 28 smokers and 32 nonsmokers.

Usage

```
data("throat.meta")
```

Format

A data frame with 60 observations on 16 variables.

Source

Charlson ES, Chen J, Custers-Allen R, Bittinger K, Li H, et al. (2010) Disordered Microbial Communities in the Upper Respiratory Tract of Cigarette Smokers. PLoS ONE 5(12): e15216.

References

R package "GUniFrac"

Examples

```
data(throat.meta)
```

---

throat.otu.tab	*OTU count table from 16S sequencing of the throat microbiome samples*

Description

This data set contains 60 subjects with 28 smokers and 32 nonsmokers. Microbiome data were collected from right and left nasopharynx and oropharynx region to form an OTU table with 856 OTUs.

Usage

```
data("throat.otu.tab")
```

Format

A data frame with 60 observations on 856 variables.

Source

Charlson ES, Chen J, Custers-Allen R, Bittinger K, Li H, et al. (2010) Disordered Microbial Communities in the Upper Respiratory Tract of Cigarette Smokers. PLoS ONE 5(12): e15216.

References

R package "GUniFrac"

Examples

```
data(throat.otu.tab)
```

---

throat.otu.tab5	*Filtered OTU count table from 16S sequencing of the throat microbiome samples*

Description

This table was derived from throat.otu.tab, after filtering out OTUs that are present (having non-zero counts) in less than five samples. This filter reduced the number of OTUs from 856 to 233.

Usage

```
data("throat.otu.tab5")
```

Format

A data frame with 60 observations on 233 variables.

Examples

```
data(throat.otu.tab5)
```

| throat.tree | *UPGMA tree of the OTUs from 16S sequencing of the throat micro-biome samples* |

Description

The OTU tree is constructed using UPGMA on the K80 distance matrix of the OTUs. It is a rooted tree of class "phylo".

Usage

```
data("throat.tree")
```

Format

List of 4 data frames.

Source

Charlson ES, Chen J, Custers-Allen R, Bittinger K, Li H, et al. (2010) Disordered Microbial Communities in the Upper Respiratory Tract of Cigarette Smokers. PLoS ONE 5(12): e15216.

References

R package "GUniFrac"

Examples

```
data(throat.tree)
```

---

unifrac.mean *Expected value of the unweighted UniFrac distance matrix*

Description

This function computes the expected value of the unweighted UniFrac distance matrix over rarefaction replicates.

Usage

```
unifrac.mean(
  otu.table,
  tree,
  rarefy.depth = min(rowSums(otu.table)),
  first.order.approx.only = FALSE,
  verbose = TRUE
)
```

Arguments

otu.table	the n.obs by n.otu matrix of read counts.
tree	the phylogeneic tree.
rarefy.depth	rarefaction depth. The default is the minimum library size observed in the OTU table.
first.order.approx.only	
	a logical value indicating whether to calculate the expected value using the first order approixmation by the delta method. The default is FALSE, using the second order approixmation.
verbose	a logical value indicating whether to generate verbose output. Default is TRUE.

Value

a list consisting of

unifrac.mean.o1	
	Expected unweighted UniFrac distance matrix by the first order approixmation.
unifrac.mean.o2	
	Expected unweighted UniFrac distance matrix by the second order approixmation.
unifrac.mean.sq.o1	
	Expected squared unweighted UniFrac distance matrix by the first order approixmation.
unifrac.mean.sq.o2	
	Expected squared unweighted UniFrac distance matrix by the second order approixmation.

Author(s)

Yi-Juan Hu <yijuan.hu@emory.edu>, Glen A. Satten <gsatten@emory.edu>

Examples

```
data(throat.otu.tab5)
data(throat.tree)
res.unifrac <- unifrac.mean( throat.otu.tab5[1:20,], throat.tree)
```

Index