

# Package ‘LDM’

January 31, 2019

**Type** Package

**Title** Testing Hypotheses about the Microbiome using the Linear  
Decomposition Model

**Version** 1.0

**Date** 2019-1-28

**Depends** GUniFrac, vegan

**Suggests** R.rsp

**VignetteBuilder** R.rsp

**Author** Yi-Juan Hu and Glen A. Satten <gas0@cdc.gov>

**Maintainer** Yi-Juan Hu <yijuan.hu@emory.edu>

**Description** The LDM package provides a single analysis path that includes distance-based ordination, global tests of any effect of the microbiome, and tests of the effects of individual OTUs (i.e., operational taxonomic units) with false discovery rate (FDR)-based correction for multiple testing. It accommodates both continuous and discrete variables (e.g., clinical outcomes, environmental factors, treatment groups) 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 correlation (e.g., repeated measurements on the same individual). It can also be applied to transformed data, and an ‘omnibus’ test can easily combine results from analyses conducted on different transformation scales.

**License** GPL (>=2)

**NeedsCompilation** no

**RoxygenNote** 6.1.1.9000

## R topics documented:

adjust.data.by.covariates . . . . .	2
ldm . . . . .	4
permanovaFL . . . . .	9
throat.meta . . . . .	11
throat.otu.tab . . . . .	12
throat.tree . . . . .	12

<b>Index</b>	<b>14</b>
--------------	-----------

---

adjust.data.by.covariates

*Adjust data 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 = NULL,
  otu.table = NULL, tree = NULL, dist.method = "bray", 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 with the form <code>~ model</code> , where <code>model</code> is specified in the same way as for <code>lm</code> or <code>glm</code> . For example, <code>~ a + b</code> specifies a model with the main effects of covariates <code>a</code> and <code>b</code> , and <code>~ a*b</code> , equivalently <code>~ a + b + a:b</code> , specifies a model with the main effects of <code>a</code> and <code>b</code> as well as their interaction.
data	an optional data frame, list or environment (or object coercible by <code>as.data.frame</code> to a data frame) containing the covariates. If not found in <code>data</code> , the covariates are taken from <code>environment(formula)</code> , typically the environment from which <code>adjust.data.by.covariates</code> is called. The default is <code>NULL</code> .
otu.table	the <code>n.obs</code> by <code>n.otu</code> matrix of read counts. If provided, the adjusted (and column-centered) <code>otu.table</code> at both the frequency scale and arcsin-root-transformed frequency scale are outputted. If provided, it is also used for calculating the distance matrix unless the distance matrix is directly imported through <code>dist</code> . The default is <code>NULL</code> .
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 <code>dist</code> . The default is <code>NULL</code> .
dist.method	method for calculating the distance measure, partial match to all methods supported by <code>vegdist</code> in the <code>vegan</code> 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 <code>dist.method</code> argument in the <code>ldm</code> function.
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 <code>square.dist=TRUE</code> and <code>center.dist=TRUE</code> are used. If <code>dist=NULL</code> , the distance matrix is calculated from the <code>otu.table</code> , using the values of <code>dist.method</code> (and <code>tree</code> if required). The default is <code>NULL</code> .
square.dist	a logical variable indicating whether to square the distance matrix. The default is <code>TRUE</code> .

`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. Applies to both frequency and transformed scales. The default is TRUE.

`freq.scale.only` a logical variable indicating whether to provide adjusted frequency-scale data matrix only (not adjusted data on 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.

`x.freq` the (column-centered) frequency-scale data matrix after adjustment of covariates.

`x.tran` the (column-centered) arcsin-root-transformed data matrix after adjustment of covariates.

### Author(s)

Yi-Juan Hu <yijuan.hu@emory.edu>, Glen A. Satten <gas0@cdc.gov>

### Examples

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

#-----
# Use the adjusted distance matrix for ordination
#-----

PCs <- eigen(adj.data$adj.dist, symmetric=TRUE)

color = rep("blue", length(throat.meta$SmokingStatus))
w = which(throat.meta$SmokingStatus=="Smoker")
color[w] = "red"

plot(PCs$vectors[,1], PCs$vectors[,2], xlab="PC1", ylab="PC2",
     col=color, main="Smokers vs. non-smokers")
legend(x="topleft", legend=c("smokers", "non-smokers"), pch=c(21,21),
      col=c("red", "blue"), lty=0)
```

ldm

*Testing hypotheses using a linear decomposition model (LDM)***Description**

This function allows you to simultaneously test the global association with the overall microbiome composition and individual OTU associations to give coherent results. It is capable of handling complex design features such as confounders, interactions, and clustered data.

**Usage**

```
ldm(formula, 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 = 20, seed = NULL, test.global = TRUE,
    test.otu = TRUE, fdr.nominal = 0.1, 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 model to be fitted. The details of model specification are given under "Details".
data	an optional data frame, list or environment (or object coercible by <code>as.data.frame</code> to a data frame) containing the covariates of interest and confounding covariates. If not found in data, the covariates are taken from <code>environment(formula)</code> , typically the environment from which <code>ldm</code> is called. The default is <code>.GlobalEnv</code> .
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 <code>dist</code> .
dist.method	method for calculating the distance measure, partial match to all methods supported by <code>vegdist</code> in the <code>vegan</code> 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". Note that the option <code>binary</code> is set to <code>FALSE</code> (i.e., not allowing calculation of presence-absence distances) when we internally call <code>vegdist</code> for its supported distances. The Hellinger distance measure ( <code>dist.method="hellinger"</code> ) takes the form $0.5 \times E$ , where $E$ is the Euclidean distance between the square-root-transformed frequency data. The weighted UniFrac distance ( <code>dist.method="wt-unifrac"</code> ) is calculated by internally calling <code>GUniFrac</code> in the <code>GUniFrac</code> package. Not used when anything other than <code>dist=NULL</code> is specified for <code>dist</code> . 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 <code>square.dist=TRUE</code> and <code>center.dist=TRUE</code> are used. If <code>dist=NULL</code> , the distance matrix is calculated from the <code>otu.table</code> , using the values of <code>dist.method</code> (and <code>tree</code> if required). The default is <code>NULL</code> .
cluster.id	character or factor variable that identifies clusters. The default value <code>cluster.id=NULL</code> if the observations are not clustered (i.e., are independent).

<code>strata</code>	a character or factor variable that defines strata (groups), within which to constrain permutations. The default is NULL.
<code>how</code>	a permutation control list, for users who want to specify their own call to the <code>how</code> function from the <code>permute</code> package. The default is NULL.
<code>perm.within.type</code>	a character string that takes values "free", "none", "series", or "grid". The default is "free" (for random permutations).
<code>perm.between.type</code>	a character string that takes values "free", "none", or "series". The default is "none".
<code>perm.within.ncol</code>	a positive integer, only used if <code>perm.within.type="grid"</code> . The default is 0. See documentation for <code>permute</code> package for additional details
<code>perm.within.nrow</code>	a positive integer, only used if <code>perm.within.type="grid"</code> . The default is 0. See documentation for <code>permute</code> package for additional details
<code>n.perm.max</code>	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 <code>n.otu</code> is the number of OTUs. If a numeric value for <code>n.otu</code> is specified, this value is used for both global and OTU-level tests.
<code>n.rej.stop</code>	the minimum number of rejections (i.e., the permutation statistic exceeds the observed statistic) to obtain before stopping. The default is 20.
<code>seed</code>	an 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 permutations.
<code>test.global</code>	a logical value indicating whether to perform the global test. The default is TRUE.
<code>test.otu</code>	a logical value indicating whether to perform the OTU-specific tests. The default is TRUE.
<code>fdr.nominal</code>	the nominal FDR value. The default is 0.1.
<code>square.dist</code>	a logical variable indicating whether to square the distance matrix. The default is TRUE.
<code>center.dist</code>	a logical variable indicating whether to center the distance matrix as described by Gower (1966). The default is TRUE.
<code>scale.otu.table</code>	a logical variable indicating whether to scale the rows of the OTU table for the freq scale. For count data, this corresponds to dividing by the library size to give relative frequencies. Does not affect the tran scale. The default is TRUE.
<code>center.otu.table</code>	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. Applies to both the frequency and transformed scales. The default is TRUE.
<code>freq.scale.only</code>	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.

## 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 #### now 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.

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=20) 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=20) 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

<code>b</code>	the matrix <code>B</code> as defined in Hu and Satten (2018)
<code>dist</code>	the (squared/centered) distance matrix
<code>x.freq</code>	the frequency-scale data matrix, scaled and centered if so specified

d.freq	a vector of the nonnegative diagonal elements of D that satisfies $b^T x.freq = D v^T$
v.freq	the v matrix with unit columns that satisfies $b^T x.freq = D v^T$
x.tran	the (column-centered) arcsin-root-transformed data matrix
d.tran	a vector of the nonnegative diagonal elements of D that satisfies $b^T x.tran = D v^T$
v.tran	the v matrix with unit columns that satisfies $b^T x.tran = D v^T$
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
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	Variance explained (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

<code>p.global.freq</code>	p-values for the global test of each set of covariates based on the frequency-scale data
<code>p.global.tran</code>	p-values for the global test of each set of covariates based on the arcsin-root-transformed frequency data
<code>p.global.omni</code>	p-values for the global test of each set of covariates based on the omnibus statistics, 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
<code>p.otu.freq</code>	p-values for the OTU-specific tests based on the frequency scale data
<code>p.otu.tran</code>	p-values for the OTU-specific tests based on the arcsin-root-transformed frequency data
<code>p.otu.omni</code>	p-values for the OTU-specific tests based on the omnibus statistics
<code>q.otu.freq</code>	q-values (i.e., FDR-adjusted p-values) for the OTU-specific tests based on the frequency scale data
<code>q.otu.tran</code>	q-values for the OTU-specific tests based on the arcsin-root-transformed frequency data
<code>q.otu.omni</code>	q-values for the OTU-specific tests based on the omnibus statistics
<code>n.perm.completed</code>	number of permutations completed
<code>global.tests.stopped</code>	a logical value indicating whether the stopping criterion has been met by all global tests
<code>otu.tests.stopped</code>	a logical value indicating whether the stopping criterion has been met by all OTU-specific tests
<code>seed</code>	a single-value integer seed that is user supplied or internally generated

### Author(s)

Yi-Juan Hu <yijuan.hu@emory.edu>, Glen A. Satten <gas0@cdc.gov>

### References

Hu YJ, Satten GA. (2019) Testing hypotheses about microbiome using the linear decomposition model. [bioRxiv:doi.org/10.1101/229831](https://doi.org/10.1101/229831).

### Examples

```
#-----
# fit only
#-----
fit <- ldm(formula=throat.otu.tab | (Sex+AntibioticUse) ~ SmokingStatus+PackYears,
           data=throat.meta, dist.method="bray", n.perm.max=0)

#-----
# test the global hypothesis only
#-----
res1.ldm <- ldm(formula=throat.otu.tab | (Sex+AntibioticUse) ~ SmokingStatus+PackYears,
                data=throat.meta, dist.method="bray",
                test.global=TRUE, test.otu=FALSE, seed=123)
```



```
#-----
# test both the global hypothese and individual OTUs
#-----
res2.ldm <- ldm(formula=throat.otu.tab | (Sex+AntibioticUse) ~ SmokingStatus+PackYears,
               data=throat.meta, dist.method="bray",
               test.global=TRUE, test.otu=TRUE, fdr.nominal=0.1, seed=123)

#-----
# clustered data
#-----
res4.ldm <- ldm(formula=sim.otu.tab | X ~ Y, data=sim.meta, dist.method="bray",
               cluster.id=ID, perm.between.type="free", perm.within.type="none",
               test.global=TRUE, test.otu=TRUE, fdr.nominal=0.1, seed=123)
```

permanovaFL

*PERMANOVA test of association*

## Description

This function performs the PERMANOVA test that can allow adjustment of confounders and control of clustered data. 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.

## Usage

```
permanovaFL(formula, data = .GlobalEnv, tree = NULL,
            dist.method = "bray", 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 = 20, seed = NULL, square.dist = TRUE,
            center.dist = TRUE, scale.otu.table = TRUE)
```

## Arguments

<code>formula</code>	a symbolic description of the model to be fitted in the form of <code>data.matrix ~ sets of covariates</code> or <code>data.matrix   confounders ~ sets of covariates</code> . The details of model specification are given in "Details" of <code>ldm</code> . Additionally, in <code>permanovaFL</code> , the <code>data.matrix</code> 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, <code>tree</code> (if required), and <code>dist.method</code> . If <code>data.matrix</code> is a distance matrix (having class <code>dist</code> or <code>matrix</code> ), it can be squared and/or centered by specifying <code>square.dist</code> and <code>center.dist</code> (described below). Distance matrices are distinguished from OTU tables by checking for symmetry of <code>as.matrix(data.matrix)</code> .
<code>data</code>	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 <code>data</code> , the covariates are taken from <code>environment(formula)</code> , typically the environment from which <code>permanovaFL</code> is called. The default is <code>.GlobalEnv</code> .
<code>tree</code>	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 <code>formula</code> .

<code>dist.method</code>	method for calculating the distance measure, partial match to all methods supported by <code>vegdist</code> in the <code>vegan</code> 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. The default is "bray". For more details, see the <code>dist.method</code> argument in the <code>ldm</code> function.
<code>cluster.id</code>	cluster identifiers. The default is value of <code>NULL</code> should be used if the observations are not in clusters (i.e., independent).
<code>strata</code>	a factor variable (or, character variable converted into a factor) to define strata (groups), within which to constrain permutations. The default is <code>NULL</code> .
<code>how</code>	a permutation control list, for users who want to specify their permutation control list using the <code>how</code> function from the <code>permute</code> R package. The default is <code>NULL</code> .
<code>perm.within.type</code>	a character string that takes values "free", "none", "series", or "grid". The default is "free" (for random permutations).
<code>perm.between.type</code>	a character string that takes values "free", "none", or "series". The default is "none".
<code>perm.within.ncol</code>	a positive integer, only used if <code>perm.within.type="grid"</code> . The default is 0. See the documentation for the R package <code>permute</code> for further details.
<code>perm.within.nrow</code>	a positive integer, only used if <code>perm.within.type="grid"</code> . The default is 0. See the documentation for the R package <code>permute</code> for further details.
<code>n.perm.max</code>	the maximum number of permutations. The default is 5000.
<code>n.rej.stop</code>	the minimum number of rejections (i.e., the permutation statistic exceeds the observed statistic) to obtain before stopping. The default is 20.
<code>seed</code>	an integer seed for the random number generator in the permutation procedure. The default is <code>NULL</code> ; 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.
<code>square.dist</code>	a logical variable indicating whether to square the distance matrix. The default is <code>TRUE</code> .
<code>center.dist</code>	a logical variable indicating whether to center the distance matrix as described by Gower (1966). The default is <code>TRUE</code> .
<code>scale.otu.table</code>	a logical variable indicating whether to scale the OTU table. For count data, this corresponds to dividing by the library size to give relative frequencies. The default is <code>TRUE</code> .

## Value

a list consisting of

<code>F.statistics</code>	F statistics for the global test of each set of covariates
<code>p.permanova</code>	p-values for the global test of each set of covariates
<code>n.perm.completed</code>	number of permutations completed

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

`seed` a single-value integer seed that is user supplied or internally generated

### Author(s)

Yi-Juan Hu <yijuan.hu@emory.edu>, Glen A. Satten <gas0@cdc.gov>

### Examples

```
res.permanova <- permanovaFL(formula=throat.otu.tab | (Sex+AntibioticUse) ~ SmokingStatus+PackYears,
                             data=throat.meta, dist.method="bray", seed=123)
```

---

throat.meta	<i>Meta data of the throat microbiome samples</i>
-------------	---

---

### 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	<i>OTU count table from 16S sequencing of the throat microbiome samples</i>
----------------	---

---

**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.tree	<i>UPGMA tree of the OTUs from 16S sequencing of the throat microbiome samples</i>
-------------	--

---

**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)
```

# Index

## \*Topic **PCA**

adjust.data.by.covariates, [2](#)

## \*Topic **datasets**

throat.meta, [11](#)

throat.otu.tab, [12](#)

throat.tree, [12](#)

## \*Topic **distance**

adjust.data.by.covariates, [2](#)

## \*Topic **microbiome**

adjust.data.by.covariates, [2](#)

ldm, [4](#)

permanovaFL, [9](#)

## \*Topic **ordination**

adjust.data.by.covariates, [2](#)

adjust.data.by.covariates, [2](#)

ldm, [4](#)

permanovaFL, [9](#)

throat.meta, [11](#)

throat.otu.tab, [12](#)

throat.tree, [12](#)