Package 'miLineage'

October 13, 2022

Type Package

Title Association Tests for Microbial Lineages on a Taxonomic Tree

Version 2.1

Date 2018-03-09

Author Zheng-Zheng Tang

Maintainer Zheng-Zheng Tang <tang@biostat.wisc.edu>

Description A variety of association tests for microbiome data analysis including Quasi-Conditional Association Tests (QCAT) described in Tang Z.-Z. et al.(2017) <doi:10.1093/bioinformatics/btw804> and Zero-Inflated Generalized Dirichlet Multinomial (ZIGDM) tests described in Tang Z.-Z. & Chen G. (2017, submitted).

License GPL (>= 2)

Depends R (>= 3.1.1), MASS, data.table, geepack

NeedsCompilation no

Repository CRAN

Date/Publication 2018-03-23 00:01:09 UTC

R topics documented:

---

miLineage-package   *Association Tests for Microbial Lineages on a Taxonomic Tree*

Description

A variety of association tests for microbiome data analysis described in Tang Z.-Z. et al.(2017)
<doi: 10.1093/bioinformatics/btw804> and Tang Z.-Z. & Chen G. (2017) <Submitted>. miLineage
allows users to (a) perform tests on multivariate taxon counts; (b) localize the covariate-associated
lineages on the taxonomic tree; and (c) assess the overall association of the microbial community
with the covariate of interest.

Details

Package:	miLineage
Type:	Package
Version:	2.1
Date:	2018-03-09
License:	GPL (>=2)

QCAT, QCAT_GEE, ZIGDM

Author(s)

Zheng-Zheng Tang

Maintainer: Zheng-Zheng Tang <tang@biostat.wisc.edu>

References

1. Tang, Z.-Z., Chen, G., Alekseyenko, A. V., and Li, H. (2017). A General Framework for Associ-
ation Analysis of Microbial Communities on a Taxonomic Tree. Bioinformatics, 33(9), 1278-1285.
2. Tang, Z.-Z. and Chen, G. (2017). Zero-Inflated Generalized Dirichlet Multinomial Regression
Model for Microbiome Compositional Data Analysis. Submitted.

Examples

```
data(data.toy)
OTU.toy = data.toy$OTU.toy
Tax.toy = data.toy$Tax.toy
case = data.toy$covariate.toy
# the OTUs should be consistent between the OTU table and the taxonomy table
OTU.toy.reorder = OTU.toy[,match(rownames(Tax.toy), colnames(OTU.toy))]
# perform QCAT test for detecting differential mean
QCAT(OTU.toy.reorder, case, 1, Tax.toy, fdr.alpha=0.05)
# perform ZIGDM test for detecting differential dispersion
ZIGDM(OTU.toy.reorder, NULL, NULL, case, test.type = "Disp", 1,
  ZI.LB = 10, Tax.toy, fdr.alpha = 0.05)
```

---

data.real	*Real data that include an OTU table, a taxonomy table, and a covariate table*

Description

The data are derived from a real gut microbiome study. The data contain an OTU table, a taxonomy table, and a covariate table. The data include 96 samples and 80 OTUs.

Usage

```
data("data.real")
```

Format

data.real contains the following objects:

- OTU.real: a matrix of OTU counts for 96 samples and 80 OTUs
- Tax.real: a matrix of taxonomy table from Rank1 (kingdom level) to Rank6 (genus level)
- covariate.real: a matrix of three variables

Source

Wu, Gary D., et al. "Linking long-term dietary patterns with gut microbial enterotypes." Science 334.6052 (2011): 105-108.

Examples

```
data(data.real)
```

---

data.toy	*Simulated data that include an OTU table, a taxonomy table, and a covariate table*

Description

This data is simulated based on a mock taxonomic tree. The data contain an OTU table, a taxonomy table, and a covariate. The data include 50 samples (25 case-control pairs) and 7 OTUs.

Usage

```
data("data.toy")
```

Format

data.toy contains the following objects:

- OTU.toy: a matrix of OTU counts for 50 samples and 7 OTUs
- Tax.toy: a matrix of taxonomy table with 3 ranks
- covariate.toy: a matrix of one variable

Source

This is a simulated data based on a mock taxonomic tree.

Examples

```
data(data.toy)
```

QCAT   *Quasi-Conditional Association Test (QCAT)*

Description

This function allows users to (a) perform QCAT on multivariate taxon counts; (b) perform QCAT on the taxonomic tree to localize the covariate-associated lineages; and (c) assess the overall association of the microbial community with the covariate of interest.

Usage

```
QCAT(OTU, X, X.index, Tax = NULL,
min.depth = 0, n.resample = NULL, fdr.alpha = 0.05)
```

Arguments

OTU	a numeric matrix contains counts with each row corresponds to a sample and each column corresponds to an OTU or a taxon. Column name is mandatory. No missing values are allowed.
X	a numeric matrix contains covariates with each column pertains to one variable. Samples in the OTU and X matrices should be identical and in the same order. No missing values are allowed.
X.index	a vector indicates the columns in X for the covariate(s) of interest. The remaining columns in X will be treated as confounders.
Tax	a matrix define the taxonomic ranks with each row corresponds to an OTU or a taxon and each column corresponds to a rank (start from the higher taxonomic level, e.g., from kingdom to genus). Row name is mandatory and should be consistent with the column name of the OTU table, Column name should be formatted as "Rank1", "Rank2", etc, representing the taxonomic levels from highest to lowest.

min.depth	lower bound for sample read depth. Samples with read depth less than min.depth will be removed before the analysis.
n.resample	number of reamplings/permutations
fdr.alpha	desired false discovery rate for multiple tests on the lineages.

Value

If Tax=NULL (Default), a test is performed using all the OTUs/taxa.

If Tax is provided, tests are performed for lineages derived from the taxonomic hierarchy. The output is a list that contains 3 components

lineage.pval	p-values for all lineages. By default (n.resample=NULL), only the asymptotic test will be performed.
sig.lineage	a vector of significant lineages
global.pval	p-values of the global tests

Author(s)

Zheng-Zheng Tang

References

Tang, Z.-Z., Chen, G., Alekseyenko, A. V., and Li, H. (2017). A General Framework for Association Analysis of Microbial Communities on a Taxonomic Tree. Bioinformatics, 33(9), 1278-1285.

Examples

```
data(data.toy)
OTU.toy = data.toy$OTU.toy
Tax.toy = data.toy$Tax.toy
case = data.toy$covariate.toy
# the OTUs should be consistent between the OTU table and the taxonomy table
OTU.toy.reorder = OTU.toy[,match(rownames(Tax.toy), colnames(OTU.toy))]
# perform QCAT
QCAT(OTU.toy.reorder, case, 1, Tax.toy, fdr.alpha=0.05)
```

QCAT_GEE   *QCAT+GEE Two-Part Test*

Description

This function performs Quasi-Conditional Association Test (QCAT) for the positive taxon counts and (Generalized Estimating Equation) GEE-based association test for the zero counts. Then, the two-part test is from the combination of those two tests. This function allows users to (a) perform QCAT+GEE on multivariate taxon counts; (b) perform QCAT+GEE on the taxonomic tree to localize the covariate-associated lineages; and (c) assess the overall association of the microbial community with the covariate of interest.

Usage

```
QCAT_GEE(OTU, X, X.index, Z, Z.index, Tax = NULL,
min.depth = 0, n.resample = NULL, fdr.alpha = 0.05)
```

Arguments

OTU
: a numeric matrix contains counts with each row corresponds to a sample and each column corresponds to an OTU or a taxon. Column name is mandatory. No missing values are allowed.

X
: a numeric matrix contains covariates for the positive-part test with each column pertains to one variable. Samples in the OTU and X matrices should be identical and in the same order. No missing values are allowed.

X.index
: a vector indicates the columns in X for the covariate(s) of interest in the positive-part test. The remaining columns in X will be treated as confounders in modeling the positive abundance.

Z
: a numeric matrix contains covariates for the zero-part test with each column pertains to one variable. Samples in the OTU and Z matrices should be identical and in the same order. No missing values are allowed.

Z.index
: a vector indicates the columns in Z for the covariate(s) of interest in the zero-part test. The remaining columns in Z will be treated as confounders in modeling the proportion of zero.

Tax
: a matrix define the taxonomic ranks with each row corresponds to an OTU or a taxon and each column corresponds to a rank (start from the higher taxonomic level, e.g., from kingdom to genus). Row name is mandatory and should be consistent with the column name of the OTU table, Column name should be formatted as "Rank1", "Rank2", etc, representing the taxonomic levels from highest to lowest.

min.depth
: lower bound for sample read depth. Samples with read depth less than min.depth will be removed before the analysis.

n.resample
: number of reamplings/permutations

fdr.alpha
: desired false discovery rate for multiple tests on the lineages.

Value

If Tax=NULL (Default), tests are performed using all the OTUs/taxa.

If Tax is provided, tests are performed for lineages derived from the taxonomic hierarchy. The output is a list that contains 3 components

lineage.pval
: a list of zero-part, positive-part and two-part (combined) p-values for all lineages. By default (n.resample=NULL), only the asymptotic test will be performed.

sig.lineage
: a list of significant lineages based on the zero-part, positive-part and two-part p-values

global.pval
: p-values of the global tests

Author(s)

Zheng-Zheng Tang

References

Tang, Z.-Z., Chen, G., Alekseyenko, A. V., and Li, H. (2017). A General Framework for Association Analysis of Microbial Communities on a Taxonomic Tree. Bioinformatics, 33(9), 1278-1285.

Examples

```
data(data.toy)
OTU.toy = data.toy$OTU.toy
Tax.toy = data.toy$Tax.toy
case = data.toy$covariate.toy
# the OTUs should be consistent between the OTU table and the taxonomy table
OTU.toy.reorder = OTU.toy[,match(rownames(Tax.toy), colnames(OTU.toy))]
# perform QCAT+GEE test
QCAT_GEE(OTU.toy.reorder, case, 1, case, 1, Tax.toy, fdr.alpha=0.05)
```

ZIGDM	*Zero-Inflated Generalized Dirichlet Multinomial (ZIGDM) Tests*

Description

Different from the distribution-free QCAT and QCAT+GEE, the ZIGDM tests are based on a parametric model (ZIGDM) for multivariate taxon counts. The ZIGDM tests can not only detect differential mean but also differential dispersion level or presence-absence frequency in microbial compositions. This function allows users to (a) perform ZIGDM tests on multivariate taxon counts; (b) perform ZIGDM tests on the taxonomic tree to localize the covariate-associated lineages; and (c) assess the overall association of the microbial community with the covariate of interest.

Usage

```
ZIGDM(OTU, X4freq, X4mean, X4disp, test.type = "Mean", X.index, ZI.LB = 10, Tax = NULL,
min.depth = 0, n.resample = NULL, fdr.alpha = 0.05)
```

Arguments

OTU   a numeric matrix contains counts with each row corresponds to a sample and each column corresponds to an OTU or a taxon. Column name is mandatory. No missing values are allowed.

X4freq   a numeric matrix contains covariates that link to presence-absence frequency in microbial compositions. Each column pertains to one variable. Set X4freq=NULL if only intercept term is needed. Samples in the OTU and X4freq matrices should be identical and in the same order. No missing values are allowed.

X4mean	a numeric matrix contains covariates that link to mean abundance in microbial compositions. Each column pertains to one variable. Set X4mean=NULL if only intercept term is needed. Samples in the OTU and X4mean matrices should be identical and in the same order. No missing values are allowed.
X4disp	a numeric matrix contains covariates that link to dispersion level in microbial compositions. Each column pertains to one variable. Set X4disp=NULL if only intercept term is needed. Samples in the OTU and X4disp matrices should be identical and in the same order. No missing values are allowed.
test.type	If test.type = "Mean", the function will test for differential mean (Default). If test.type = "Disp", the function will test for differential dispersion. If test.type = "Freq", the function will test for differential presence-absence frequency.
X.index	If test.type = "Mean", X.index is a vector indicates the columns in X4mean for the covariate(s) of interest. The remaining columns in X4mean will be treated as confounders in modeling the abundance of the present taxa. If test.type = "Disp", X.index is a vector indicates the columns in X4disp for the covariate(s) of interest. The remaining columns in X4disp will be treated as confounders in modeling the dispersion level of the present taxa. If test.type = "Freq", X.index is a vector indicates the columns in X4freq for the covariate(s) of interest. The remaining columns in X4freq will be treated as confounders in modeling the presence-absence frequency.
ZI.LB	lower bound of zero counts for the taxon that needs zero-inflated model. The counts for a taxon is assumed to be zero-inflated if the number of zero observations for the taxon is greater than ZI.LB (Default is 10). If ZI.LB=NULL, GDM model (i.e. non-zero-inflated version of ZIGDM) will be applied.
Tax	a matrix define the taxonomic ranks with each row corresponds to an OTU or a taxon and each column corresponds to a rank (start from the higher taxonomic level, e.g., from kingdom to genus). Row name is mandatory and should be consistent with the column name of the OTU table, Column name should be formatted as "Rank1", "Rank2", etc, representing the taxonomic levels from highest to lowest.
min.depth	lower bound of sample read depth. Samples with read depth less than min.depth will be removed before the analysis.
n.resample	number of reamplings/permutations
fdr.alpha	desired false discovery rate for multiple tests on the lineages.

Value

If Tax=NULL (Default), a test is performed using all the OTUs/taxa.

If Tax is provided, tests are performed for lineages derived from the taxonomic hierarchy. The output is a list that contains 3 components

lineage.pval	p-values for all lineages. By default (n.resample=NULL), only the asymptotic test will be performed.
sig.lineage	a vector of significant lineages
global.pval	p-values of the global tests

Author(s)

Zheng-Zheng Tang

References

Tang, Z.-Z. and Chen, G. (2017). Zero-Inflated Generalized Dirichlet Multinomial Regression Model for Microbiome Compositional Data Analysis. Submitted.

Examples

```
data(data.toy)
OTU.toy = data.toy$OTU.toy
Tax.toy = data.toy$Tax.toy
case = data.toy$covariate.toy
# the OTUs should be consistent between the OTU table and the taxonomy table
OTU.toy.reorder = OTU.toy[,match(rownames(Tax.toy), colnames(OTU.toy))]
# perform ZIGDM test for detecting differential dispersion
ZIGDM(OTU.toy.reorder, NULL, NULL, case, test.type = "Disp", 1,
  ZI.LB = 10, Tax.toy, fdr.alpha = 0.05)
```

Index