Package 'MiHC'

February 26, 2020

Version 1.0

Title Microbiome Higher Criticism Analysis

Date 2020-02-26

Author Hyunwook Koh

Maintainer Hyunwook Koh <hkoh7@jhu.edu>

Description This R package provides facilities for the microbiome higher criticism analysis (MiHC) which tests the association between a microbial group (e.g., community or clade) composition and a host phenotype of interest. MiHC is a data-driven omnibus test taken in a search space spanned by tailoring the higher criticism test to incorporate phylogenetic information and/or modulate sparsity levels and including the Simes test for excessively high sparsity levels.

NeedsCompilation no

Depends R(>= 3.4.1), cluster, compositions, permute, phyloseq, knitr, rmarkdown

Suggests knitr, rmarkdown

VignetteBuilder knitr

License GPL-2

URL https://github.com/hk1785/MiHC

R topics documented:

MiHC   *Microbiome higher criticism analysis*

Description

This function tests the association between a microbial group (e.g., community or clade) composition and a host phenotype of interest using MiHC.

Usage

```
MiHC(y, covs=NULL, otu.tab, tree, model, hs=c(1,3,5,7,9), W=TRUE,
comp=FALSE, CLR=FALSE, opt.ncl=30, n.perm=5000)
```

Arguments

y	A numeric vector of the host outcomes. Gaussian (e.g., body mass index), Binomial (e.g., disease status, treatment/placebo) or Poisson (e.g., number of tumors/treatments) outcomes.
covs	A data.frame (or matrix/vector) for covariate (e.g., age, gender) adjustment(s). Default is cov=NULL for no covariate adjustment.
otu.tab	A matrix of the OTU table. (1. Rows are samples and columns are OTUs. 2. Monotone/singletone OTUs need to be removed.)
tree	A rooted phylogenetic tree.
model	"gaussian" for Gaussian outcomes, "binomial" for Binomial outcomes, "poisson" for Poisson outcomes.
hs	A vector of the candidate modulation schema for lower sparsity levels. Default is hc=c(1,3,5,7,9).
W	An indicator to consider weighted high criticism tests or not. Default is W=TRUE to consider weighted higher criticism tests.
comp	An indicator if the OTU table contains absolute abundances (i.e., counts) or relative abundances (i.e., proportions). Default is comp=FALSE for absolute abundances.
CLR	An indicator if the OTU table needs to be converted using the centered log-ratio (CLR) transformation. Default is CLR=FALSE for no CLR transformation.
opt.ncl	A upper limit to find the optimal number of clusters. Default is opt.ncl=30.
n.perm	A number of permutations. Default is n.perm=5000.

Value

simes.pv: The p-value for the Simes test.

ind.pvs: The p-values for the item-by-item unweighted and weighted higher criticism tests.

ada.pvs: The p-values for the local (i.e., uHC(A) and wHC(A)) and global (i.e., MiHC) omnibus higher criticism tests.

Author(s)

Hyunwook Koh

References

Koh and Zhao. A powerful microbial group association test based on the higher criticism analysis for sparse microbial association signals. (Under revision).

Simes (1986). An improved Bonferroni procedure for multiple tests of significance. Biometrika.73(3):751-754

Examples

```
# Import requisite R packages
require(cluster)
require(compositions)
require(permute)
require(phyloseq)

# Import example microbiome data
data(phy)
otu.tab <- otu_table(phy)
tree <- phy_tree(phy)
y <- sample_data(phy)$y
covs <- data.frame(matrix(NA, length(y), 2))
covs[,1] <- as.numeric(sample_data(phy)$x1)
covs[,2] <- as.factor(sample_data(phy)$x2)

# Fit MiHC
set.seed(123)
out <- MiHC(y, covs=covs, otu.tab=otu.tab, tree=tree, model="binomial", n.perm=1000)
out
```

MiHC.plot   *The Q-Q plots for the microbiome higher criticism analysis*

Description

This function produces Q-Q plots for the microbiome higher criticism analysis.

Usage

```
MiHC.plot(MiHC.out, leg.loc="bottomright", pdf.filename=NULL)
```

Arguments

MiHC.out   An output obtained using the MiHC function.

leg.loc   The legend location to list the top 10 influential OTUs. Default is leg.loc="bottomright".

pdf.filename   The PDF filename to print the figure as a PDF file. Default is pdf.filename=NULL
                to print the figure on the R graphics window.

Value

The Q-Q plots between the expected and observed quantiles for the unweighted and weighted higher criticism tests. Blue dots represent individual OTUs and a red diagonal line represents no influential points; as such, the OTUs that fall along the diagonal line have no influence on the host phenotype while the OTUs that have larger deviations from the diagonal line are more influential on the host phenotype. Darker to lighter vertical lines represent more to less influential OTUs in rank order among the 10 most influential OTUs that correspond to the 10 largest deviations from the red diagonal line.

Author(s)

Hyunwook Koh

References

Koh and Zhao. A powerful microbial group association test based on the higher criticism analysis for sparse microbial association signals. (Under revision).

Simes (1986). An improved Bonferroni procedure for multiple tests of significance. Biometrika.73(3):751-754

Examples

```
# Import requisite R packages
require(cluster)
require(compositions)
require(permute)
require(phyloseq)

# Import example microbiome data
data(phy)
otu.tab <- otu_table(phy)
tree <- phy_tree(phy)
y <- sample_data(phy)$y
covs <- data.frame(matrix(NA, length(y), 2))
covs[,1] <- as.numeric(sample_data(phy)$x1)
covs[,2] <- as.factor(sample_data(phy)$x2)

# Fit MiHC
set.seed(123)
out <- MiHC(y, covs=covs, otu.tab=otu.tab, tree=tree, model="binomial", n.perm=1000)
out

# Create plots
MiHC.plot(MiHC.out=out)
```

MiHC.stat	*uHC(h)'s and wHC(h)'s*

Description

This function produces the values of the test statistics of uHC(h)'s and wHC(h)'s.

Usage

```
MiHC.stat(Zs, hs=hs, Ws=Ws)
```

Arguments

Zs	The Z-score values for individual OTUs.
hs	A vector of the candidate modulation schema for lower sparsity levels. Default is hc=c(1,3,5,7,9).
Ws	A vector of the weights for invidual OTUs.

Value

The values of the test statistics of uHC(h)'s and wHC(h)'s.

Author(s)

Hyunwook Koh

References

Koh and Zhao. A powerful microbial group association test based on the higher criticism analysis for sparse microbial association signals. (Under revision).

Examples

```
# Import requisite R packages
require(cluster)
require(permute)
require(phyloseq)

Zs <- rnorm(50)
hs <- c(1,3,5,7,9)
Ws <- runif(50)

MiHC.stat(Zs, hs=hs, Ws=Ws)
```

---

phy	*An example microbiome and meta data*

Description

This is an example data in the phyloseq format.

Usage

```
data("phy")
```

Format

phyloseq

Author(s)

Hyunwook Koh

References

McMurdie and Holmes (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE. 8(4):e61217

Examples

```
require(phyloseq)

data(phy)
phy
```