Package 'massMap'

May 19, 2020

Title massMap: a two-stage microbial association mapping framework with advanced FDR control

Version 2.0

Date 2020-05-19

Author Jiyuan Hu, Hyunwook Koh, Linchen He, Menghan Liu, Martin J. Blaser, Huilin Li.

Maintainer Jiyuan Hu <Jiyuan.Hu@nyumc.org>

Description This package is developed to powerfully discover trait-associated taxa at the taget rank, such as Genus or species. A two-stage microbial association mapping framework (massMap) is implemented which uses grouping information from the taxonomic tree to strengthen statistical power in association tests at the target rank. MassMap can apply to binary, continuous and survival traits. MassMap first screens the association of taxonomic groups at a pre-selected higher taxonomic rank using a powerful microbial group test OMiAT for the binary/continuous trait or optimal microbiome-based survival analysis tool OMiSA for the survival trait respectively. Then it proceeds to test the association for each candidate taxon at the target rank within the significant taxonomic groups identified in the first stage. Hierarchical BH (HBH) and selected subset testing (SST) procedures are evaluated to control the FDR for the two-stage structured tests. Both simulation studies and real data applications demonstrated marked statistical power improvement of massMap over competing methods.

Depends dirmult, phyloseq, robustbase, robCompositions, BiasedUrn, CompQuadForm, MiRKAT, GUniFrac, ecodist, survival, OMiSA, R (>= 3.3.0)

License GPL (>= 2)

Encoding UTF-8

LazyData true

URL https://github.com/JiyuanHu/massMap

BugReports http://github.com/JiyuanHu/massMap/issues

R topics documented:

massMap-package *Massmap: a two-stage microbial association mapping framework with advanced FDR control*

Description

This package is developed to powerfully discover trait-associated taxa at the taget rank, such as Genus or species. A two-stage microbial association mapping framework (massMap) is implemented which uses grouping information from the taxonomic tree to strengthen statistical power in association tests at the target rank. MassMap can apply to binary, continuous and survival traits. MassMap first screens the association of taxonomic groups at a pre-selected higher taxonomic rank using a powerful microbial group test OMiAT for the binary/continuous trait or optimal microbiome-based survival analysis tool OMiSA for the survival trait respectively. Then it proceeds to test the association for each candidate taxon at the target rank within the significant taxonomic groups identified in the first stage. Hierarchical BH (HBH) and selected subset testing (SST) procedures are evaluated to control the FDR for the two-stage structured tests. Both simulation studies and real data applications demonstrated marked statistical power improvement of massMap over competing methods.

Details

Package:	massMap
Type:	Package
Version:	2.0
Date:	2020-05-19
License:	GPL (>= 2)

Author(s)

Jiyuan Hu, Hyunwook Koh, Linchen He, Menghan Liu, Martin J. Blaser, Huilin Li.

Maintainer: Jiyuan Hu <Jiyuan.Hu@nyumc.org>

References

Hu, Jiyuan, Hyunwook Koh, Linchen He, Menghan Liu, Martin J. Blaser, and Huilin Li. "A two-stage microbial association mapping framework with advanced FDR control." Microbiome 6, no. 1 (2018): 131.

Examples

```
require(massMap)
###Illustration 1: binary trait
##MassMap for the continuous trait is the same as the binary trait.
data(phy)
map = as.data.frame(sample_data(phy))
X = map[,c(age,gender)] # a data frame, each element could either numeric or factor.
Y = map[[ABH]]#a vector instead of a data frame.
otu.tab = otu_table(phy)
```

```
tax.tab = tax_table(phy)
tree = phy_tree(phy)
############
# Not run:
#res = massMap(X=X, Y=Y,otu.tab=otu.tab,is.count.otu.tab=TRUE,
#   tax.tab=tax.tab,tree=tree,outcome.trait="binary",
#   screening.rank = "Family", target.rank="Species",alpha=0.05,n.perm=1e2)
#res$res.screening ##The group association test at the screening rank
#res$res.target ##The microbial association test at the target rank
###

###Illustration 2: survival trait
##The observed survival time (obstime) and the event indicator (delta) are required
##when running massMap for the surval trait
data("MiSurv.Data",package= OMiSA)
otu.tab <- otu_table(MiSurv.Data)
tax.tab <- tax_table(MiSurv.Data)
tree <- phy_tree(MiSurv.Data)
map = sample_data(MiSurv.Data)
obstime <- as.numeric(unlist(map[,1]))
delta <- as.numeric(unlist(map[,2]))
X = data.frame(map[,3:4])
############
# Not run:
#res.Surv = massMap(X=X, obstime = obstime,delta= delta,otu.tab=otu.tab, is.count.otu.tab=TRUE,
# tax.tab=tax.tab, tree=tree,outcome.trait="survival",
#   screening.rank = "Family",target.rank="Species",alpha=0.05,n.perm=1e4)

#res$res.screening ##The group association test at the screening rank
#res$res.target ##The microbial association test at the target rank
```

`massMap` *The main function of the two-stage microbial association mapping framework.*

Description

This function detects the trait-associated taxa at the specified taget rank. We recommend Family to be the screening rank. The results of Hierarchical Benjamini-Hochberge (HBH), selected subset testing with BH procedures (SST) and the traditional one-stage BH procedure are reported.

Usage

```
##binary or continuous trait
massMap(X, Y, otu.tab, is.count.otu.tab = TRUE, tax.tab, tree,
outcome.trait = outcome.traits, screening.rank = "Family", target.rank = ranks,
alpha = 0.05, n.perm = 1e4)
##survival trait
massMap(X, obstime, delta, otu.tab, is.count.otu.tab = TRUE, tax.tab, tree,
   outcome.trait = outcome.traits, screening.rank= "Family", target.rank= ranks,
   alpha=0.05,n.perm=1e4)
```

```
outcome.traits
#c("binary", "continuous", "survival")
ranks
#c("Kingdom", "Phylum", "Class", "Order", "Family", "Genus", "Species", "OTU")
```

Arguments

X	A data frame which contains covariates to be adjusted in the regression model. The elements of the data frame must be numeric or factor. Set X = NULL if there is no covariate.
Y	A numeric vector of the binary/continuous outcome trait with length = sample size.
obstime	A numeric vector of the survival time for the survival outcome trait with length = sample size.
delta	A numeric vector of the status indicator for the survival outcome trait with length = sample size.
otu.tab	The taxonomic table of the microbiome data. Each row represents the taxonomy alignment of each taxon on Kingdom (Domain), Phylum, Class, Order, Family, Genus and Species respectively. Best if it is a class otu_table from package "phyloseq".
is.count.otu.tab	An indicator of whether the OTU table contains the count data or relative abundance data for OTUs.
tax.tab	The OTU table of the microbiome data. Each row represents a subject and each column represents the OTU. Best if it is a class taxonomyTable from package "phyloseq".
tree	The phylogenetic tree of the microbiome data. Best if it is a class phylogenetic tree from package "phyloseq".
outcome.trait	Specify the type of outcome trait. Must be either "binary", "continuous" or "survival".
screening.rank	Specify the screening rank. The recommend and default setting is "Family".
target.rank	Specify the target rank.The default setting is "Species".
alpha	Significance level for the adjusted p-values. The default setting is 0.05.
n.perm	Numver of permutations in order to calculate p-values. The default setting 1e4 is large enough to obtain accurate p-values.

Value

res.screening	A data frame which contains the group association test results at the screening rank. There are four elements: **lineage** The lineage of each taxonomic group; **size** The number of taxa within the lineage; **pval.raw** The raw p-value of OMiAT test; **pval.adj** The adjusted p-value of OMiAT test.
res.target	A data frame which contains the association test results at the target rank. There are eight elements: **lineage** The lineage of each taxon at the target rank; **p.raw** The raw p-value of the non-parametric score association test;

p.BH The BH adjusted p-values;

p.HBH Adjusted p-values using Hierarchical BH procedure;

p.SST Adjusted p-values using selected subset testing procedure;

status.BH The association status of each taxon after adjustment of p-values using BH procedure. status.BH=1 if the taxon is significantly associated with the outcome trait. status.BH=0 otherwise;

status.HBH The association status of each taxon after adjustment of p-values using two-stage HBH procedure;

status.SST The association status of each taxon after adjustment of p-values using two-stage SST procedure.
The p-value of HBH and SST is NA if the corresponding group test at the screening rank is insignificant.

Author(s)

Jiyuan Hu, Hyunwook Koh, Linchen He, Menghan Liu, Martin J. Blaser, Huilin Li.

References

Hu, Jiyuan, Hyunwook Koh, Linchen He, Menghan Liu, Martin J. Blaser, and Huilin Li. "A two-stage microbial association mapping framework with advanced FDR control." Microbiome 6, no. 1 (2018): 131.

Examples

```
require(massMap)
###Illustration 1: binary trait
##MassMap for the continuous trait is the same as the binary trait.
data(phy)
map = as.data.frame(sample_data(phy))
X = map[,c(age,gender)] # a data frame, each element could either numeric or factor.
Y = map[[ABH]]#a vector instead of a data frame.
otu.tab = otu_table(phy)
tax.tab = tax_table(phy)
tree = phy_tree(phy)
############
# Not run:
#res = massMap(X=X, Y=Y,otu.tab=otu.tab,is.count.otu.tab=TRUE,
#  tax.tab=tax.tab,tree=tree,outcome.trait="binary",
#  screening.rank = "Family", target.rank="Species",alpha=0.05,n.perm=1e2)
#res$res.screening ##The group association test at the screening rank
#res$res.target ##The microbial association test at the target rank
###

###Illustration 2: survival trait
##The observed survival time (obstime) and the event indicator (delta) are required
##when running massMap for the surval trait
data("MiSurv.Data",package= OMiSA)
otu.tab <- otu_table(MiSurv.Data)
tax.tab <- tax_table(MiSurv.Data)
tree <- phy_tree(MiSurv.Data)
map = sample_data(MiSurv.Data)
obstime <- as.numeric(unlist(map[,1]))
delta <- as.numeric(unlist(map[,2]))
X = data.frame(map[,3:4])
```

```
############
# Not run:
#res.Surv = massMap(X=X, obstime = obstime,delta= delta,otu.tab=otu.tab, is.count.otu.tab=TRUE,
# tax.tab=tax.tab, tree=tree,outcome.trait="survival",
#   screening.rank = "Family",target.rank="Species",alpha=0.05,n.perm=1e4)

#res$res.screening ##The group association test at the screening rank
#res$res.target ##The microbial association test at the target rank
```

---

phy *A phyloseq-class experiment-level object example data.*

Description

This data is a phyloseq-class object which contains the OTU table, sample data, taxonomy table and phylogenetic tree information. Details about the phyloseq class please see the R package "phyloseq". We are interested in the association between microbial taxa and the binary outcome trait antibiotic history (ABH) adjusting for age and gender.

Usage

```
data(phy)
```

Format

phyloseq-class experiment-level object

`otu_table()` OTU Table: [90 taxa and 100 samples]

`..@ .Data`: a matrix where each row represents the suject,each column represents the OTU.

`..@ taxa_are_rows`: logi FALSE

`sample_data()` Sample Data: [100 samples by 3 sample variables]

A data frame with 100 observations for one outcome variable ABH and two covariates age and gender.

`ABH`: a numeric vector indicating the antibiotic history of each individual, where ABH =0 indicating the individual without antibiotic usage and ABH=1 indicating the individual having antibiotic usage in the preceding year;

`age`: a numeric vector;

`gender`: a factor with levels `female`, `male`.

`tax_table()` Taxonomy Table: [90 taxa by 7 taxonomic ranks]

`..@ .Data`: a matrix where each row represents the taxonomy alignment on Kingdom (Domain), Phylum, Class, Order, Family, Genus and Species represently for each taxon.

`phy_tree()` Phylogenetic Tree: [90 tips and 89 internal nodes]

The phylogenetic tree of the example microbiome data. Look for details in the help file of phy_tree() from "phyloseq" package.

Index