

# Package ‘massMap’

February 22, 2018

**Title** A two-stage microbial association mapping framework with advanced FDR control

**Version** 1.0

**Date** 2018-02-22

**Author** Jiyuan Hu

**Maintainer** Jiyuan Hu <jiyuan\_hu@fudan.edu.cn>

**Description** This package is developed to powerfully discover trait-associated taxa at the target 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 first screens the association of taxonomic groups at a pre-selected higher taxonomic rank using a powerful microbial group test OMiAT. 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, 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 . . . . . | 2 |
| ABH . . . . .             | 3 |
| massMap . . . . .         | 3 |
| otu.tab . . . . .         | 5 |
| tax.tab . . . . .         | 5 |
| tree . . . . .            | 6 |
| X . . . . .               | 6 |
| <b>Index</b>              | 7 |

---

|                 |  |
|-----------------|--|
| massMap-package | <i>A two-stage microbial association mapping framework with advanced FDR control</i> |
|-----------------|--|

---

## Description

This package is developed to powerfully discover trait-associated taxa at the target 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 first screens the association of taxonomic groups at a pre-selected higher taxonomic rank using a powerful microbial group test OMiAT. 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: 1.0
Date: 2018-02-22
License: GPL (>= 2)
```

## Author(s)

Jiyuan Hu

Maintainer: Jiyuan Hu <jiyuan\_hu@fudan.edu.cn>

## References

HU J, Koh H, He L, Liu M, Blaser J M, Li H(2018). A two-stage microbial association mapping framework with advanced FDR control.

## Examples

```
data(X)#covariates of age and gender
data(ABH)#the antibiotic history of 100 subjects
data(otu.tab)# the otu table of microbiome data for 100 subjects and 90 species.
data(tax.tab)#the taxonomy alignment of 90 species
data(tree) #the phylogenetic tree of the microbiome data
#####
# Not run:
#res = massMap(X=X, Y=ABH,otu.tab=otu.tab,tax.tab=tax.tab,tree=tree,outcome.trait="binary",
# screening.rank = "Family", target.rank="Species",alpha=0.2,n.perm=1e4)
```

ABH

*The binary outcome trait antibiotic history (ABH) for each subject.***Description**

A numeric vector which contains the antibiotic history of each subject. ABH = 0 if the subject did not receive antibiotic treatment in the preceding year. ABH= 1 otherwise.

**Usage**

```
data("ABH")
```

**Format**

A numeric vector.

**References**

HU J, Koh H, He L, Liu M, Blaser J M, Li H(2018). A two-stage microbial association mapping framework with advanced FDR control.

massMap

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

This function detects the trait-associated taxa at the specified target 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**

```
massMap(X, Y, otu.tab, tax.tab, tree, outcome.trait = outcome.traits,
screening.rank = ranks, target.rank = ranks,
alpha = 0.05, n.perm = 10000)
outcome.traits
#c("binary", "continuous")
ranks
#c("Domain", "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 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 <code>otu_table</code> from package "phyloseq". |
| 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 <code>taxonomyTable</code> from package "phyloseq".   |
| tree           | The phylogenetic tree of the microbiome data. Best if it is a class <code>phylogenetic tree</code> from package "phyloseq".   |
| outcome.trait  | Specify the type of outcome trait. Must be either "binary" or "continuous".   |
| 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 | <p>A data frame which contains the group association test results at the screening rank. There are four elements:</p> <p><b>lineage</b> The lineage of each taxonomic group;</p> <p><b>size</b> The number of taxa within the lineage;</p> <p><b>pval.raw</b> The raw p-value of OMiAT test;</p> <p><b>pval.adj</b> The adjusted p-value of OMiAT test.</p>  |
| res.target    | <p>A data frame which contains the association test results at the target rank. There are eight elements:</p> <p><b>lineage</b> The lineage of each taxon at the target rank;</p> <p><b>p.raw</b> The raw p-value of the non-parametric score association test;</p> <p><b>p.BH</b> The BH adjusted p-values;</p> <p><b>p.HBH</b> Adjusted p-values using Hierarchical BH procedure;</p> <p><b>p.SST</b> Adjusted p-values using selected subset testing procedure;</p> <p><b>status.BH</b> The association status of each taxon after adjustment of p-values using BH procedure. <code>status.BH=1</code> if the taxon is significantly associated with the outcome trait. <code>status.BH=0</code> otherwise;</p> <p><b>status.HBH</b> The association status of each taxon after adjustment of p-values using two-stage HBH procedure;</p> <p><b>status.SST</b> The association status of each taxon after adjustment of p-values using two-stage SST procedure.</p> <p>The p-value of HBH and SST is NA if the corresponding group test at the screening rank is insignificant.</p> |

### Author(s)

Jiyuan Hu

### References

HU J, Koh H, He L, Liu M, Blaser J M, Li H(2018). A two-stage microbial association mapping framework with advanced FDR control.

## Examples

```
data(X)#covariates of age and gender
data(ABH)#the antibiotic history of 100 subjects
data(otu.tab)# the otu table of microbiome data for 100 subjects and 90 species.
data(tax.tab)#the taxonomy alignment of 90 species
data(tree) #the phylogenetic tree of the microbiome data
#####
# Not run:
#res = massMap(X=X, Y=ABH,otu.tab=otu.tab,tax.tab=tax.tab,tree=tree,outcome.trait="binary",
# screening.rank = "Family", target.rank="Species",alpha=0.2,n.perm=1e4)
```

---

otu.tab

*The OTU table of the example microbiome data.*

---

## Description

A class 'otu\_table' from the R package "phyloseq". Each row is the microbial abundance profile for each individual. Each column is the abundance for each OTU.

## Usage

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

## Format

A Formal class 'otu\_table' [package "phyloseq"] with 2 slots

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

..@ taxa\_are\_rows: logi FALSE

## References

HU J, Koh H, He L, Liu M, Blaser J M, Li H(2018). A two-stage microbial association mapping framework with advanced FDR control.

---

tax.tab

*The taxonomic table of the example microbiome data.*

---

## Description

A class 'taxonomyTable' from the R package "phyloseq".

## Usage

```
data("tax.tab")
```

## Format

A Formal class 'taxonomyTable' [package "phyloseq"] with 1 slot

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

---

`tree`*The phylogenetic tree of the example microbiome data.*

---

**Description**

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

**Usage**

```
data("tree")
```

---

`X`*Covariates of the associated microbiome data.*

---

**Description**

A data frame which contains covariates to be adjusted in the regression models.

**Usage**

```
data("X")
```

**Format**

A data frame with 100 observations for two covariates.

`age` a numeric vector;

`gender` a factor with levels female, male.

**References**

HU J, Koh H, He L, Liu M, Blaser J M, Li H(2018). A two-stage microbial association mapping framework with advanced FDR control.

# Index

## \*Topic **datasets**

- ABH, [3](#)
- otu.tab, [5](#)
- tax.tab, [5](#)
- tree, [6](#)
- X, [6](#)

## \*Topic **package**

- massMap-package, [2](#)

## \*Topic **two-stage**

- massMap, [3](#)

ABH, [3](#)

massMap, [3](#)

massMap-package, [2](#)

otu.tab, [5](#)

tax.tab, [5](#)

tree, [6](#)

X, [6](#)