Package 'miMediation'

February 27, 2026

Version 1.0

Date 2026-02-27

Title Mediation Tests for Microbime Data

Description Testing mediation effects of microbiome.

Author Qilin Hong, Yunfei Peng

Maintainer Qilin Hong <qhong8@wisc.edu>, Yunfei Peng <peng228@wisc.edu>

Depends R (>= 3.5.0)

Encoding UTF-8

License GPL-3

LazyData true

biocViews

Imports ape,
data.table,
data.tree,
fdrtool,
ggplot2,
ggtree,
harmonicmeanp,
MASS,
phyloseq,
SKAT,
TreeTools,
PALM,
compositions,
hdi,
cp4p,
Iso,
gtools,
glmnet,
HDMT,
stats

Suggests knitr,
prettydoc,
rmarkdown

RoxygenNote 7.3.1

VignetteBuilder knitr,
prettydoc

R topics documented:

`CAMRA` *CAMRA: Causal Absolute-abundance Mediation from Relative-Abundance data*

Description

CAMRA is a method for taxon-level microbiome mediator discovery that delivers well-calibrated error control. Unlike existing approaches that test mediation effects directly on the relative-abundance (RA) scale and can suffer from compositionality-induced false discoveries, CAMRA targets absolute-abundance (AA)–scale mediation effects by reconstructing AA-level exposure–microbiome and microbiome–outcome associations from RA inputs. CAMRA then combines the two path-specific signals under a composite-null mediation framework and reports taxon-wise q-values for mediator discovery.

Usage

```
CAMRA(
  mediators,
  treatment,
  outcome,
  confounders = NULL,
  pseudo = 0.5,
  fdr.alpha = 0.05,
  hdmt.exact = 0,
  screen = FALSE,
  const = 2,
  CClasso = FALSE,
  seed = 42
)
```

Arguments

`mediators`	A named numeric matrix containing microbiome abundance. Each row is a subject and each column is a taxon. Column name contains the taxon name.
`outcome`	A numeric vector of the outcome variable (currently treated as a continuous variable).
`confounders`	An optional numeric vector or matrix containing confounders that may affect the treatment, mediators and outcome. Each row is a subject and each column is a specific confounder, e.g., age or sex. Default is NULL.
`pseudo`	A numeric value for the pseudo-count added before log-ratio transformations. Default is 0.5.
`fdr.alpha`	An optional numeric value for the desired FDR significance level in identifying mediating nodes on the tree. Default is 0.05.

hdmt.exact	An integer flag passed to HDMT::fdr_est controlling how mixture-null quantiles are computed: 0 = approximation method, 1 = exact method. Default is 0.
screen	A logical value indicating whether to screen high-dimensional mediators before testing. Default is FALSE.
const	A numeric multiplier used for the number of filtered bacteria during screening (const * n/log(n)). Default is 2.
CClasso	A logical value indicating whether to use fastCCLasso to compute correlations instead of SparCC. Default is FALSE.
seed	An integer seed for reproducibility. Default is 42.
exposure	A numeric vector of the exposure.

Details

CAMRA implements taxon-level mediation testing that targets absolute-abundance (AA)–scale mediation effects using standard sequencing relative-abundance (RA) count inputs. For each taxon k, CAMRA first estimates the exposure-to-microbiome association by applying PALM to obtain a path-specific p-value p_{α_k}. It then estimates the microbiome-to-outcome association conditional on exposure using a PALAR-style regression with debiased inference to obtain p_{β_k}.

These two path-specific signals are combined using the max-P joint-significance statistic $\max\{p_{\alpha_k}, p_{\beta_k}\}$ for testing the composite mediation null $H_{0k} : \alpha_k\beta_k = 0$. Finally, CAMRA applies HDMT to obtain taxon-level q-values by explicitly accounting for the mixture structure of the composite null, enabling well-calibrated multiple-testing adjustment for mediator discovery.

Value

A list containing the following components:

pval.alpha	A numeric vector of p-values from the exposure-microbiome association test.
pval.beta	A numeric vector of p-values from the microbiome-outcome association test (conditional on exposure).
qval.med	A numeric vector of taxon-level mediation test q-values.
sig.mediators	A vector of selected microbial mediators significant at the ‘fdr.alpha‘ level.
index_detected	The column indices of the significant mediators in the original matrix.
runtime_sec	The total execution time in seconds.

Author(s)

Qiyu Wang <wang_qy@mail.ustc.edu.cn>, Yunfei Peng <peng228@wisc.edu>

References

Wang Q, Li Y, Peng Y, Tang, ZZ (2026). Error control in microbiome mediator discovery: benchmark and remedy. Submitted.

Wei Z, Hong Q, Chen G, Hartert TV, Rosas-Salazar C, Das SR, Shilts MH, Levin AM, Tang ZZ (2026). Fast and reliable association discovery in large-scale microbiome studies and meta-analyses using PALM. Genome Biology, accepted.

Li Y, Wang Q, Feng Z, Wang X, Tang ZZ (2025). PALAR: Estimation of absolute abundance effects in regression with relative abundance predictors. Journal of the American Statistical Association (JASA), DOI: 10.1080/01621459.2025.2596250.

Dai JY, Stanford JL, LeBlanc M (2020). A multiple-testing procedure for high-dimensional mediation hypotheses. Journal of the American Statistical Association (JASA), DOI: 10.1080/01621459.2020.1765785.

Examples

```
# Load the real dataset
data(data.bmi)

# Run CAMRA mediation screening
CAMRA_res <- CAMRA(
  mediators = data.camra$mediators,
  treatment = data.camra$treatment,
  outcome = data.camra$outcome,
  confounders = NULL,
  fdr.alpha = 0.05,
  seed = 123
)

# View significant microbial mediators
print(CAMRA_res$sig.mediators)
```

---

data.bmi *Real data for CAMRA example*

Description

This data is derived from 'curatedMetagenomicData' R package, which includes 863 samples from China and USA.

Usage

```
data(data.bmi)
```

Format

data.bmi contains the following components:

treatment treatment indicator.

mediators an abundance matrix of microbial taxa.

outcome continuous outcome variable.

---

data.cecal *Real data*

Description

This data is derived from a real microbiome dataset (Cho et al., 2012), which includes 48 samples (38 in antibiotics vs. 10 in controls).

Usage

```
data(data.cecal)
```

Format

`data.cecal` contains the following components:

treatment treatment indicator: Antibiotics group and control group are coded as 1 and 0, respectively.

mediators an abundance matrix with the top 100 most abundant taxa that have at least 20% non-zero observations.

outcome body fat percentage.

tree a phylogenetic tree.

Source

https://doi.org/10.1038/nature11400

References

Cho, I. et al. (2012). Antibiotics in early life alter the murine colonic microbiome and adiposity. *Nature* 488:621-626.

Examples

```
data(data.cecal)
```

`data.zeeviD` *Real data*

Description

This data is derived from a real microbiome dataset (Zeevi, D. et al., 2015), which includes microbiome samples from 200 healthy subjects.

Usage

```
data(data.zeeviD)
```

Format

`data.zeeviD` contains the following components:

treatment treatment indicator.

mediators an abundance matrix with the top 100 most abundant taxa.

outcome continuous outcome.

tree a taxonomy table.

Source

https://doi.org/10.1016/j.cell.2015.11.001

References

Zeevi, D. et al. (2015). Personalized nutrition by prediction of glycemic responses. *Cell* 163:1079-1094.

Examples

```
data(data.zeeviD)
```

---

phyloMed *Phylogeny-based test of mediation effect in microbiome (PhyloMed)*

Description

phyloMed enables us to test the mediation effect in high-dimensional microbial composition. The method leverages the hierarchical phylogeny relationship among different microbial taxa to decompose the complex mediation model on the full microbial composition into multiple simple independent local mediation models on subcompositions. The phyloMed function (a) performs the mediation test for the subcomposition at each internal node of the phylogenetic tree and pinpoint the mediating nodes with significant test p-values; and (b) combine all subcomposition p-values to assess the overall mediation effect of the entire microbial community.

Usage

```
phyloMed(
  treatment,
  mediators,
  outcome,
  confounders = NULL,
  interaction = FALSE,
  tree,
  pi.method = "product",
  fdr.alpha = 0.05,
  n.perm = NULL,
  verbose = FALSE,
  graph = NULL
)
```

Arguments

treatment	A numeric vector of the treatment.
mediators	A named numeric matrix containing microbiome abundance. Each row is a subject and each column is a taxon. Column name contains the taxon name.
outcome	A numeric vector of continuous or binary outcome.
confounders	An optional numeric vector or matrix containing confounders that may affect the treatment, mediators and outcome. Each row is a subject and each column is a specific confounder, e.g., age or sex. Default is NULL.
interaction	An optional logical value. If TRUE, the interaction term between treatment and mediator will be taken into account. Default is FALSE.
tree	A phylogenetic tree (phylo-class object) or a taxonomy table (matrix-class object). The tip labels in the phylogenetic tree or the row names in the taxonomy table should overlap with the column names in the mediators matrix. The column names in the taxonomy table should start from the higher level to lower level, e.g., from kingdom to genus. See Details.

pi.method	An optional character string denotes the method to used in estimate proportion of null. Default method is "product", an alternative method is "maxp". Can be abbreviated.
fdr.alpha	An optional numeric value for the desired FDR significance level in identifying mediating nodes on the tree. Default is 0.05.
n.perm	An optional numeric value for the maximum number of permutations. Default is NULL. See Details.
verbose	An optional logical value. If TRUE, information of the test on each node will be printed. Default is FALSE.
graph	An optional character string denotes the layout of the graph, which contains a phylogenetic tree or taxonomic tree with identified mediating nodes highlighted. Can be "circular" or "rectangular". Can be abbreviated. Default is NULL. See Details.

Details

phyloMed could leverage phylogeny or taxonomy relationship among taxa. If the tree is a unrooted and/or non-binary phylogenetic tree, phyloMed will preprocess the tree: (a) root the tree with the longest tip branch as outgroup if it is unrooted; and/or (b) resolve multichotomies into dichotomies based on the order they appear if it is non-binary. If the tree is a taxonomy table, phyloMed will group taxa based on different levels of taxonomic ranks. phyloMed uses the treatment-mediator association test p-value and mediation-outcome association test p-value to construct the subcomposition mediation test statistic at each local model (Hong et al., Manuscript). The two p-values can come from either the asymptotic test or the permutation test. Asymptotic test is faster but less accurate when the study sample size is small. By default (n.perm=NULL), only asymptotic test will be performed. Otherwise, if n.perm is set to a positive number, results from two versions of PhyloMed will be output, one based on the asymptotic p-value and the other based on the permutation p-value. Graph only highlights the mediating nodes identified from permutation version when both versions are performed. By default (graph=NULL), graph will not be plotted.

Value

A phyloseq-class object named clean.data and a list named rslt.

clean.data contains the following components:

sample_data	Input treatment, outcome and confounders.
otu_table	The abundance data for the taxa that are present on the tips of the phy_tree or on the rows of the tax_table.
tax_table	The taxonomy table with rows exactly match the taxa in the otu_table. NULL if input tree is a phylogeny tree.
phy_tree	The binary and rooted phylogenetic tree with tips exactly match the taxa in the otu_table: (a) The internal nodes are numbered with value larger than the number of tips; (b) The internal nodes are numbered sequentially, with values increasing away from the root. NULL if input tree is a taxonomy table.

If n.perm is not NULL, the function will return two lists in rslt named PhyloMed.A and PhyloMed.P, respectively. Otherwise, only one list named PhyloMed.A will be returned.

Each list contains the following components:

node.pval	A numeric vector of subcomposition mediation p-values for all internal nodes.
sig.clade	A list of significant nodes with their descendants.

`null.prop` A vector of the estimated proportion of different types of null hypotheses across all local mediation tests.

`global.pval` A global test p-value using harmonic mean.

If `graph` is not `NULL`, the phylogenetic or taxonomic tree will be plotted. The layout depends on the input of `graph`. The size of the circle at each internal node is proportional to $-\log_{10}$(subcomposition p-value), the larger circle indicates a smaller p-value. The significant nodes are highlighted by blue rectangle.

Author(s)

Qilin Hong <qhong8@wisc.edu>

References

Hong, Q., Chen, G., & Tang, Z. Z. (2023). PhyloMed: a phylogeny-based test of mediation effect in microbiome. *Genome Biology*, 24(1), 1-21.

Examples

```
# Load real data
data(data.cecal)
# Run test with phylogeny tree
Trt = data.cecal$treatment
M = data.cecal$mediators
Y = data.cecal$outcome
tree = data.cecal$tree
rslt.phylomed = phyloMed(Trt, M, Y, tree = tree, graph = "rectangular")
# Run test with taxonomy table
Trt = data.zeeviD$treatment
M = data.zeeviD$mediators
Y = data.zeeviD$outcome
tree = data.zeeviD$tree
rslt.phylomed = phyloMed(Trt, M, Y, tree = tree, graph = "circular")
```