Package 'metamicrobiomeR'

March 2, 2019

Title an R package for analysis of microbiome relative abundance data using zero inflated beta GAMLSS and meta-analysis across studies using random effect model

Version 1.1

YEAR 2019

Description The metamicrobiomeR package implements Generalized Additive Model for Location, Scale and Shape (GAMLSS)
with zero inflated beta (BEZI) family for analysis of microbiome relative abundance data (with various options for data transformation/normalization to address compositional effects) and random effect meta-analysis models for meta-
analysis pooling estimates across microbiome studies.
Random Forest model to predict microbiome age based on relative abundances of shared bacterial genera with the Bangladesh data (Subramanian et al 2014), comparison of multiple diversity indexes using linear/linear mixed effect models and some data display/visualization are also implemented.

Depends R (>= 3.4.2)

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1.9000

Imports gdata, plyr, dplyr, ggplot2, gridExtra, RColorBrewer, lme4, lmerTest, mgcv, meta, gamlss, reshape2, caret, randomForest, grDevices, gplots, magrittr, tools, sas7bdat, foreign, knitr, rmarkdown, zCompositions, compositions, matrixStats, RCurl, httr, repmis, jsonlite

Suggests knitr, rmarkdown

VignetteBuilder knitr

URL https://github.com/nhanhocu/metamicrobiomeR

BugReports https://github.com/nhanhocu/metamicrobiomeR/issues

RemoteType github

RemoteHost api.github.com

RemoteRepo metamicrobiomeR

RemoteUsername nhanhocu

RemoteRef master

RemoteSha 3ab29318c414691554545f1d070b0a7eee2d8963

GithubRepo metamicrobiomeR

GithubUsername nhanhocu

GithubRef master

GithubSHA1 3ab29318c414691554545f1d070b0a7eee2d8963

NeedsCompilation no

Author Nhan Ho [aut, cre]

Maintainer Nhan Ho <nhanhocumc@gmail.com>

R topics documented:

`alpha.compare` *Compare multiple alpha diversity indexes between groups*

Description

This function calculates average of alpha diversity indexes for a specific rarefaction depth, standardize diversity indexes and compare between groups using linear/linear mixed effect model and adjust for covariates.

Usage

```
alpha.compare(datlist, depth, mapfile, mapsampleid, comvar, adjustvar,
  personid = "personid", longitudinal = "yes", age.limit = 1e+06,
  standardize = FALSE, ...)
```

Arguments

`datlist`	the list of your dataframe.
`depth`	the rarefaction depth of choice. Depth can be "max" (highest depth) or an order (number) of the depth in the list generated by alpha_rarefaction.py
`mapfile`	mapping file
`mapsampleid`	sample id in the mapping file

comvar	variable for comparison
adjustvar	variables that need to be adjusted in the model
personid	name of variable for person id. Default is "personid" (applicable when longitudinal="yes").
longitudinal	longitudinal data or one time data. Options are c("yes","no"). Default is "yes".
age.limit	age upper limit for included samples. Default is 1000000 months (~no upper limit).
standardize	standardization of diversity indexes before comparison or not. Default is FALSE.

Value

list of comparison result matrices and mean diversity of all indexes for each samples (with or without standardization)

Examples

```
#Read in outputs from "alpha_rarefaction.py" in QIIME
patht<-system.file("extdata/QIIME_outputs/Bangladesh/alpha_div_collated", package = "metamicrobiomeR", mustW
alpha.rm<-read.multi(patht=patht,patternt=".txt",assignt="no",study="Bangladesh")
#Bangladesh extra metadata
data(sam.rm)
samfile<-merge(samde, he50[,c("child.id","gender","month.exbf","month.food")],by="child.id")
samfile$age.sample<-samfile$age.months
samfile$bf<-factor(samfile$bf,levels=c("ExclusiveBF","Non_exclusiveBF","No_BF"))
samfile$personid<-samfile$child.id
samfile$sampleid<-tolower(samfile$fecal.sample.id)
#comparison of standardized alpha diversity indexes between genders adjusting for breastfeeding and infant age
alphacom6.rm.sexsg<-alpha.compare(datlist=alpha.rm,depth=3,mapfile=samfile,mapsampleid="fecal.sample.id",co
alphacom6.rm.sexsg$alphasum[,1:5]
```

`meta.niceplot` *Nice meta-analysis plots.*

Description

This function displays meta-analysis results of relative abundance as a nice combined heatmap and forest plot. More flexibility/options for plot will be added.

Usage

```
meta.niceplot(metadat, sumtype = "taxa", level = "main", p, p.adjust,
  phyla.col = c("rainbow", "select"),
  phyla.select = c("actinobacteria", "bacteroidetes", "cyanobacteria",
  "firmicutes", "fusobacteria", "proteobacteria", "verrucomicrobia",
  ".thermi."), col.select = c("#dd1c77", "#31a354", "#91003f", "#d95f0e",
  "#636363", "#2ef0e7", "#862ef0", "#000"), est.break = c(-Inf, -1, -0.5,
  -0.1, 0, 0.1, 0.5, 1, Inf), est.break.label = c("<-1", "[-1,-0.5)",
  "[-0.5,-0.1)", "[-0.1,0)", "[0,0.1)", "[0.1,0.5)", "[0.5,1)", ">=1"),
  neg.palette = "PuBu", pos.palette = "YlOrRd", p.sig.heat = "no",
  p.break = c(0, 1e-04, 0.05, 1), p.break.label = c("**", "*", ""),
  p.pool.break = c(0, 0.05, 1), p.pool.break.label = c("[0,0.05)",
```

```
"[0.05,1]"), padjust.pool.break = c(0, 0.1, 1),
padjust.pool.break.label = c("[0,0.1)", "[0.1,1]"),
forest.est.shape = c("17", "16"), forest.est.col = c("red", "black"),
forest.col = c("by.pvalue", "by.estimate"), leg.key.size = 1,
leg.text.size = 8, heat.text.x.size = 8, heat.text.x.angle = 0,
forest.axis.text.y = 8, forest.axis.text.x = 8,
heat.forest.width.ratio = c(1, 1), point.ratio = c(3, 1),
line.ratio = c(2, 1))
```

Arguments

metadat	output data from metatab.show.
sumtype	Either "taxa" for taxa and "path" for pathway.
level	"main" for main level such as phylum or "sub" for higher level such as species. Default is "main".
p	name of variable for p-values
p.adjust	name of variable for multiple testing adjusted p-values
phyla.col	type of color for main level (phylum). Options are "rainbow" (default) or "select".
phyla.select	selected phyla for selected colors (only when phyla.col="select"). Default are c("actinobacteria","bacteroidetes","cyanobacteria","firmicutes","fusobacteria","proteobacteria","verr
col.select	selected colors for selected phyla (only when phyla.col="select"). Corresponding default are c("#dd1c77","#31a354","#91003f","#d95f0e","#636363","#2ef0e7","#862ef0","#000'
est.break	breaks for estimates to generate color categories on heatmap. Default are c(-Inf, -1,-0.5,-0.1,0,0.1,0.5,1, Inf). For pathway, recommended breaks are c(-Inf, -0.5,-0.1,-0.05,0,0.05,0.1,0.5, Inf).
est.break.label	labels for corresponding color categories on heatmap generated by est.break. Default corresponding to default est.break are c("<-1","[-1,-0.5)","[-0.5,-0.1)","[-0.1,0)", "[0,0.1)", "[01,0.5)", "[0.5,1)", ">=1"). For pathway, corresponding recommended labels are c("<-0.5)", "[-0.5,-0.1)","[-0.1,-0.05)","[-0.05,0)","[0,0.05)","[0.05,0.1)", "[0.1,0.5)", ">=0.5").
neg.palette	color palette for negative estimate values. Default is "PuBu". Use display.brewer.all() of RColorBrewer package for other options.
pos.palette	color palette for positive estimate values. Default is "YlOrRd". Use display.brewer.all() of RColorBrewer package for other options.
p.sig.heat	whether or not show significant p values on heatmap. Default is "yes".
p.break	breaks for significant levels of p values. Default is c(0, 0.0001,0.05, 1).
p.break.label	labels to be showed on heatmap for different levels of p-values from p.break. Default is c("**", "*","") for p breaks at c(0, 0.0001,0.05, 1).
p.pool.break	breaks for pooled p-values to be distinguished in forest plot. Default are c(0,0.05,1).
p.pool.break.label	labels for pooled p-value breaks. Corresponding default are c("[0,0.05)","[0.05,1]").
padjust.pool.break	breaks for multiple testing adjusted p-values to be distinguished in forest plot. Default are c(0,0.1,1).
padjust.pool.break.label	labels for multiple testing adjusted p-value breaks. Corresponding default are c("[0,0.1)","[0.1,1]").

`forest.est.shape`
: point shape of pooled estimates in forest plot. Default are c("17","16") for corresponding significant and non-significant pooled estimates.

`forest.est.col`
: colors of point (pooled estimates) and 95 CI bars in forest plot. Default are c("red", "black") for significant and non-significant estimates.

`forest.col`
: Color of forest plot (point estimates and 95 CI). Options are "by.pvalue" (distinguished by signficant vs. non-significant p-value) or "by.estimate" (color scaled similarly to heatmap color).

`leg.key.size`
: legdend key size for heatmap.

`leg.text.size`
: legend text size for heatmap.

`heat.text.x.size`
: heatmap x label text size.

`heat.text.x.angle`
: heatmap x label text angle.

`forest.axis.text.y`
: forest plot y label text size.

`forest.axis.text.x`
: forest plot x label text size.

`heat.forest.width.ratio`
: ratio of width between heatmap and forest plot to be used in grid.arrange. Dedault is c(1,1).

`point.ratio`
: ratio of point size between significant pooled estimate and non-significant pooled estimate. Default is c(3,1).

`line.ratio`
: ratio of error bar line size between significant pooled estimate and non-significant pooled estimate. Default is=c(2,1).

Value

combined heatmap forest plot.

Examples

```
#Load saved results of four studies for the comparison of bacterial taxa relative abundance between genders adju
data(taxacom.rm.sex.adjustbfage)
data(taxacom.ha.sex.adjustbfage)
data(taxacom6.zi.usbmk.sex.adjustbfage)
data(taxacom6.unc.sex.adjustedbfage)
taxacom6.zi.rm.sex.adjustbfage$study<-"Subramanian et al 2014 (Bangladesh)"
taxacom6.zi.rm.sex.adjustbfage$pop<-"Bangladesh"
taxacom.zi.ha.sex.adjustbfage$study<-"Bender et al 2016 (Haiti)"
taxacom.zi.ha.sex.adjustbfage$pop<-"Haiti"
taxacom6.zi.usbmk.sex.adjustbfage$study<-"Pannaraj et al 2017 (USA(CA_FL))"
taxacom6.zi.usbmk.sex.adjustbfage$pop<-"USA(CA_FL)"
taxacom6.zi.unc.sex.adjustedbfage$study<-"Thompson et al 2015 (USA(NC))"
taxacom6.zi.unc.sex.adjustedbfage$pop<-"USA(NC)"
tabsex4<-plyr::rbind.fill(taxacom6.zi.rm.sex.adjustbfage,taxacom.zi.ha.sex.adjustbfage,taxacom6.zi.usbmk.se
#Meta-analysis (take time to run)
metab.sex<-meta.taxa(taxcomdat=tabsex4,summary.measure="RR",pool.var="id",studylab="study",backtransform=FA
#nice plot phylum level
metadat<-metatab.show(metatab=metab.sex$random,com.pooled.tab=tabsex4,tax.lev="l2",showvar="genderMale",p.c
meta.niceplot(metadat=metadat,sumtype="taxa",level="main",p="p",p.adjust="p.adjust",phyla.col="rainbow",leg
#nice plot family level
```

```
metadat<-metatab.show(metatab=metab.sex$random,com.pooled.tab=tabsex4,tax.lev="l5",showvar="genderMale",p.c
meta.niceplot(metadat=metadat,sumtype="taxa",level="sub",p="p",p.adjust="p.adjust",phyla.col="rainbow",leg
```

`meta.taxa` *Meta-analysis of taxa/pathway abundance comparison.*

Description

This function does meta-analysis based on estimates and standard errors from taxa/pathway abundance comparison using random effect and fixed effect meta-analysis models.

Usage

```
meta.taxa(taxcomdat, estimate.pattern = "Estimate.",
  se.pattern = "Std. Error.", summary.measure = "RR",
  pool.var = "id", studylab = "study", backtransform = FALSE,
  percent.meta = 0.5, p.adjust.method = "fdr", ...)
```

Arguments

`taxcomdat`	matrice of estimates and SE of all taxa/pathways combined from all included studies.
`estimate.pattern`	string pattern for estimates. Default is "Estimate.".
`se.pattern`	string pattern for standard error. Default is "Std. Error.".
`summary.measure`	"RR" for estimates from GAMLSS with BEZI family and "RD" for estimates from Linear/linear mixed effect model. Default is "RR"
`pool.var`	name of id variable for meta-analysis. Default is "id".
`studylab`	name of variable characterizing included studies. Default is "study".
`backtransform`	whether or not to perform backtransformation of the estimates. Default is FALSE.
`percent.meta`	the threshold percentage of number of studies that a taxa is available to do meta-analysis. Default is 0.5
`p.adjust.method`	method for multiple testing adjustment (available methods of the function p.adjust). Default is "fdr".

Value

a list of matrices of results for all variables in the comparison models.

Examples

```
#Load saved results of four studies for the comparison of bacterial taxa relative abundance between genders adju
data(taxacom.rm.sex.adjustbfage)
data(taxacom.ha.sex.adjustbfage)
data(taxacom6.zi.usbmk.sex.adjustbfage.)
data(taxacom6.unc.sex.adjustedbfage)
taxacom6.zi.rm.sex.adjustbfage$study<-"Subramanian et al 2014 (Bangladesh)"
taxacom6.zi.rm.sex.adjustbfage$pop<-"Bangladesh"
taxacom.zi.ha.sex.adjustbfage$study<-"Bender et al 2016 (Haiti)"
```

```
taxacom.zi.ha.sex.adjustbfage$pop<-"Haiti"
taxacom6.zi.usbmk.sex.adjustbfage$study<-"Pannaraj et al 2017 (USA(CA_FL))"
taxacom6.zi.usbmk.sex.adjustbfage$pop<-"USA(CA_FL)"
taxacom6.zi.unc.sex.adjustedbfage$study<-"Thompson et al 2015 (USA(NC))"
taxacom6.zi.unc.sex.adjustedbfage$pop<-"USA(NC)"
tabsex4<-plyr::rbind.fill(taxacom6.zi.rm.sex.adjustbfage,taxacom.zi.ha.sex.adjustbfage,taxacom6.zi.usbmk.se
#Meta-analysis (take time to run)
metab.sex<-meta.taxa(taxcomdat=tabsex4,summary.measure="RR",pool.var="id",studylab="study",backtransform=FA
metatab.show(metatab=metab.sex$random,com.pooled.tab=tabsex4,tax.lev="l2",showvar="genderMale",p.cutoff.typ
```

`metatab.show` *Display meta-analysis results.*

Description

This function displays meta-analysis results of relative abundance as heatmap, forest plot, table or data.

Usage

```
metatab.show(metatab, com.pooled.tab, sumvar = "taxa", tax.lev = "l2",
  showvar, estimate.pattern = "Estimate.", se.pattern = "Std. Error.",
  p.pattern = "Pr(>|t|)", readjust.p = FALSE, p.cutoff.type = "p",
  p.cutoff = 0.05, display = "plot", plot = "heatmap",
  fill.value = "log(OR)", grid = FALSE, digit = 2, p.digit = 4,
  ...)
```

Arguments

`metatab`	matrice of taxa/pathway abundance comparison meta-analysis results generated from meta.taxa.
`com.pooled.tab`	matrice of taxa/pathway abundance comparison generated from taxa.compare or pathway.compare combined from all included studies.
`sumvar`	Either "taxa" for taxa and "path" for pathway.
`tax.lev`	taxa level to be displayed. Options are from "l2" (phylum) to "l7" (species). Default is "l2".
`showvar`	variable (string pattern) in the model to be displayed.
`estimate.pattern`	string pattern for estimates. Default is "Estimate.".
`se.pattern`	string pattern for standard error. Default is "Std. Error.".
`readjust.p`	multiple testing re-adjustment for only the level to be displayed (TRUE) or keep original multiple testing adjustment for all taxa of all levels (FALSE).Default is FALSE.
`p.cutoff.type`	type of p-value for cutoff. Options are "p" for p-value or "p.adjust" for multiple testing adjusted p-value. Default is "p".
`p.cutoff`	cutoff p-value to be displayed. Default is 0.05.
`display`	type of display. Options are display=c("plot","table","data")
`plot`	type of plot. Options are plot=c("heatmap","forest").

fill.value	name of legend.
grid	whether multiple plots will be displayed alongside. Default is FALSE.
digit	digit for estimates and 95 CI. Default is 2.
p.digit	digit for p-values. Default is 4.

Value

plot table or data.

Examples

```
#Load saved results of four studies for the comparison of bacterial taxa relative abundance between genders adju
data(taxacom.rm.sex.adjustbfage)
data(taxacom.ha.sex.adjustbfage)
data(taxacom6.zi.usbmk.sex.adjustbfage.)
data(taxacom6.unc.sex.adjustedbfage)
taxacom6.zi.rm.sex.adjustbfage$study<-"Subramanian et al 2014 (Bangladesh)"
taxacom6.zi.rm.sex.adjustbfage$pop<-"Bangladesh"
taxacom.zi.ha.sex.adjustbfage$study<-"Bender et al 2016 (Haiti)"
taxacom.zi.ha.sex.adjustbfage$pop<-"Haiti"
taxacom6.zi.usbmk.sex.adjustbfage$study<-"Pannaraj et al 2017 (USA(CA_FL))"
taxacom6.zi.usbmk.sex.adjustbfage$pop<-"USA(CA_FL)"
taxacom6.zi.unc.sex.adjustedbfage$study<-"Thompson et al 2015 (USA(NC))"
taxacom6.zi.unc.sex.adjustedbfage$pop<-"USA(NC)"
tabsex4<-plyr::rbind.fill(taxacom6.zi.rm.sex.adjustbfage,taxacom.zi.ha.sex.adjustbfage,taxacom6.zi.usbmk.se
#Meta-analysis (take time to run)
metab.sex<-meta.taxa(taxcomdat=tabsex4,summary.measure="RR",pool.var="id",studylab="study",backtransform=FA
metatab.show(metatab=metab.sex$random,com.pooled.tab=tabsex4,tax.lev="l2",showvar="genderMale",p.cutoff.typ
metatab.show(metatab=metab.sex$random,com.pooled.tab=tabsex4,tax.lev="l2",showvar="genderMale",p.cutoff.typ
```

---

microbiomeage *Predict microbiome age.*

Description

This function predicts microbiome age using Random Forest model based on relative abundances of bacterial genera shared with the Bangladesh study (Subramanian et al 2014). This function gets the shared genera list between the Bangladesh study and all other included studies, get the training and test sets from Bangladesh data based on the shared genera list, fit the train Random Forest model and predict microbiome age in the test set of Bangladesh data and data from all included studies, check for performance of the model based on the shared genera list on Bangladesh healthy cohort data, reproduce the findings of the Bangladesh malnutrition study.

Usage

```
microbiomeage(l6.relabundtab)
```

Arguments

l6.relabundtab	list of taxa summary table from phylum up to genus level merged to mapping file outputed from QIIME of all included studies.

Value

list of training and test sets of Bangladesh data, shared genera list, relative abundance data of shared genera, randomforest fit, RF model performance plot,predicted microbiome age of Bangladesh data and data of other included studies.

Examples

```
#Load data from each study and put in a list
#load Bangladesh taxa relative abundance summary up to genus level merged with mapping file (output from QIIME)
bal6<-read.delim(system.file("extdata/QIIME_outputs/Bangladesh/tax_mapping7", "Subramanian_et_al_mapping_fi
colnames(bal6)<-tolower(colnames(bal6))
#View(bal6)
#format for data of other studies should be similar to Bangladesh data, must have 'age.sample' variable as age of
# Load data of 3 other studies
data(gtab.3stud)
names(gtab.3stud)
#predict microbiome age on Bangladesh data and data of other three studies based on shared genera across 4 studie
#(take time to run)
miage<-microbiomeage(l6.relabundtab=gtab.3stud)
#list of shared genera that are available in the Bangladesh study and other included studies
miage$sharedgenera.importance
#check performance
gridExtra::grid.arrange(miage$performanceplot$ptrain, miage$performanceplot$ptest,nrow=1)
#replicate the findings of Subramanian et al paper
ggplot2::ggplot() +geom_point(data=miage$microbiomeage.bangladesh$all,aes(x=age.sample, y=age.predicted, col
```

`pathway.compare` *Compare (kegg) pathway abundance*

Description

This is a slightly modified version of the taxa.compare function. It compares pathway abundance generated from PICRUSt analysis between groups using different methods (apply to pathway summary tables already merged to mapping file and put in a list (e.g.level 1, 2 and 3)). Specifically, it compares relative abundances of bacterial functional pathways at all levels using GAMLSS or LM/LMEM and compares of log(absolute abundances) of bacterial functional pathways at all levels using LM/LMEM.

Usage

```
pathway.compare(pathtab, mapfile, sampleid = "sampleid",
  pathsum = "rel", stat.med = "gamlss", transform = "none", comvar,
  adjustvar, personid = "personid", longitudinal = "yes",
  p.adjust.method = "fdr", percent.filter = 0.05,
  relabund.filter = 5e-05, pooldata = FALSE, age.limit = 1e+06,
  age.lowerlimit = 0, ...)
```

Arguments

`pathtab`	list of pathway abundance table of all levels.
`mapfile`	mapping file or file containing covariates.

sampleid	variable containing sample id to be matched between pathway abundance table and mapping file.
pathsum	type of abundance to be compared. Options are "rel" for relative abundance or "log" for log of absolute abundance. Default is "rel".
stat.med	statistical method for comparison. stat.med can be "lm" for LM/LMEM (usable for both pathsum="rel" or "log") or "gamlss" for GAMLSS with BEZI family (gamlss only make sense if pathsum="rel").
comvar	main variable for comparison.
adjustvar	variables to be adjusted.
personid	name of variable for person id in mapping file (applicable for longitudinal data)
longitudinal	whether the data is longitudinal. Default is "yes".
p.adjust.method	method for multiple testing adjustment. Available options are those of the p.adjust function. Default is "fdr".
percent.filter	prevalence threshold (the percentage of number of samples the taxa/pathway available). Default is 0.05.
relabund.filter	relative abundance threshold (the minimum of the average relative abundance for a taxa/pathway to be retained). Default is 0.00005.
pooldata	whether the data is pooled from multiple studies. Default is FALSE.
age.limit	upper age limit for data to be included. Default is 1000000 months (~no upper age limit).
age.lowerlimit	lower age limit for data to be included. Default is 0 month.

Value

matrice of coefficients, standard errors, p-values and multiple testing adjusted p-values of all variables in the models.

Examples

```
#Load Bangladesh extra metadata
data(sam.rm)
#Read in PICRUSt output for KEGG pathways level 1-3
patht<-system.file("extdata/QIIME_outputs/Bangladesh/picrust", package = "metamicrobiomeR", mustWork = TRUE)
kegg<-read.multi(patht=patht,patternt=".txt",assignt="no")
kegg.rm<-list()
for (i in 1:length(kegg)){
  rownames(kegg[[i]])<-kegg[[i]][,"kegg_pathways"]
 kegg[[i]]<-kegg[[i]][,colnames(kegg[[i]])[!colnames(kegg[[i]]) %in% c("otu.id","kegg_pathways")]]
  kegg.rm[[i]]<-as.data.frame(t(kegg[[i]]))
}
covar.rm<-merge(samde, he50[,c("child.id","gender","zygosity","day.firstsample","day.lastsample","n.sample"
                    "n.diarrhea.yr","percent.time.diarrhea","fraction.antibiotic","subject.allocation")],
covar.rm<-dplyr::rename(covar.rm,sampleid=fecal.sample.id, personid=child.id ,age.sample=age.months)
covar.rm$bf<-factor(covar.rm$bf, levels=c('ExclusiveBF','Non_exclusiveBF','No_BF'))
covar.rm$personid<-as.factor(covar.rm$personid)
#Comparison of pathway relative abundances (take time to run)
pathcom.rm6.rel.gamlss.sexg<-pathway.compare(pathtab=kegg.rm,mapfile=covar.rm,sampleid="sampleid",pathsum="
taxcomtab.show(taxcomtab=pathcom.rm6.rel.gamlss.sexg$l3, sumvar="path",tax.lev="l3",tax.select="none",showv
```

read.multi *Read multiple files*

Description

This function reads multiple files of the same pattern in a directory into R.

Usage

```
read.multi(patht, patternt = ".txt", assignt = "no", study = NULL,
  tolower.colnames = TRUE)
```

Arguments

patht	path to files.
patternt	pattern of files. Options are ".txt" or ".csv" or ".sav" or ".sas7bdat".
assignt	assign files. Default is "no".
study	name of the study. Default is NULL.
tolower.colnames	turn all column names to lower case. Default is TRUE.

Value

list of all data files in the path

Examples

```
patht<-system.file("extdata/QIIME_outputs/Bangladesh/alpha_div_collated", package = "metamicrobiomeR", mustW
alpha.ba<-read.multi(patht=patht,patternt=".txt",assignt="no",study="Bangladesh")
```

taxa.compare *Compare taxa relative abundance*

Description

This function compares taxa relative abundance summary tables at all levels between groups using GAMLSS with BEZI or Linear/Linear Mixed Effect models (LM/LMEM) after filtering (using prevalence and relative abundance thresholds).

Usage

```
taxa.compare(taxtab, propmed.rel = "gamlss", transform = "none",
  zeroreplace.method = "none", comvar, adjustvar,
  personid = "personid", longitudinal = "yes", percent.filter = 0.05,
  relabund.filter = 5e-05, p.adjust.method = "fdr", ...)
```

Arguments

taxtab	taxa relative abundance table (already merged to mapping file) from phylum to species or any preferred highest taxa level.
propmed.rel	statistical method for comparing relative abundance. Options are "lm" for LM/LMEM or "gamlss" for GAMLSS with BEZI family.
transform	transformation of relative abundance data. Options are "none" for no transformation, "gmpr" for Geometric Mean of Pairwise Ratios (GMPR) normalization, "asin.sqrt" for arcsine transformation, "logit" for logit transformation, "clr" for centered log ratio transformation. Default is "none".
zeroreplace.method	Method for zero replacement implemented in R package *zCompositions*. Options are "none" for no replacement, "multKM" for Multiplicative Kaplan-Meier smoothing spline (KMSS) replacement, "multLN" for Multiplicative lognormal replacement, "multRepl" for Multiplicative simple replacement, "lrEM" for Log-ratio EM algorithm, "lrDA" for Log-ratio DA algorithm. Default is "none".
comvar	main variable for comparison
adjustvar	variables to be adjusted.
personid	name of variable for person id (applicable for longitudinal data)
longitudinal	whether data is longitudinal? Options are "yes" or "no". Default is "yes".
percent.filter	prevalence threshold (the percentage of number of samples the taxa/pathway available). Default is 0.05.
relabund.filter	relative abundance threshold (the minimum of the average relative abundance for a taxa/pathway to be retained). Default is 0.00005.
p.adjust.method	method for multiple testing adjustment. Options are those of the p.adjust.methods of stats:: p.adjust function. Default for this function is "fdr".
pooldata	whether data is pooled from multiple studies. Default is FALSE.

Value

matrice of coefficients, standard errors, p-values and multiple testing adjusted p-values of all variables in the models.

Examples

```
#Load summary tables of bacterial taxa relative abundance from Bangladesh data
data(taxtab.rm7)
#Comparison of bacterial taxa relative abundance using LMEM or GAMLSS (take time to run)
taxacom6.rmg<-taxa.compare(taxtab=taxtab6.rm[[5]],propmed.rel="lm",comvar="bf",adjustvar="age.sample",longi
taxacom6.zi.rmg<-taxa.compare(taxtab=taxtab6.rm[[5]],propmed.rel="gamlss",comvar="bf",adjustvar="age.sample
taxcomtab.show(taxcomtab=taxacom6.zi.rmg,tax.select="none", showvar="bfNon_exclusiveBF", tax.lev="l2",readj
```

taxa.filter *Filter relative abundance data*

Description

This function filters bacterial taxa or pathway relative abundance tables based on the percentage of samples with their availability (prevalence) and relative abundance thresholds. It will remove taxa/pathway with relative abundance <relabund.filter and available in <percent.filter of number of samples.

Usage

```
taxa.filter(taxtab, percent.filter = 0.05, relabund.filter = 5e-05)
```

Arguments

taxtab	taxa/pathway relative abundance table.
percent.filter	prevalence threshold (the percentage of number of samples the taxa/pathway available). Default is 0.05.
relabund.filter	relative abundance threshold (the minimum of the average relative abundance for a taxa/pathway to be retained). Default is 0.00005.

Value

list of all taxa/pathways retained after filtering.

Examples

```
#Load summary tables of bacterial taxa relative abundance from Bangladesh data
data(taxtab.rm7)
taxlist.rm<-taxa.filter(taxtab=taxtab.rm[[5]],percent.filter = 0.05, relabund.filter = 0.00005)
```

taxa.mean.plot *Plot mean taxa abundance*

Description

This function visualize mean relative abundance by group as stacked plots.

Usage

```
taxa.mean.plot(tabmean, sumvar = "taxa", tax.select = "none",
  tax.lev = "l2", comvar, groupvar, mean.filter = 0.005,
  pallete.by.phylum = FALSE, show.taxname = "full",
  legend.position = "right", xlab = "Chronological age (month)",
  ylab = "Relative abundance")
```

Arguments

tabmean	table of mean abundance generated from taxa.meansdn.
sumvar	variable to be plotted. Options are c("taxa","path"). Default is "taxa"
tax.select	list of selected taxa/pathways to be plotted. Default is "none" or plot all taxa/pathways.
tax.lev	taxa level to be visualized. Options are from "l2" (phylum) to "l7" (species). Default is "l2". If sumvar="path", all pathways will be visualized.
comvar	main variable for comparison.
groupvar	variable for stratifying.
mean.filter	mean abundance filtering threshold (only plot those with mean abundance>threshold and plot all those with mean abundance <threshold as "other").
pallete.by.phylum	whether each pallete of color for each phylum. Default is FALSE (plot distinc colors).
show.taxname	whether show "full" taxa name or "short" name. Default is "full".
legend.position	position of legend. Options are c("right", "left","bottom","top","none") as in ggplot2. Default is "right".

Value

a list of ggplot2 object and list of taxa/pathways plotted (those with mean abundance >mean.filter).

Examples

```
#Load summary tables of bacterial taxa relative abundance from Bangladesh data
data(taxtab.rm7)
library(magrittr)
taxa.meansdn.rm<-taxa.meansdn(taxtab=taxtab.rm[[5]],sumvar="bf",groupvar="age.sample")
p.bf.l2<-taxa.mean.plot(tabmean=taxa.meansdn.rm,tax.lev="l2", comvar="bf", groupvar="age.sample",mean.filte
p.bf.l2$p
```

---

taxa.meansdn *Summarize abundance by group*

Description

This function summarizes taxa/pathway abundance tables to provide mean, sd, count by groups.

Usage

```
taxa.meansdn(taxtab, sumvar, groupvar, percent.filter = 0.05,
  relabund.filter = 5e-05, othervar = "none")
```

Arguments

taxtab	taxa/pathway abundance table from phylum to species or any preferred highest taxa level.
sumvar	main variable for summary
groupvar	variable to be stratified.
percent.filter	prevalence threshold (the percentage of number of samples the taxa/pathway available). Default is 0.05.
relabund.filter	relative abundance threshold (the minimum of the average relative abundance for a taxa/pathway to be retained). Default is 0.00005.
othervar	list of variables that are not abundance variables to be summarized. Default is "none".

Value

table of mean, sd, count by group.

Examples

```
#Load summary tables of bacterial taxa relative abundance from Bangladesh data
data(taxtab.rm7)
library(magrittr)
taxa.meansdn.rm<-taxa.meansdn(taxtab=taxtab.rm[[5]],sumvar="bf",groupvar="age.sample")
```

---

taxcomtab.show *Display abundance comparison results.*

Description

This function displays taxa/pathway abundance comparison results as table.

Usage

```
taxcomtab.show(taxcomtab, sumvar = "taxa", tax.lev = "l2",
  tax.select = "none", showvar, readjust.p = FALSE,
  p.adjust.method = "fdr", p.cutoff = 0.05, digit = 2, p.digit = 4,
  ...)
```

Arguments

taxcomtab	table of taxa abundance comparison generated from taxa.compare.
sumvar	Options are "taxa" for bacterial taxa and "path" for pathway. Default is "taxa"
tax.lev	taxa level to be displayed. Options are from "l2" (phylum) to "l7" (species). Default is "l2".
tax.select	selected list of taxa to be displayed. Default is "none" or display all available taxa.
showvar	variable (pattern) in the model to be displayed.
readjust.p	multiple testing re-adjustment for only the level to be displayed (TRUE) or keep original multiple testing adjustment for all taxa of all levels (FALSE).

`p.adjust.method`
: method for multiple testing adjustment. Available options are those of the p.adjust function. Default is "fdr".

`p.cutoff`
: cutoff p-value to be displayed. Default is 0.05.

`digit`
: digit for estimates and 95 CI. Default is 2.

`p.digit`
: digit for p-values. Default is 4.

Value

a table of results.

Examples

```
#Load summary tables of bacterial taxa relative abundance from Bangladesh data
data(taxtab.rm7)
#Comparison of bacterial taxa relative abundance using LMEM or GAMLSS (take time to run)
taxacom6.rmg<-taxa.compare(taxtab=taxtab6.rm[[5]],taxsum="rel",propmed.rel="lm",comvar="bf",adjustvar="age
taxcomtab.show(taxcomtab=taxacom6.rmg,tax.select="none", showvar="bfNon_exclusiveBF", tax.lev="l2",readjust
```

Index