

Package ‘MCMC.OTU’

February 4, 2019

Type Package

Title Bayesian Analysis of Multivariate Counts Data in DNA
Metabarcoding and Ecology

Version 1.0.10

Date 2016-02-10

Author Mikhail V. Matz

Maintainer Mikhail V. Matz <matz@utexas.edu>

Description

Poisson-lognormal generalized linear mixed model analysis of multivariate counts data using MCMC, aiming to infer the changes in relative proportions of individual variables. The package was originally designed for sequence-based analysis of microbial communities (“metabarcoding”, variables = operational taxonomic units, OTUs), but can be used for other types of multivariate counts, such as in ecological applications (variables = species). The results are summarized and plotted using ‘ggplot2’ functions. Includes functions to remove sample and variable outliers and reformat counts into normalized log-transformed values for correlation and principal component/coordinate analysis. Walkthrough and examples: http://www.bio.utexas.edu/research/matz_lab/matzlab/Methods_files/walkthroughExample_mcmcOTU_R.txt.

License GPL-3

Depends MCMCglmm,ggplot2,coda

NeedsCompilation no

Repository CRAN

Date/Publication 2016-02-12 00:53:04

R topics documented:

| | |
|--------------------------------|----|
| MCMC.OTU-package | 2 |
| getNormalizedOTUdata | 3 |
| green.data | 4 |
| logLin | 9 |
| mcmc.otu | 10 |
| mcmc.pval | 13 |
| otuByAutocorr | 14 |

| | |
|--------------------------|----|
| otuStack | 14 |
| OTUsummary | 15 |
| padjustOTU | 17 |
| panel.cor | 18 |
| panel.cor.pval | 18 |
| purgeOutliers | 19 |
| signifOTU | 20 |
| startedLog | 20 |
| summaryPlotOTU | 21 |

Index 23

| | |
|------------------|--|
| MCMC.OTU-package | <i>Bayesian analysis of multivariate counts data in DNA metabarcoding and ecology.</i> |
|------------------|--|

Description

This package enables MCMC-based generalized linear mixed model analysis of multivariate counts data, such as common in DNA metabarcoding and community ecology. The results are summarized and plotted using ggplot2 functions.

Details

Package: MCMC.OTU
 Type: Package
 Version: 1.0.10
 Date: 2016-02-10
 License: GPL-3

At the moment, the package handles experimental design with a single multilevel fixed factor or two fully crossed multilevel fixed factors. Any number of scalar (OTU-specific) random factors (i.e. blocking factors) are allowed. By default, it is assumed that variation in total counts per sample is not biologically relevant (reflects sequencing or survey effort). See help for the core function `mcmc.otu()` for more details.

Author(s)

Mikhail V. Matz <matz@utexas.edu>

References

Elizabeth A. Green, Sarah W. Davies, Mikhail V. Matz, Monica Medina Quantifying cryptic Symbiodinium diversity within *Orbicella faveolata* and *Orbicella franksi* at the Flower Garden Banks, Gulf of Mexico. PeerJ 2014 2:e386 <https://peerj.com/articles/386/>

Examples

```
# Symbiodinium sp diversity in two coral species at two reefs (banks)
data(green.data)

# removing outliers
goods=purgeOutliers(
  data=green.data,
  count.columns=c(4:length(green.data[1,])),
  zero.cut=0.25 # remove this line for real analysis
)

# stacking the data table
gs=otuStack(
  data=goods,
  count.columns=c(4:length(goods[1,])),
  condition.columns=c(1:3)
)

# fitting the model
mm=mcmc.otu(
  fixed="bank+species+bank:species",
  data=gs,
  nitt=3000,burnin=2000 # remove this line for real analysis!
)

# selecting the OTUs that were modeled reliably
acpass=otuByAutocorr(mm,gs)

# calculating effect sizes and p-values:
ss=OTUsummary(mm,gs,summ.plot=FALSE)

# correcting for mutliple comparisons (FDR)
ss=padjustOTU(ss)

# getting significatly changing OTUs (FDR<0.05)
sigs=signifOTU(ss)

# plotting them
ss2=OTUsummary(mm,gs,otus=sigs)

# bar-whiskers graph of relative changes:
# ssr=OTUsummary(mm,gs,otus=signifOTU(ss),relative=TRUE)

# displaying effect sizes and p-values for significant OTUs
ss$otuWise[sigs]
```

Description

Generates a table of model-derived log₂-transformed OTU abundances without sample effects (i.e., corresponding to predicted normalized data)

Usage

```
getNormalizedOTUdata(model, data, log.base=10)
```

Arguments

| | |
|----------|---|
| model | mcmc.otu model: the output of mcmc.otu function fitted with pr=TRUE option |
| data | The dataset that was analysed to generate the model (output of otuStack function) |
| log.base | base of the logarithm to use |

Value

The function returns a list of two data frames. The first one, normData, is the model-predicted log₂-transformed OTU abundances table. It has one column per OTU and one row per sample. The second data frame, conditions, is a table of experimental conditions corresponding to the normData table.

Author(s)

Mikhail V. Matz, University of Texas at Austin <matz@utexas.edu>

References

Elizabeth A. Green, Sarah W. Davies, Mikhail V. Matz, Monica Medina Next-generation sequencing reveals cryptic Symbiodinium diversity within *Orbicella faveolata* and *Orbicella franksi* at the Flower Garden Banks, Gulf of Mexico. PeerJ 2014 <https://peerj.com/preprints/246/>

| | |
|------------|---|
| green.data | <i>Symbiodinium sp. ITS2 OTUs from Orbicella franksi and O. faveolata</i> |
|------------|---|

Description

Symbiodinium sp. ITS2 OTUs from Orbicella franksi and O. faveolata corals at the Flower Garden Banks. Two banks (east and west), two coral species, five OTUs.

Usage

```
data(green.data)
```

Format

A data frame with 58 observations on the following 156 variables.

species a factor with levels faveolata franksi

bank a factor with levels east west

sample a factor with levels EFAV117 EFAV123 EFAV128 EFAV137 EFAV140 EFAV142 EFAV143 EFAV145
EFAV150 EFAV153 EFAV154 EFAV157 EFAV161 EFAV162 EFRA101 EFRA102 EFRA105 EFRA216
EFRA218 EFRA221 EFRA222 EFRA224 EFRA231 EFRA232 EFRA233 EFRA236 EFRA82 EFRA90
EFRA91 WFAV102 WFAV103 WFAV107 WFAV111 WFAV112 WFAV188 WFAV191 WFAV89 WFAV91
WFAV94 WFAV95 WFAV97 WFAV98 WFAV99 WFRA106 WFRA107 WFRA112 WFRA114 WFRA115 WFRA116
WFRA117 WFRA198 WFRA199 WFRA200 WFRA201 WFRA202 WFRA203 WFRA205 WFRA210

H35JRAZ01A0C9J a numeric vector

H35JRAZ01A0D1B a numeric vector

H35JRAZ01A0RQ0 a numeric vector

H35JRAZ01A1FKL a numeric vector

H35JRAZ01A1GT3 a numeric vector

H35JRAZ01A1NMH a numeric vector

H35JRAZ01A1S03 a numeric vector

H35JRAZ01A1VUD a numeric vector

H35JRAZ01A22NW a numeric vector

H35JRAZ01A2ZJK a numeric vector

H35JRAZ01A2ZYY a numeric vector

H35JRAZ01A30IE a numeric vector

H35JRAZ01A32BS a numeric vector

H35JRAZ01A3DXY a numeric vector

H35JRAZ01A3UJS a numeric vector

H35JRAZ01A4BHG a numeric vector

H35JRAZ01A4F30 a numeric vector

H35JRAZ01A4G60 a numeric vector

H35JRAZ01A408A a numeric vector

H35JRAZ01A4U50 a numeric vector

H35JRAZ01A5LTB a numeric vector

H35JRAZ01A509U a numeric vector

H35JRAZ01A6ITE a numeric vector

H35JRAZ01A771F a numeric vector

H35JRAZ01A7EP4 a numeric vector

H35JRAZ01A7SMY a numeric vector

H35JRAZ01A8A59 a numeric vector

H35JRAZ01A8X2Y a numeric vector

H35JRAZ01A93HV a numeric vector
H35JRAZ01A9TZY a numeric vector
H35JRAZ01AMZP0 a numeric vector
H35JRAZ01AN7JF a numeric vector
H35JRAZ01ANHED a numeric vector
H35JRAZ01ANUPT a numeric vector
H35JRAZ01A07GS a numeric vector
H35JRAZ01A0N2S a numeric vector
H35JRAZ01APP5A a numeric vector
H35JRAZ01APPG5 a numeric vector
H35JRAZ01APWEH a numeric vector
H35JRAZ01APYLM a numeric vector
H35JRAZ01AQIU4 a numeric vector
H35JRAZ01AR1R8 a numeric vector
H35JRAZ01AR8Q4 a numeric vector
H35JRAZ01ARD6M a numeric vector
H35JRAZ01ARKYW a numeric vector
H35JRAZ01ARSTD a numeric vector
H35JRAZ01ARZ7L a numeric vector
H35JRAZ01AS8ZF a numeric vector
H35JRAZ01ASV28 a numeric vector
H35JRAZ01ATSK2 a numeric vector
H35JRAZ01ATZCY a numeric vector
H35JRAZ01AU38T a numeric vector
H35JRAZ01AUMOE a numeric vector
H35JRAZ01AUTH5 a numeric vector
H35JRAZ01AUX8X a numeric vector
H35JRAZ01AVB0Q a numeric vector
H35JRAZ01AVFAX a numeric vector
H35JRAZ01AVX70 a numeric vector
H35JRAZ01AWGQC a numeric vector
H35JRAZ01AXA3Z a numeric vector
H35JRAZ01AXGLD a numeric vector
H35JRAZ01AXVIR a numeric vector
H35JRAZ01AY2DE a numeric vector
H35JRAZ01AY6KB a numeric vector
H35JRAZ01AZFWS a numeric vector

H35JRAZ01AZTST a numeric vector
H35JRAZ01BB9FU a numeric vector
H35JRAZ01BBITP a numeric vector
H35JRAZ01BBSV6 a numeric vector
H35JRAZ01BD2HV a numeric vector
H35JRAZ01BDBD6 a numeric vector
H35JRAZ01BDTVF a numeric vector
H35JRAZ01BE95X a numeric vector
H35JRAZ01BEHEK a numeric vector
H35JRAZ01BEMYI a numeric vector
H35JRAZ01BEP0J a numeric vector
H35JRAZ01BFNRT a numeric vector
H35JRAZ01BFVYJ a numeric vector
H35JRAZ01BG87N a numeric vector
H35JRAZ01BGG1A a numeric vector
H35JRAZ01BGN2J a numeric vector
H35JRAZ01BGUP9 a numeric vector
H35JRAZ01BH7VV a numeric vector
H35JRAZ01BIHI3 a numeric vector
H35JRAZ03C08W3 a numeric vector
H35JRAZ03C0K1K a numeric vector
H35JRAZ03C141R a numeric vector
H35JRAZ03C16B4 a numeric vector
H35JRAZ03C1INY a numeric vector
H35JRAZ03C1WD2 a numeric vector
H35JRAZ03C2344 a numeric vector
H35JRAZ03C2919 a numeric vector
H35JRAZ03C2CBI a numeric vector
H35JRAZ03C2DMJ a numeric vector
H35JRAZ03C2MH3 a numeric vector
H35JRAZ03C2PQY a numeric vector
H35JRAZ03C32WG a numeric vector
H35JRAZ03C3ANP a numeric vector
H35JRAZ03C5910 a numeric vector
H35JRAZ03C5CKR a numeric vector
H35JRAZ03C5G30 a numeric vector
H35JRAZ03C60I8 a numeric vector

H35JRAZ03C6AF1 a numeric vector
H35JRAZ03C6E0W a numeric vector
H35JRAZ03C6LU3 a numeric vector
H35JRAZ03C6Q3V a numeric vector
H35JRAZ03C750M a numeric vector
H35JRAZ03C7GCE a numeric vector
H35JRAZ03C7P6J a numeric vector
H35JRAZ03C8JJT a numeric vector
H35JRAZ03C9MVI a numeric vector
H35JRAZ03CX46J a numeric vector
H35JRAZ03CX5BC a numeric vector
H35JRAZ03CXVYS a numeric vector
H35JRAZ03DA80G a numeric vector
H35JRAZ03DALPN a numeric vector
H35JRAZ03DAW5J a numeric vector
H35JRAZ03DAXEI a numeric vector
H35JRAZ03DC5UZ a numeric vector
H35JRAZ03DCCVA a numeric vector
H35JRAZ03DCGCE a numeric vector
H35JRAZ03DCXOX a numeric vector
H35JRAZ03DE7WW a numeric vector
H35JRAZ03DED1S a numeric vector
H35JRAZ03DEJ2S a numeric vector
H35JRAZ03DF9NA a numeric vector
H35JRAZ03DFE9W a numeric vector
H35JRAZ03DFZXI a numeric vector
H35JRAZ03DG0DK a numeric vector
H35JRAZ03DGHF2 a numeric vector
H35JRAZ03DHRCR a numeric vector
H35JRAZ03DIOTH a numeric vector
H35JRAZ03DJAMF a numeric vector
H35JRAZ03DJLV9 a numeric vector
H35JRAZ03DJU1V a numeric vector
H35JRAZ03DKE4J a numeric vector
H35JRAZ03DKR09 a numeric vector
H35JRAZ03DLA3U a numeric vector
H35JRAZ03DLJW2 a numeric vector

H35JRAZ03DLRGI a numeric vector
 H35JRAZ03DMDRM a numeric vector
 H35JRAZ03DMGA2 a numeric vector
 H35JRAZ03DMGAO a numeric vector
 H35JRAZ03D0292 a numeric vector
 H35JRAZ03DOWHJ a numeric vector
 H35JRAZ03DP0DM a numeric vector
 H35JRAZ03DQ55S a numeric vector
 H35JRAZ03DQK7F a numeric vector
 H35JRAZ03DR909 a numeric vector
 H35JRAZ03DRJ6E a numeric vector
 H35JRAZ03DSA3E a numeric vector
 H35JRAZ03DSJQA a numeric vector
 H35JRAZ03DSR5L a numeric vector

Source

Elizabeth A. Green, Sarah W. Davies, Mikhail V. Matz, Monica Medina Next-generation sequencing reveals cryptic Symbiodinium diversity within *Orbicella faveolata* and *Orbicella franksi* at the Flower Garden Banks, Gulf of Mexico. PeerJ 2014 <https://peerj.com/preprints/246/>

References

Elizabeth A. Green, Sarah W. Davies, Mikhail V. Matz, Monica Medina Next-generation sequencing reveals cryptic Symbiodinium diversity within *Orbicella faveolata* and *Orbicella franksi* at the Flower Garden Banks, Gulf of Mexico. PeerJ 2014 <https://peerj.com/preprints/246/>

| | |
|--------|--|
| logLin | <i>prepares OTU counts data for PCA analysis using log-linear-hybrid transform</i> |
|--------|--|

Description

Normalizes counts based on total counts per sample, then uses log-linear-hybrid transformation on the data.

Usage

```
logLin(data,count.columns,k=10,zero.na=FALSE)
```

Arguments

| | |
|----------------------------|---|
| <code>data</code> | OTU counts table. Rows: samples, columns: OTUs. |
| <code>count.columns</code> | numbers of columns containing OTU counts. |
| <code>k</code> | count threshold to transition from linear to log transform. |
| <code>zero.na</code> | With <code>'zero.na=TRUE'</code> , the original zero count datapoints become <code>'NA'</code> (useful for analysis of pairwise correlations between OTUs). |

Value

A table of transformed normalized values.

Author(s)

Mikhail V. Matz, University of Texas at Austin <matz@utexas.edu>

References

Log-linear-hybrid transform: Rocke DM and Durbin B *Bioinformatics* 2003 19:966-972 (DOI: 10.1093/bioinformatics/btg107) MCMC.OTU: Elizabeth A. Green, Sarah W. Davies, Mikhail V. Matz, Monica Medina Next-generation sequencing reveals cryptic *Symbiodinium* diversity within *Orbicella faveolata* and *Orbicella franksi* at the Flower Garden Banks, Gulf of Mexico. *PeerJ* 2014 <https://peerj.com/preprints/246/>

| | |
|----------|--|
| mcmc.otu | <i>Analyzes multivariate counts data using poisson-lognormal mixed model</i> |
|----------|--|

Description

Wrapper function for MCMCglmm by Jarrod Hadfield, designed for multivariate counts data such as in sequence-based analysis of microbial communities ("metabarcoding", variables = operational taxonomic units, OTUs), or in ecological applications (variables = species). The function aims to infer the changes in relative proportions of individual variables. The maximum number of variables that can be processed on a laptop computer is about 200; more memory is required for larger numbers.

Usage

```
mcmc.otu(fixed=NULL, random=NULL, data, y.scale="proportion",
globalMainEffects="remove", vprior="uninf",...)
```

Arguments

| | |
|--------------------------------|--|
| <code>fixed</code> | combination of fixed effects, as a text string. Do not use "*" symbol, list it fully, such as: 'factor1+factor2+factor1:factor2'. |
| <code>random</code> | A vector of names for variable-specific scalar random effects, such as 'c("effect1","effect2")'. |
| <code>data</code> | output of the <code>otuStack()</code> function |
| <code>y.scale</code> | By default, the modeled abundances will be expressed relative to the total counts in the sample, effectively corresponding to proportions of total. Specify 'y.scale="absolute"' to express the results as absolute abundances. |
| <code>globalMainEffects</code> | By default, the model will assume that the samples can vary systematically in abundance among factor combinations (i.e., there is an effect of a factor combination applicable to all variables) and remove these effects; this is analogous to normalizing the samples to total counts. Specify 'globalMainEffects="keep"' to switch this off. |
| <code>vprior</code> | Prior for variance of user-specified random effects. By default an inverse Wishart prior with assumed variance (V) at 1 and the degree of belief parameter (nu) at 0. With 'prior="iw"' and 'prior="iw01"' nu is the number of OTUs minus 0.998, resulting in a weakly informative prior that is commonly used in this type of inference. 'vprior="iw"' will assume large prior variance (1), 'vprior="iw01"' will assume small prior variance (0.1). If the model has trouble converging, specify 'vtior="iw"'. |
| <code>...</code> | other options for <code>MCMCglmm</code> function, such as <code>nitt</code> (number of iterations), <code>thin</code> (thinning interval), and <code>burnin</code> (number of initial iterations to disregard). For a more precise inference, specify 'nitt=50000, thin=25, burnin=5000'. See <code>MCMCglmm</code> documentation for more details. |

Details

This function constructs priors and runs an MCMC chain to fit a Poisson-lognormal generalized linear mixed model to the multivariate counts data.

The fixed effects for the model by default include a variable-specific intercept, global (non-variable-specific) main effects of fixed factors, and variable-specific effect for each of the listed fixed factors. With `globalMainEffects="keep"` the model will not include the global main effects, resulting in them being absorbed into the variable-specific effects.

The user-specified random effects are all assumed to be variable-specific with no covariances.

The model includes one universal random factor: the scalar random effect of sample, which accounts for the unequal counting effort among samples.

Residual variances are assumed to be variable-specific with no covariances, with weakly informative inverse Wishart prior with `variance=1` and `nu=(number of variables)-0.998`.

The priors for fixed effects are diffuse gaussians with a mean at 0 and very large variances ($1e+8$),

Value

An `MCMCglmm` object. `OTUsummary()` function within this package summarizes these data, calculates all variable-wise credible intervals and p-values, and plots the results either as line-point-whiskers graph or a bar-whiskers graph using `ggplot2` functions.

OTUsummary() only works for experiments with a single multilevel factor or two fully crossed multilevel factors.

For more useful operations on MCMCglmm objects, such as posterior.mode(), HPDinterval(), and plot(), see documentation for MCMCglmm package.

Author(s)

Mikhail V. Matz, University of Texas at Austin <matz@utexas.edu>

References

Elizabeth A. Green, Sarah W. Davies, Mikhail V. Matz, Monica Medina Next-generation sequencing reveals cryptic Symbiodinium diversity within Orbicella faveolata and Orbicella franksi at the Flower Garden Banks, Gulf of Mexico. PeerJ 2014 <https://peerj.com/preprints/246/>

See Also

OTUsummary(),MCMCglmm()

Examples

```
# Symbiodinium sp diversity in two coral species at two reefs (banks)
data(green.data)

# removing outliers
goods=purgeOutliers(
  data=green.data,
  count.columns=c(4:length(green.data[1,])),
  zero.cut=0.25 # remove this line for real analysis
)

# stacking the data table
gs=otuStack(
  data=goods,
  count.columns=c(4:length(goods[1,])),
  condition.columns=c(1:3)
)

# fitting the model
mm=mcmc.otu(
  fixed="bank+species+bank:species",
  data=gs,
  nitt=3000,burnin=2000 # remove this line for real analysis!
)

# selecting the OTUs that were modeled reliably
acpass=otuByAutocorr(mm,gs)

# calculating effect sizes and p-values:
ss=OTUsummary(mm,gs,summ.plot=FALSE)
```

```
# correcting for multiple comparisons (FDR)
ss=padjustOTU(ss)

# getting significantly changing OTUs (FDR<0.05)
sigs=signifOTU(ss)

# plotting them
ss2=OTUsummary(mm,gs,otus=sigs)

# bar-whiskers graph of relative changes:
# ssr=OTUsummary(mm,gs,otus=signifOTU(ss),relative=TRUE)

# displaying effect sizes and p-values for significant OTUs
ss$otuWise[sigs]
```

`mcmc.pval`*calculates p-value based on Bayesian z-score or MCMC sampling*

Description

Accessory function to OTUsummary function

Usage

```
mcmc.pval(dat, testlim = 0, sided = 2, ptype="z")
```

Arguments

| | |
|----------------------|---|
| <code>dat</code> | a table of MCMC samples |
| <code>testlim</code> | positive value to test whether the posterior crosses it (for variance components) |
| <code>sided</code> | sidedness of the test |
| <code>ptype</code> | 'z' for z-score based p-values, 'mcmc' for standard sampling based p-values |

Value

A vector of calculated Bayesian p-values. For standard 'mcmc' p-values, the minimal possible value is 1/(MCMC sample size). z-score based p-values are useful to approximate very low p-values with limited MCMC sample sizes.

Author(s)

Mikhail Matz, UT Austin <matz@utexas.edu>

| | |
|---------------|--|
| otuByAutocorr | <i>Selects OTUs for which MCMC-based parameter estimates are reliable.</i> |
|---------------|--|

Description

Looks at autocorrelation among sampled parameter values.

Usage

```
otuByAutocorr(model, data, ac.cut=0.1)
```

Arguments

| | |
|--------|--|
| model | Model generated by <code>mcmc.otu()</code> |
| data | Dataset used to fit the model |
| ac.cut | Autocorrelation cutoff |

Value

Vector of OTU names for which autocorrelation is less than `ac.cut` argument.

Author(s)

Mikhail V. Matz, University of Texas at Austin <matz@utexas.edu>

References

Elizabeth A. Green, Sarah W. Davies, Mikhail V. Matz, Monica Medina Next-generation sequencing reveals cryptic Symbiodinium diversity within *Orbicella faveolata* and *Orbicella franksi* at the Flower Garden Banks, Gulf of Mexico. PeerJ 2014 <https://peerj.com/preprints/246/>

| | |
|----------|---|
| otuStack | <i>Prepares OTU counts data for MCMC model fitting using <code>mcmc.otu()</code>.</i> |
|----------|---|

Description

Stacks all OTU columns into one (count), creates a column for OTU names and columns for experimental conditions. Adds a fake OTU called "summ" - sum of all counts per sample.

Usage

```
otuStack(data, count.columns, condition.columns)
```

Arguments

`data` The input data table: one column per OTU, plus columns containing experimental conditions and the required column 'sample' denoting biological replicates. See `green.data` included with the package.

`count.columns` vector of numbers identifying OTU counts columns

`condition.columns` vector of numbers identifying columns containing experimental conditions

Value

A dataset that can be fed into `mcmc.otu()` function.

Author(s)

Mikhail V. Matz, University of Texas at Austin <matz@utexas.edu>

References

Elizabeth A. Green, Sarah W. Davies, Mikhail V. Matz, Monica Medina Next-generation sequencing reveals cryptic Symbiodinium diversity within *Orbicella faveolata* and *Orbicella franksi* at the Flower Garden Banks, Gulf of Mexico. PeerJ 2014 <https://peerj.com/preprints/246/>

Examples

```
# Symbiodinium sp diversity in two coral species at two reefs (banks)
data(green.data)
green.data

# stacking the data table
gs=otuStack(green.data,count.columns=c(4:8),condition.columns=c(1:3))
head(gs,30)
```

OTUsummary

Summarizes and plots results of mcmc.otu() function series.

Description

Calculates abundances of each OTU across factor combinations; calculates pairwise differences between all factor combinations and their significances for each OTU; plots results as bar or line graphs with credible intervals (ggplot2) NOTE: only works for experiments involving a single multi-level fixed factor or two fully crossed multi-level fixed factors.

Usage

```
OTUsummary(model, data, otus = NA, relative = FALSE,
log.base = 10, summ.plot = TRUE, ptype = "z", xgroup=NULL, ...)
```

Arguments

| | |
|-----------|---|
| model | Model generated by <code>mcmc.otu()</code> or <code>mcmc.otu.normalized()</code> |
| data | Dataset used to build the model (returned by <code>otuStack()</code> or <code>otuStackNormalize()</code>) |
| otus | A vector of OTU names to summarize and plot. If left unspecified, all OTUs will be summarized. |
| relative | Whether to plot OTU abundances as $\log(\text{proportion of total})$ (default) or fold-changes relative to the sample that is considered to be "global control" (<code>relative = TRUE</code>). The "global control" is the combination of factors that served as a reference during model fitting, either because it is alphanumericly first (that happens by default) or because it has been explicitly designated as such using <code>relevel()</code> function. |
| log.base | Base of the logarithm to use. |
| summ.plot | By default, the function generates a summary plot, which is a line-points-95% credible intervals plot of $\log(\text{fraction of total})$ with <code>'relative=FALSE'</code> and a bar graph of $\log(\text{fold change relative to the control})$, again with 95% credible intervals, with <code>'relative=TRUE'</code> . Specify <code>'summ.plot=FALSE'</code> if you don't want the summary plot. |
| ptype | Which type of p-values to use. By default p-values based on the Bayesian z-score are used. Specify <code>'ptype="mcmc"'</code> to output more conventional p-values based on MCMC sampling (these will be limited on the lower end by the size of MCMC sample). |
| xgroup | For two-factor designs: which of the factors to use to form the x-axis. The other one will be used to form facets. |
| ... | Additional options for <code>summaryPlotOTU()</code> function. Among those, <code>'x.order'</code> can be a vector specifying the order of factor levels on the x-axis. |

Value

A list of three items:

| | |
|---------|--|
| summary | Summary table containing calculated abundances, their SD and 95% credible limits |
| otuWise | A series of matrices listing pairwise differences between factor combinations (upper triangle) and corresponding p-values (lower triangle) |
| ggPlot | the <code>ggplot2</code> object for plotting. See http://docs.ggplot2.org/0.9.2.1/theme.html for ways to modify it, such as add text, rotate labels, change fonts, etc. |

Author(s)

Mikhail V. Matz, University of Texas at Austin <matz@utexas.edu>

References

Elizabeth A. Green, Sarah W. Davies, Mikhail V. Matz, Monica Medina Next-generation sequencing reveals cryptic *Symbiodinium* diversity within *Orbicella faveolata* and *Orbicella franksi* at the Flower Garden Banks, Gulf of Mexico. *PeerJ* 2014 <https://peerj.com/preprints/246/>

See Also

mcmc.otu(),MCMCglmm()

Examples

```
# see example in ?MCMC.OTU
```

| | |
|------------|--|
| padjustOTU | <i>Adjusts p-values in the OTU summary for multiple comparisons.</i> |
|------------|--|

Description

Collects all p-values from the summary, performs FDR correction (Benjamini-Hochberg), returns the summary with adjusted p-values.

Usage

```
padjustOTU(otu.summary, method="BH")
```

Arguments

| | |
|-------------|---|
| otu.summary | The output of OTUsummary function. |
| method | Method of adjustment for multiple comparisons (see ?p.adjust) |

Value

An OTU summary list (see ?OTUsummary) with adjusted p-values

Author(s)

Mikhail V. Matz, University of Texas at Austin <matz@utexas.edu>

References

Elizabeth A. Green, Sarah W. Davies, Mikhail V. Matz, Monica Medina Next-generation sequencing reveals cryptic Symbiodinium diversity within *Orbicella faveolata* and *Orbicella franksi* at the Flower Garden Banks, Gulf of Mexico. PeerJ 2014 <https://peerj.com/preprints/246/>

panel.cor *accessory function for pairs() to display Pearson correlations*

Description

works as upper.panel or lower.panel argument of pairs() (package graphics).

Usage

```
panel.cor(x, y, digits=2, cex.cor)
```

Arguments

| | |
|---------|---|
| x | x element of the pairs() matrix |
| y | y element of the pairs() matrix |
| digits | number of non-zero digits to leave at the end |
| cex.cor | scaling factor for displayed text |

References

cannibalized from an example in ?pairs (package graphics)

panel.cor.pval *accessory function for pairs() to display pvalue of the Pearson correlation*

Description

works as upper.panel or lower.panel argument of pairs() (package graphics). Displays pvalues better than 0.1.

Usage

```
panel.cor.pval(x, y, digits = 2, cex.cor, p.cut=0.1)
```

Arguments

| | |
|---------|---|
| x | x element of the pairs() matrix |
| y | y element of the pairs() matrix |
| digits | number of non-zero digits to leave at the end |
| cex.cor | scaling factor for displayed text |
| p.cut | p-value cutoff |

References

cannibalized from an example in ?pairs (package graphics)

| | |
|---------------|--|
| purgeOutliers | <i>Removes outlier samples and OTUs.</i> |
|---------------|--|

Description

Samples are removed if their total counts fall below the specified z-score cutoff (default -2.5); OTUs are removed if their sum of counts is less than specified fraction of total counts (default 0.001), or if they are observed in less than specified fraction of all samples (default 0.2). Samples are removed first, then OTUs.

Usage

```
purgeOutliers(data, count.columns, sampleZcut=(-2.5), otu.cut=0.001, zero.cut=0)
```

Arguments

| | |
|---------------|--|
| data | A data frame containing counts, with samples as rows and OTUs as columns |
| count.columns | Vector of OTU column numbers. |
| sampleZcut | z-score cutoff for removing low-coverage outlier samples. |
| otu.cut | cutoff for a fraction of total counts the OTU has to represent. This would depend on the methodology and nature of OTUs; for coral symbionts (<i>Symbiodinium</i> sp) the reasonable cutoff is 0.001 (Quigley et al, PLoS ONE 2014) |
| zero.cut | fraction of samples in which the OTU is found. |

Value

The data table with outlier samples and OTUs removed.

Author(s)

Mikhail V. Matz, University of Texas at Austin <matz@utexas.edu>

References

Elizabeth A. Green, Sarah W. Davies, Mikhail V. Matz, Monica Medina Next-generation sequencing reveals cryptic *Symbiodinium* diversity within *Orbicella faveolata* and *Orbicella franksi* at the Flower Garden Banks, Gulf of Mexico. PeerJ 2014 <https://peerj.com/preprints/246/>

| | |
|-----------|---|
| signifOTU | <i>Finds differentially represented OTUs.</i> |
|-----------|---|

Description

Trawls otuWise summary tables in the data returned by OTUsummary and returns the names of OTUs that have p-values better than the specified cutoff.

Usage

```
signifOTU(otu.summary, p.cutoff=0.05)
```

Arguments

| | |
|-------------|------------------------------------|
| otu.summary | The output of OTUsummary function. |
| p.cutoff | p-value cutoff to report OTUs. |

Value

A vector of significant OTUs.

Author(s)

Mikhail V. Matz, University of Texas at Austin <matz@utexas.edu>

References

Elizabeth A. Green, Sarah W. Davies, Mikhail V. Matz, Monica Medina Next-generation sequencing reveals cryptic Symbiodinium diversity within *Orbicella faveolata* and *Orbicella franksi* at the Flower Garden Banks, Gulf of Mexico. PeerJ 2014 <https://peerj.com/preprints/246/>

| | |
|------------|--|
| startedLog | <i>prepares OTU counts data for PCA analysis using started-log transform</i> |
|------------|--|

Description

Normalizes counts based on total counts per sample, adds a constant number (argument logstart) to get rid of zeroes, then log-transforms the data.

Usage

```
startedLog(data, count.columns, logstart=0.1)
```

Arguments

| | |
|---------------|--|
| data | OTU counts table. Rows: samples, columns: OTUs. |
| count.columns | numbers of columns containing OTU counts. |
| logstart | number to add to zero counts before log transform. With 'logstart=0', the zero count datapoints become 'NA' (useful for analysis of pairwise correlations between OTUs). |

Value

A table of log-transformed normalized values.

Author(s)

Mikhail V. Matz, University of Texas at Austin <matz@utexas.edu>

References

Started Log transform: Rocke DM and Durbin B *Bioinformatics* 2003 19:966-972 (DOI: 10.1093/bioinformatics/btg107) MCMC.OTU: Elizabeth A. Green, Sarah W. Davies, Mikhail V. Matz, Monica Medina Next-generation sequencing reveals cryptic Symbiodinium diversity within *Orbicella faveolata* and *Orbicella franksi* at the Flower Garden Banks, Gulf of Mexico. *PeerJ* 2014 <https://peerj.com/preprints/246/>

| | |
|----------------|---|
| summaryPlotOTU | <i>Wrapper function for ggplot2 to make bar and line graphs of mcmc.otu() results</i> |
|----------------|---|

Description

This function is called automatically by OTUsummary() and also can be used separately to plot the results produced by OTUsummary() with more plotting options

Usage

```
summaryPlotOTU(data, xgroup, facet = NA, type = "bar",
x.order = NA, whiskers = "ci", otus = NA, log.base=10)
```

Arguments

| | |
|--------|---|
| data | A summary table generated by HPDplot(), it is the first element in the returned list. |
| xgroup | Which factor will be used to form the x axis (for 2-way designs). |
| facet | The factor by which the plot will be split into facets (for 2-way designs). |

| | |
|----------|--|
| type | Two types are supported: "bar" and "line" ("line" also has points). "bar" is more useful to plot fold-changes returned when OTUsummary() is run with the option 'relative=TRUE'. "line" is better for plotting actual inferred OTU abundances across factor levels; it is particularly good for time courses and other cases when multiple factor levels must be compared to each other. "bar" is good to plot log(fold-changes) relative to global control. |
| x.order | A vector giving the order of factor levels on the x-axis. If unspecified, alphanumeric order will be used. |
| whiskers | The interval indicated by the whiskers. Default is "ci", the 95% credible interval; another option is "sd" - standard deviation of the posterior. |
| otus | Vector of OTU names to plot. By default, all OTUs in the summary will be plotted. |
| log.base | Base of the logarithm to indicate on the y-axis label. |

Details

The function invokes ggplot() function from the ggplot2 package to plot the results either as a single panel (one-way designs) or a multi-panel (2-way designs, one panel per level of the factor specified by 'facet' argument).

Value

A ggplot object. See <http://docs.ggplot2.org/0.9.2.1/theme.html> for ways to modify it, such as add text, rotate labels, change fonts, etc.

Author(s)

Mikhail V. Matz, University of Texas at Austin <matz@utexas.edu>

References

Elizabeth A. Green, Sarah W. Davies, Mikhail V. Matz, Monica Medina Next-generation sequencing reveals cryptic Symbiodinium diversity within *Orbicella faveolata* and *Orbicella franksi* at the Flower Garden Banks, Gulf of Mexico. PeerJ 2014 <https://peerj.com/preprints/246/>

Index

[getNormalizedOTUdata](#), 3
[green.data](#), 4

[logLin](#), 9

[MCMC.OTU \(MCMC.OTU-package\)](#), 2
[mcmc.otu](#), 10
[MCMC.OTU-package](#), 2
[mcmc.pval](#), 13

[otuByAutocorr](#), 14
[otuStack](#), 14
[OTUsummary](#), 15

[padjustOTU](#), 17
[panel.cor](#), 18
[panel.cor.pval](#), 18
[purgeOutliers](#), 19

[signifOTU](#), 20
[startedLog](#), 20
[summaryPlotOTU](#), 21