Package 'CoDaSeq'

August 16, 2016

Type Package

Title Compositional Data Analysis of High Throughput Sequencing

Version 0.99.1

Date 2016-05-21

Author Greg Gloor, Wallace Chan

Maintainer Greg Gloor <ggloor@uwo.ca>

Description A set of common functions for the analysis of high throughput sequencing count data

License file LICENSE

NeedsCompilation no

R topics documented:

`codaSeq.clr` *Center Log-Ratio Function.*

Description

Returns a matrix of center log-ratio transformed data with samples by row. Equivalent to log(x/gx) for every value where gx is the geometic mean of the vector X.

Usage

```
codaSeq.clr <- function(x, samples.by.row=TRUE)
```

Arguments

`x`	A matrix or dataframe with samples by rows or columns.
`samples.by.row`	TRUE if samples are by row, FALSE if samples are by column.

Details

Natural log is used for biplots and other exploratory analyses.

Value

returns a matrix of clr tranformed values with samples in the rows and variables in columns

Author(s)

Greg Gloor, Jean Macklaim, Wallace Chan

References

Please use the citation given by `citation(package="CoDaSeq")`

See Also

`codaSeq.filter`, `codaSeq.outlier`, `codaSeq.rarefy`, `codaSeq.propr.phismy`, `codaSeq.propr.aldex.phi`

`codaSeq.filter` *Filter compositional dataset for 0 values and abundance.*

Description

Returns a reduced able of counts where the samples must contain a minimum number of reads, and OTUs must be found with a minimum abundance in all remaining samples.

Usage

```
codaSeq.filter <- function(x, y=tax.vector, min.reads=5000, min.prop=0.001, max.prop=0.025,
  min.occurrence=0, samples.by.row=TRUE)
```

Arguments

`x`	A matrix or dataframe containing a count table.
`min.reads`	The minimum reads per sample. Default=5000.
`min.prop`	The minimum proportional abundance of a read in any sample. Default=0.001.
`max.prop`	The maximum proportional abundance of a read in any sample. Default=0.025.
`min.fraction`	The minimum sample proportion of non-0 reads for each variable.
`sample.by.row`	True if rows contain samples, false if rows contain variables.

Details

Filters min/max.prop first, min.fraction second, min/max.prop third. Requires numeric data only.

Value

Returns a dataframe with the following information:

`data.0`

`data.1`

`data.2` Returns a reduced vector with filtered samples by rows.

Author(s)

Greg Gloor, Jean Macklai, Wallace Chan

References

Please use the citation given by `citation(package="CoDaSeq")`

See Also

`codaSeq.clr`, `codaSeq.outlier`, `codaSeq.rarefy`, `codaSeq.propr.phismy`, `codaSeq.propr.aldex.phi`

---

codaSeq.outlier *Identifying sample Outliers.*

Description

Returns a list of proportional contribution to group variance, sample names that are outliers, and sample names that are not outliers.

Usage

```
codaSeq.outlier <- function(x, plot.me=TRUE, col=rgb(1,0,0,0.3))
```

Arguments

`x`	A matrix or data frame with clr transformed values, with sample by row.
`plot.me`	A logical value determining if a histogram should be plotted of the variance contribution per sample.
`col`	RGB values for your selection of colour.

Details

Samples must be grouped. This approach makes no sense across groups. If you do not know if you have natural groups, ignore this step and exam your data by PCA. Outliers are defined as those contributing greater than the median plus twice the interquartile range of the sample variance to the total.

Value

Returns list

`sample.var`	Proportional variance contributions for each sample.
`bad`	Rownames of outlier samples.
`good`	Rownames of non-outlier samples.

Author(s)

Greg Gloor, Jean Macklai, Wallace Chan

References

Please use the citation given by `citation(package="CoDaSeq")`

See Also

codaSeq.clr, codaSeq.filter, codaSeq.rarefy, codaSeq.propr.phismy, codaSeq.propr.aldex.phi

---

propr.aldex.phi *Expected Value of Phi From Dirichlet Log-Ratio Distribution*

Description

Returns data frame of the lower-triangle of symmetrical phi metric, where the value of phi is the expected value of a number of Dirichlet Monte-Carlo replicates of the data. This reduces the problem of 0-count and low-count features being highly variable because their values range wildly and so the expected value is always large.

Usage

```
propr.aldex.phi <- function(aldex.clr)
```

Details

Requires aldex.clr function from ALDEx2 Package. Param aldex.clr is an S3 object from the aldex.clr function. We ignore all the other measures that are used for trouble-shooting phi. The sma.df function in particular is very time and memory intensive

Value

sym.phi	Calculated sum of phi values through all DIR MR instances.
lt	Indice of correct size.
lt.int	Indice of correct size.
sma.df	Dataframe to hold info.
sma.df$phi	Dataframe to hold the lower triangle because the matrix is symmetrical

Author(s)

Greg Gloor, Jean Macklaim, Wallace Chan

References

Please use the citation given by citation(package="CoDaSeq")

See Also

codaSeq.clr, codaSeq.filter, codaSeq.rarefy, codaSeq.outlier, codaSeq.propr.phismy

propr.phisym *Symmetric Phi Statistics*

Description

Returns a matrix where element (i,j) is the symmetric phi statistic between columns i and j of X.

Usage

```
propr.phismy <- function (X)
```

Arguments

X	A matrix or data frame of centered log ratio transformation.

Details

X should be the result of a centered log-ratio transformation.

Author(s)

Greg Gloor, Jean Macklai, Wallace Chan

References

Please use the citation given by `citation(package="CoDaSeq")`

See Also

codaSeq.clr, codaSeq.filter, codaSeq.outlier, codaSeq.rarefy, codaSeq.propr.aldex.phi

Examples

```
N <- 10
# Number of observations
# Make a data frame with columns a and b roughly proportional
# and columns c and d roughly proportional
  X <- data.frame(a=(1:N), b=(1:N) * rnorm(N, 10, 0.1),
                  c=(N:1), d=(N:1) * rnorm(N, 10, 1.0))
 round(propr.phisym(clr(X)),2)
```

Index