Package 'CoDaSeq'

August 19, 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>

Depends ALDEx2

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.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.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 samples 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.aldex.phi`

`codaSeq.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

```
codaSeq.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 MC 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.rarefy *Subsample dataset without replacement.*

Description

Returns a table of counts where samples have been sampled without replacement. This is included for compatibility, but in general is a bad idea, since it results in a loss of information, and distorts the underlying data somewhat.

Usage

```
codaSeq.rarefy <- function(x, n=1000, samples.by.row=TRUE)
```

Arguments

x	A matrix or dataframe containing a count table.
n	The desired target number of reads per sample. Default=1000.
sample.by.row	True if rows contain samples, false if rows contain OTUs.

Value

Returns a matrix in the same orientation as the original with counts per OTU reduced to the common sampling depth. This is a constant sum operation.

Author(s)

Greg Gloor, Jean Macklaim, Wallace Chan

References

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

See Also

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

Index