

Package ‘enveomics.R’

March 7, 2016

Version 1.0.1

Title Various R Functions from the Kostas Lab

Description A collection of R functions of common use in the Kostas Lab for microbial ecology.

Author Luis M. Rodriguez-R [aut, cre]

Maintainer Luis M. Rodriguez-R <lmrodriguezr@gmail.com>

URL <http://enve-omics.gatech.edu/>

Depends R (>= 2.9), stats, methods, parallel, modeest, fitdistrplus, sn

Suggests tools, vegan, ape, picante, gplots, optparse

License Artistic-2.0

LazyLoad yes

Encoding UTF-8

NeedsCompilation no

Repository CRAN

Date/Publication 2016-03-07 01:09:09

R topics documented:

enveomics.R-package	2
enve.barplot	2
enve.cliopts	4
enve.df2dist	5
enve.df2dist.group	5
enve.df2dist.list	6
enve.prune.dist	7
enve.recplot	8
enve.recplot2	10
enve.RecPlot2-class	11
enve.recplot2.corePeak	12
enve.recplot2.findPeaks	12

enve.RecPlot2.Peak-class	14
enve.recplot2.__counts	15
enve.recplot2.__findPeak	15
enve.recplot2.__findPeaks	16
enve.recplot2.__peakHist	17
enve.recplot2.__whichClosestPeak	17
enve.tribs	18
enve.TRIBS-class	19
enve.TRIBS.merge	20
enve.tribs.test	20
enve.TRIBStest-class	21
enve.__prune.iter	22
enve.__prune.reduce	22
enve.__tribs	23
phyla.counts	23
plot.enve.RecPlot2	24
plot.enve.TRIBS	25
plot.enve.TRIBStest	26
summary.enve.TRIBS	27
summary.enve.TRIBStest	27
\$.methods	28

Index 29

enveomics.R-package *Various R Functions from the Kostas Lab*

Description

A collection of R functions of common use in the Kostas Lab for microbial ecology.

Author(s)

Luis M. Rodriguez-R <lmrodriguezr@gmail.com> [aut, cre]

enve.barplot *enve barplot*

Description

Creates nice barplots from tab-delimited tables

Usage

```
enve.barplot(x, sizes, top = 25, colors.per.group = 9, bars.width = 4,
  legend.ncol = 1, other.col = "#000000", add.trend = FALSE,
  organic.trend = FALSE, sort.by = median, min.report = 101,
  order = NULL, col, ...)
```

Arguments

x	Can be either the input data or the path to the file containing the table. If it contains the data, it must be a data frame or an object coercible to data frame. If it is a path, it must point to a tab-delimited file containing header (first row) and row names (first column).
sizes	A numeric vector containing the real size of the samples (columns) in the same order of the input table. If set, the values are assumed to be 100%, otherwise the sum of the columns is used.
top	Maximum number of categories to display. Any additional categories will be listed as "Others".
colors.per.group	Number of categories in the first two saturation groups of colors. The third group contains the remaining categories if needed.
bars.width	Width of the barplot with respect to the legend.
legend.ncol	Number of columns in the legend.
other.col	Color of the "Others" category.
add.trend	Controls if semi-transparent areas are to be plotted between the bars to connect the regions (trend regions).
organic.trend	Controls if the trend regions are to be smoothed (curves). By default, trend regions have straight edges. If TRUE, forces add.trend=TRUE.
sort.by	Any function that takes a numeric vector and returns a numeric scalar. This function is applied to each row, and the resulting values are used to sort the rows (decreasingly). Good options include: sd, min, max, mean, median.
min.report	Minimum percentage to report the value in the plot. Any value above 100 indicates that no values are to be reported.
order	Controls how the rows should be ordered. If NULL (default), sort.by is applied per row and the results are sorted decreasingly. If NA, no sorting is performed, i.e., the original order is respected. If a vector is provided, it is assumed to be the custom order to be used (either by numeric index or by row names).
col	Colors to use. If provided, overrides the variables 'top' and 'colors.per.group', but 'other.col' is still used if the vector is insufficient for all the rows.
...	Any additional parameters to be passed to 'barplot'.

Author(s)

Luis M. Rodriguez-R [aut, cre]

Examples

```
# Load data
data("phyla.counts", package="enveomics.R", envir=environment())
# Create a barplot sorted by variance with organic trends
enve.barplot(phyla.counts, # Counts of phyla in four sites
  sizes=c(250,100,75,200), # Total sizes of the datasets of each site
  bars.width=2, # Decrease from default, so the names are fully displayed
```

```

    organic.trend=TRUE, # Nice curvy background
    sort.by=var # Sort by variance across sites
  )

```

enve.cliopts

enve cliopts

Description

Generates nicely formatted command-line interfaces for functions (`_closures_` only).

Usage

```

enve.cliopts(fx, rd_file, positional_arguments, usage, mandatory = c(),
  vectorize = c(), ignore = c(), number = c(), defaults = list(),
  o_desc = list(), p_desc = "")

```

Arguments

<code>fx</code>	Function for which the interface should be generated.
<code>rd_file</code>	(Optional) .Rd file with the standard documentation of the function.
<code>positional_arguments</code>	(Optional) Number of <code>_positional_</code> arguments passed to <code>parse_args</code> (package:optparse).
<code>usage</code>	(Optional) Usage passed to <code>OptionParser</code> (package:optparse).
<code>mandatory</code>	Mandatory arguments.
<code>vectorize</code>	Arguments of the function to vectorize (comma-delimited). If numeric, use also <code>'number'</code> .
<code>ignore</code>	Arguments of the function to ignore.
<code>number</code>	Force these arguments as numerics. Useful for numeric vectors (see <code>'vectorize'</code>) or arguments with no defaults.
<code>defaults</code>	Defaults to use instead of the ones provided by the formals.
<code>o_desc</code>	Descriptions of the options. Help from <code>'rd'</code> is ignored for arguments present in this list.
<code>p_desc</code>	Description of the function. Help from <code>'rd'</code> is ignored for the function description unless this value is an empty string.

Value

Returns a `'list'` with keys: `'options'`, a named list with the values for the function's arguments; and `'args'`, a vector with zero or more strings containing the positional arguments.

Author(s)

Luis M. Rodriguez-R [aut, cre]

enve.df2dist	<i>enve df2dist</i>
--------------	---------------------

Description

Transform a dataframe (or coercible object, like a table) into a ‘dist’ object.

Usage

```
enve.df2dist(x, obj1.index = 1, obj2.index = 2, dist.index = 3,
            default.d = NA, max.sim = 0)
```

Arguments

x	A table (or coercible object) with at least three columns: (1) ID of the object 1, (2) ID of the object 2, and (3) distance between the two objects.
obj1.index	Index of the column containing the ID of the object 1.
obj2.index	Index of the column containing the ID of the object 2.
dist.index	Index of the column containing the distance.
default.d	Default value (for missing values)
max.sim	If not-zero, assumes that the values are similarity (not distance) and this is the maximum similarity (corresponding to distance 0). Applies transformation: distance = (max.sim - values)/max.sim.

Value

Returns a ‘dist’ object.

Author(s)

Luis M. Rodriguez-R [aut, cre]

enve.df2dist.group	<i>enve df2dist group</i>
--------------------	---------------------------

Description

Transform a dataframe (or coercible object, like a table) into a ‘dist’ object, where there are 1 or more distances between each pair of objects.

Usage

```
enve.df2dist.group(x, obj1.index = 1, obj2.index = 2, dist.index = 3,
                  summary = median, empty.rm = TRUE)
```

Arguments

x	A dataframe (or coercible object) with at least three columns: (1) ID of the object 1, (2) ID of the object 2, and (3) distance between the two objects.
obj1.index	Index of the column containing the ID of the object 1.
obj2.index	Index of the column containing the ID of the object 2.
dist.index	Index of the column containing the distance.
summary	Function summarizing the different distances between the two objects.
empty.rm	Remove rows with empty or NA groups

Value

Returns a ‘dist’ object.

Author(s)

Luis M. Rodriguez-R [aut, cre]

enve.df2dist.list *enve df2dist list*

Description

Transform a dataframe (or coercible object, like a table) into a ‘dist’ object.

Usage

```
enve.df2dist.list(x, groups, obj1.index = 1, obj2.index = 2,
  dist.index = 3, empty.rm = TRUE, ...)
```

Arguments

x	A dataframe (or coercible object) with at least three columns: (1) ID of the object 1, (2) ID of the object 2, and (3) distance between the two objects.
groups	Named array where the IDs correspond to the object IDs, and the values correspond to the group.
obj1.index	Index of the column containing the ID of the object 1.
obj2.index	Index of the column containing the ID of the object 2.
dist.index	Index of the column containing the distance.
empty.rm	Remove incomplete matrices
...	Any other parameters supported by ‘enve.df2dist.group’.

Value

Returns a ‘list’ of ‘dist’ object.

Author(s)

Luis M. Rodriguez-R [aut, cre]

enve.prune.dist *enve prune dist*

Description

Automatically prunes a tree, to keep representatives of each clade.

Usage

```
enve.prune.dist(t, dist.quantile = 0.25, min_dist, quiet = FALSE,  
               max_iters = 100, min_nodes_random = 40000, random_nodes_frx = 1)
```

Arguments

t	A 'phylo' object or a path to the Newick file.
dist.quantile	The quantile of edge lengths.
min_dist	The minimum distance to allow between two tips. If not set, dist.quantile is used instead to calculate it.
quiet	Boolean indicating if the function must run without output.
max_iters	Maximum number of iterations.
min_nodes_random	Minimum number of nodes to trigger "tip-pairs" nodes sampling. This sampling is less reproducible and more computationally expensive, but it's the only solution if the cophenetic matrix exceeds $2^{31}-1$ entries; above that, it cannot be represented in R.
random_nodes_frx	Fraction of the nodes to be sampled if more than 'min_nodes_random'.

Value

Returns a pruned phylo object.

Author(s)

Luis M. Rodriguez-R [aut, cre]

 enve.recplot

enve recplot

Description

Produces recruitment plots provided that BlastTab.catsbj.pl has been previously executed. Requires the gplots library.

Usage

```
enve.recplot(prefix, id.min = NULL, id.max = NULL, id.binsize = NULL,
             id.splines = 0, id.metric = "id", id.summary = "sum", pos.min = 1,
             pos.max = NULL, pos.binsize = 1000, pos.splines = 0, rec.col1 = "white",
             rec.col2 = "black", main = NULL, contig.col = grey(0.85),
             ret.recplot = FALSE, ret.hist = FALSE, ret.mode = FALSE,
             id.cutoff = NULL, verbose = TRUE, ...)
```

Arguments

prefix	Path to the prefix of the BlastTab.catsbj.pl output files. At least the files .rec and .lim must exist with this prefix.
id.min	Minimum identity to be considered. By default, the minimum detected identity. This value is a percentage.
id.max	Maximum identity to be considered. By default, 100.
id.binsize	Size of the identity bins (vertical histograms). By default, 0.1 for identity metrics and 5 for bit score.
id.splines	Smoothing parameter for the splines in the identity histogram. Zero (0) for no splines. A generally good value is 1/2. If non-zero, requires the stats package.
id.metric	Metric of identity to be used (Y-axis). It can be any unambiguous prefix of "identity", "corrected identity", or "bit score".
id.summary	Method used to build the identity histogram (Horizontal axis of the right panel). It can be any unambiguous prefix of "sum", "average", "median", "90% lower bound", "90% upper bound", "95% lower bound", and "95% upper bound". The last four options correspond to the upper and lower boundaries of the 90% and 95% empirical confidence intervals.
pos.min	Minimum (leftmost) position in the reference (concatenated) genome (in bp).
pos.max	Maximum (rightmost) position in the reference (concatenated) genome (in bp). By default: Length of the genome.
pos.binsize	Size of the position bins (horizontal histograms) in bp.
pos.splines	Smoothing parameter for the splines in the position histogram. Zero (0) for no splines. If non-zero, requires the stats package.
rec.col1	Lightest color in the recruitment plot.
rec.col2	Darkest color in the recruitment plot.

<code>main</code>	Title of the plot.
<code>contig.col</code>	Color of the Contig boundaries. Set to NA to ignore Contig boundaries.
<code>ret.recplot</code>	Indicates if the matrix of the recruitment plot is to be returned.
<code>ret.hist</code>	Indicates if the vectors of the identity and position histograms are to be returned.
<code>ret.mode</code>	Indicates if the mode of the identity is to be computed. It requires the <code>modeest</code> package.
<code>id.cutoff</code>	Minimum identity to consider an alignment as "top". By default, it is 0.95 for the identity metrics and 95% of the best scoring alignment for bit score.
<code>verbose</code>	Indicates if the function should report the advance.
<code>...</code>	Any additional graphic parameters to be passed to <code>plot</code> for all panels except the recruitment plot (lower-left).

Value

A list with the following elements:

`pos.marks`: Midpoints of the position histogram.

`id.matrix`: Midpoints of the identity histogram.

`recplot` (if `ret.recplot=TRUE`): Matrix containing the recruitment plot values.

`id.mean`: Mean identity.

`id.median`: Median identity.

`id.mode` (if `ret.mode=TRUE`): Mode of the identity.

`id.hist` (if `ret.hist=TRUE`): Values of the identity histogram.

`pos.hist.low` (if `ret.hist=TRUE`): Values of the position histogram (depth) with "low" identity (i.e., below `id.cutoff`).

`pos.hist.top` (if `ret.hist=TRUE`): Values of the position histogram (depth) with "top" identity (i.e., above `id.cutoff`).

`id.max`: Value of `id.max`. This is returned because `id.max=NULL` may vary.

`id.cutoff`: Value of `id.cutoff`. This is returned because `id.cutoff=NULL` may vary.

`seqdepth.mean.top`: Average sequencing depth with identity above `id.cutoff`.

`seqdepth.mean.low`: Average sequencing depth with identity below `id.cutoff`.

`seqdepth.mean.all`: Average sequencing depth without identity filtering.

`seqdepth.median.top`: Median sequencing depth with identity above `id.cutoff`.

`seqdepth.median.low`: Median sequencing depth with identity below `id.cutoff`.

`seqdepth.median.all`: Median sequencing depth without identity filtering.

`id.metric`: Full name of the used identity metric.

`id.summary`: Full name of the summary method used to build the identity plot.

Author(s)

Luis M. Rodriguez-R [aut, cre]

enve.recplot2 *enve recplot2*

Description

Produces recruitment plots provided that BlastTab.catsbj.pl has been previously executed.

Usage

```
enve.recplot2(prefix, plot = TRUE, pos.breaks = 1000, id.breaks = 300,
  id.metric = c("identity", "corrected identity", "bit score"),
  id.summary = sum, id.cutoff = 95, threads = 2, verbose = TRUE,
  ...)
```

Arguments

prefix	Path to the prefix of the BlastTab.catsbj.pl output files. At least the files .rec and .lim must exist with this prefix.
plot	Should the object be plotted?
pos.breaks	Breaks in the positions histogram. It can also be a vector of break points, and values outside the range are ignored. If zero (0), it uses the sequence breaks as defined in the .lim file, which means one bin per contig (or gene, if the mapping is against genes).
id.breaks	Breaks in the identity histogram. It can also be a vector of break points, and values outside the range are ignored.
id.metric	Metric of identity to be used (Y-axis). Corrected identity is only supported if the original BLAST file included sequence lengths.
id.summary	Function summarizing the identity bins. Other recommended options include: 'median' to estimate the median instead of total bins, and 'function(x) mlv(x,method='parzen')\$M' to estimate the mode.
id.cutoff	Cutoff of identity metric above which the hits are considered 'in-group'. The 95% identity corresponds to the expectation of ANI<95% within species.
threads	Number of threads to use.
verbose	Indicates if the function should report the advance.
...	Any additional parameters supported by 'plot.enve.RecPlot2'.

Value

Returns an object of class 'enve.RecPlot2'.

Author(s)

Luis M. Rodriguez-R [aut, cre]

enve.RecPlot2-class *enve.RecPlot2 S4 class*

Description

Enve-omics representation of Recruitment plots. This object can be produced by ‘enve.recplot2’ and supports S4 method plot.

Objects from the Class

Objects can be created by calls of the form `new(enve.RecPlot2 ...)`

Slots

`counts:` (matrix) Counts as a two-dimensional histogram.
`pos.counts.in:` (numeric) Counts of in-group hits per position bin.
`pos.counts.out:` (numeric) Counts of out-group hits per position bin.
`id.counts:` (numeric) Counts per ID bin.
`id.breaks:` (numeric) Breaks of identity bins.
`pos.breaks:` (numeric) Breaks of position bins.
`seq.breaks:` (numeric) Limits of the subject sequences after concatenation.
`seq.names:` (character) Names of the subject sequences.
`id.metric:` (character) Metric used as ‘identity’.
`id.ingroup:` (logical) Identity bins considered in-group.
`call:` (call) Call producing this object.

Methods

`$ signature(x = "enve.RecPlot2"): ...`

Author(s)

Luis M. Rodriguez-R [aut, cre]

`enve.recplot2.corePeak`*enve recplot2 corePeak*

Description

Finds the peak in a list of peaks that is most likely to represent the "core genome" of a population.

Usage`enve.recplot2.corePeak(x)`**Arguments**

`x` 'list' of 'enve.RecPlot2.Peak' objects.

Author(s)

Luis M. Rodriguez-R [aut, cre]

`enve.recplot2.findPeaks`*enve recplot2 findPeaks*

Description

Identifies peaks in the population histogram potentially indicating sub-population mixtures.

Usage

```
enve.recplot2.findPeaks(x, min.points = 50, quant.est = c(0.005,
  0.995), mlv.opts = list(method = "parzen"), fitdist.opts.sn = list(distr = "sn",
  method = "qme", probs = c(0.1, 0.5, 0.8), start = list(omega = 1,
  alpha = -1), lower = c(1e-06, -Inf, 0), upper = c(Inf,
  0, Inf)), fitdist.opts.norm = list(distr = "norm", method = "qme",
  probs = c(0.4, 0.6), start = list(sd = 1), lower = c(1e-08,
  0)), rm.top = 0.05, with.skewness = TRUE, optim.rounds = 200,
  optim.epsilon = 1e-08, merge.logdist = log(1.25), verbose = FALSE)
```

Arguments

<code>x</code>	An 'enve.RecPlot2' object.
<code>min.points</code>	Minimum number of points in the quantile-estimation-range ('quant.est') to estimate a peak.
<code>quant.est</code>	Range of quantiles to be used in the estimation of a peak's parameters.
<code>mlv.opts</code>	Options passed to 'mlv' to estimate the mode.
<code>fitdist.opts.sn</code>	Options passed to 'fitdist' to estimate the standard deviation if with.skewness=TRUE. Note that the 'start' parameter will be ammended with xi=estimated mode for each peak.
<code>fitdist.opts.norm</code>	Options passed to 'fitdist' to estimate the standard deviation if with.skewness=FALSE. Note that the 'start' parameter will be ammended with mean=estimated mode for each peak.
<code>rm.top</code>	Top-values to remove before finding peaks, as a quantile probability. This step is useful to remove highly conserved regions, but can be turned off by setting <code>rm.top=0</code> . The quantile is determined <i>after</i> removing zero-coverage windows.
<code>with.skewness</code>	Allow skewness correction of the peaks. Typically, the sequencing-depth distribution for a single peak is left-skewed, due partly (but not exclusively) to fragmentation and mapping sensitivity. See Lindner et al 2013, <i>Bioinformatics</i> 29(10):1260-7 for an alternative solution for the first problem (fragmentation) called "tail distribution".
<code>optim.rounds</code>	Maximum rounds of peak optimization.
<code>optim.epsilon</code>	Trace change at which optimization stops (unless 'optim.rounds' is reached first). The trace change is estimated as the sum of square differences between parameters in one round and those from two rounds earlier (to avoid infinite loops from approximation).
<code>merge.logdist</code>	Maximum value of \log -ratio between centrality parameters in peaks to attempt merging. The default of ~ 0.22 corresponds to a maximum difference of 25%.
<code>verbose</code>	Display (mostly debugging) information.

Value

Returns a list of 'enve.RecPlot2.Peak' objects.

Author(s)

Luis M. Rodriguez-R [aut, cre]

enve.RecPlot2.Peak-class

enve.RecPlot2.Peak S4 class

Description

Enve-omics representation of a peak in the sequencing depth histogram of a Recruitment plot (see 'enve.recplot2.findPeaks').

Objects from the Class

Objects can be created by calls of the form `new(enve.RecPlot2.Peak ...)`

Slots

dist: (character) Distribution of the peak. Currently supported: 'norm' (normal) and 'sn' (skew-normal).

values: (numeric) Sequencing depth values predicted to conform the peak.

values.res: (numeric) Sequencing depth values not explained by this or previously identified peaks.

mode: (numeric) Seed-value of mode anchoring the peak.

param.hat: (list) Parameters of the distribution. A list of two values if `dist='norm'` (sd and mean), or three values if `dist='sn'` (`omega=scale`, `alpha=shape`, and `xi=location`). Note that the "dispersion" parameter is always first and the "location" parameter is always last.

n.hat: (numeric) Number of bins estimated to be explained by this peak. This should ideally be equal to the length of 'values', but it's not and integer.

n.total: (numeric) Total number of bins from which the peak was extracted.

err.res: (numeric) Error left after adding the peak.

merge.logdist: (numeric) Attempted 'merge.logdist' parameter.

Methods

`$ signature(x = "enve.RecPlot2.Peak"): ...`

Author(s)

Luis M. Rodriguez-R [aut, cre]

enve.recplot2.__counts
enve recplot2 counts

Description

Internal ancilliary function (see 'enve.recplot2').

Usage

```
enve.recplot2.__counts(x, pos.breaks, id.breaks, rec.idcol)
```

Arguments

x
pos.breaks
id.breaks
rec.idcol

Author(s)

Luis M. Rodriguez-R [aut, cre]

enve.recplot2.__findPeak
enve recplot2 findPeak

Description

Internall ancilliary function (see 'enve.recplot2.findPeaks').

Usage

```
enve.recplot2.__findPeak(lsd1, min.points, quant.est, mlv.opts,  
  fitdist.opts, with.skewness, optim.rounds, optim.epsilon,  
  n.total, merge.logdist, verbose)
```

Arguments

lsd1
min.points
quant.est
mlv.opts
fitdist.opts
with.skewness
optim.rounds
optim.epsilon
n.total
merge.logdist
verbose

Author(s)

Luis M. Rodriguez-R [aut, cre]

enve.recplot2.__findPeaks
enve recplot2 findPeaks

Description

Internal ancilliary function (see 'enve.recplot2.findPeaks').

Usage

```
enve.recplot2.__findPeaks(peaks.opts)
```

Arguments

peaks.opts

Author(s)

Luis M. Rodriguez-R [aut, cre]

`enve.recplot2.__peakHist`
enve recplot2 peakHist

Description

Internal ancilliary function (see ‘enve.RecPlot2.Peak’).

Usage

`enve.recplot2.__peakHist(x, mids, counts = TRUE)`

Arguments

x
mids
counts

Author(s)

Luis M. Rodriguez-R [aut, cre]

`enve.recplot2.__whichClosestPeak`
enve recplot2 whichClosestPeak

Description

Internal ancilliary function (see ‘enve.recplot2.findPeaks’).

Usage

`enve.recplot2.__whichClosestPeak(peak, peaks)`

Arguments

peak
peaks

Author(s)

Luis M. Rodriguez-R [aut, cre]

enve.tribs

enve tribs

Description

Subsample any objects in "distance space" to reduce the effect of sample-clustering. This function was originally designed to subsample genomes in "phylogenetic distance space", a clear case of strong clustering bias in sampling, by Luis M. Rodriguez-R and Michael R Weigand.

Usage

```
enve.tribs(dist, selection = labels(dist), replicates = 1000,
  summary.fx = median, dist.method = "euclidean", subsamples = seq(0,
    1, by = 0.01), dimensions = ceiling(length(selection) *
    0.05), metaMDS.opts = list(), threads = 2, verbosity = 1,
  points, pre.tribs)
```

Arguments

dist	Distances as a 'dist' object.
selection	Objects to include in the subsample. By default, all objects are selected.
replicates	Number of replications per point
summary.fx	Function to summarize the distance distributions in a given replicate. By default, the median distance is estimated.
dist.method	Distance method between random points and samples in the transformed space. See 'dist'.
subsamples	Subsampling fractions
dimensions	Dimensions to use in the NMDS. By default, 5% of the selection length.
metaMDS.opts	Any additional options to pass to metaMDS, as 'list'.
threads	Number of threads to use.
verbosity	Verbosity. Use 0 to run quietly, increase for additional information.
points	Optional. If passed, the MDS step is skipped and this object is used instead. It can be the '\$points' slot of class 'metaMDS' (from 'vegan'). It must be a matrix or matrix-coercible object, with samples as rows and dimensions as columns.
pre.tribs	Optional. If passed, the points are recovered from this object (except if 'points' is also passed. This should be an 'enve.TRIBS' object estimated on the same objects (the selection is unimportant).

Value

Returns an 'enve.TRIBS' object.

Author(s)

Luis M. Rodriguez-R [aut, cre]

`enve.TRIBS-class``enve.TRIBS S4 class`

Description

Enve-omics representation of "Transformed-space Resampling In Biased Sets (TRIBS)". This object represents sets of distances between objects, sampled nearly-uniformly at random in "distance space". Subsampling without selection is trivial, since both the distances space and the selection occur in the same transformed space. However, it's useful to compare randomly subsampled sets against a selected set of objects. This is intended to identify overdispersion or overclustering (see `'enve.TRIBStest'`) of a subset against the entire collection of objects with minimum impact of sampling biases. This object can be produced by `'enve.tribs'` and supports S4 methods `'plot'` and `'summary'`.

Objects from the Class

Objects can be created by calls of the form `new(enve.TRIBS . . .)`

Slots

`distance:` (numeric) Centrality measurement of the distances between the « selected objects (without subsampling).

`points:` (matrix) Position of the different objects in distance « space.

`distances:` (matrix) Subsampled distances, where the rows are replicates « and the columns are subsampling levels.

`spaceSize:` (numeric) Number of objects.

`selSize:` (numeric) Number of selected objects.

`dimensions:` (numeric) Number of dimensions in the distance space.

`subsamples:` (numeric) Subsampling levels (as fractions, from 0 to 1).

`call:` (call) Call producing this object.

Methods

No methods defined with class "enve.TRIBS" in the signature.

Author(s)

Luis M. Rodriguez-R [aut, cre]

```
enve.TRIBS.merge      enve TRIBS merge
```

Description

Merges two ‘enve.TRIBS’ objects generated from the same objects at different subsampling levels.

Usage

```
enve.TRIBS.merge(x, y)
```

Arguments

x	First ‘enve.TRIBS’ object.
y	Second ‘enve.TRIBS’ object.

Value

Returns an ‘enve.TRIBS’ object.

Author(s)

Luis M. Rodriguez-R [aut, cre]

```
enve.tribs.test      enve tribs test
```

Description

Estimates the empirical difference between all the distances in a set of objects and a subset, together with its statistical significance.

Usage

```
enve.tribs.test(dist, selection, bins = 50, ...)
```

Arguments

dist	Distances as ‘dist’ object.
selection	Selection defining the subset.
bins	Number of bins to evaluate in the range of distances.
...	Any other parameters supported by ‘enve.tribs’, except ‘subsamples’.

Value

Returns an ‘enve.TRIBStest’ object.

Author(s)

Luis M. Rodriguez-R [aut, cre]

enve.TRIBStest-class *enve.TRIBStest S4 class*

Description

Test of significance of overclustering or overdispersion in a selected set of objects with respect to the entire set (see 'enve.TRIBS'). This object can be produced by 'enve.tribs.test' and supports S4 methods 'plot' and 'summary'.

Objects from the Class

Objects can be created by calls of the form `new(enve.TRIBStest ...)`

Slots

`pval.gt`: (numeric) P-value for the overdispersion test.

`pval.lt`: (numeric) P-value for the overclustering test.

`all.dist`: (numeric) Empiric PDF of distances for the entire dataset (subsampled at selection size).

`sel.dist`: (numeric) Empiric PDF of distances for the selected objects (without subsampling).

`diff.dist`: (numeric) Empiric PDF of the difference between 'all.dist' and 'sel.dist'. The p-values are estimating by comparing areas in this PDF greater than and lesser than zero.

`dist.mids`: (numeric) Midpoints of the empiric PDFs of distances.

`diff.mids`: (numeric) Midpoints of the empiric PDF of difference of distances.

`call`: (call) Call producing this object.

Methods

No methods defined with class "enve.TRIBStest" in the signature.

Author(s)

Luis M. Rodriguez-R [aut, cre]

enve.__prune.iter *enve prune iter*

Description

Internal function for enve.prune.dist

Usage

```
enve.__prune.iter(t, dist, min_dist, quiet)
```

Arguments

t
dist
min_dist
quiet

Author(s)

Luis M. Rodriguez-R [aut, cre]

enve.__prune.reduce *enve prune reduce*

Description

Internal function for enve.prune.dist

Usage

```
enve.__prune.reduce(t, nodes, min_dist, quiet)
```

Arguments

t
nodes
min_dist
quiet

Author(s)

Luis M. Rodriguez-R [aut, cre]

enve.__tribs	<i>enve tribs</i>
--------------	-------------------

Description

Internal ancilliary function (see 'enve.tribs').

Usage

```
enve.__tribs(rep, frx, selection, dimensions, dots, dist.method,  
             summary.fx, dist)
```

Arguments

rep
frx
selection
dimensions
dots
dist.method
summary.fx
dist

Author(s)

Luis M. Rodriguez-R [aut, cre]

phyla.counts	<i>Counts of microbial phyla in four sites</i>
--------------	--

Description

This data set gives the counts of phyla in three different sites.

Usage

```
phyla.counts
```

Format

A data frame with 9 rows (phyla) and 4 rows (sites).

plot.enve.RecPlot2 *plot enve RecPlot2*

Description

Plots an ‘enve.RecPlot2’ object.

Usage

```
## S3 method for class 'enve.RecPlot2'
plot(x, layout = matrix(c(5, 5, 2, 1, 4, 3), nrow = 2),
     widths = c(1, 7, 2), heights = c(1, 2), palette = grey((100:0)/100),
     underlay.group = TRUE, peaks.col = "darkred", id.lim = range(x$id.breaks),
     pos.lim = range(x$pos.breaks), pos.units = c("Mbp", "Kbp",
     "bp"), mar = list(`1` = c(5, 4, 1, 1) + 0.1, `2` = c(ifelse(any(layout ==
     1), 1, 5), 4, 4, 1) + 0.1, `3` = c(5, ifelse(any(layout ==
     1), 1, 4), 1, 2) + 0.1, `4` = c(ifelse(any(layout ==
     1), 1, 5), ifelse(any(layout == 2), 1, 4), 4, 2) + 0.1,
     `5` = c(5, 3, 4, 1) + 0.1, `6` = c(5, 4, 4, 2) + 0.1),
     pos.splines = 0, id.splines = 1/2, in.lwd = ifelse(pos.splines >
     0, 1/2, 2), out.lwd = ifelse(pos.splines > 0, 1/2, 2),
     id.lwd = ifelse(id.splines > 0, 1/2, 2), in.col = "darkblue",
     out.col = "lightblue", id.col = "black", breaks.col = "#AAAAAA40",
     peaks.opts = list(), ...)
```

Arguments

x	‘enve.RecPlot2’ object to plot.
layout	Matrix indicating the position of the different panels in the layout, where: 0: Empty space, 1: Counts matrix, 2: position histogram (sequencing depth), 3: identity histogram, 4: Populations histogram (histogram of sequencing depths), 5: Color scale for the counts matrix (vertical), 6: Color scale of the counts matrix (horizontal). Only panels indicated here will be plotted. To plot only one panel simply set this to the number of the panel you want to plot.
widths	Relative widths of the columns of ‘layout’.
heights	Relative heights of the rows of ‘layout’.
palette	Colors to be used to represent the counts matrix, sorted from no hits to the maximum sequencing depth.
underlay.group	If TRUE, it indicates the in-group and out-group areas coloured based on ‘in.col’ and ‘out.col’. Requires support for semi-transparency.
peaks.col	If not NA, it attempts to represent peaks in the population histogram in the specified color. Set to NA to avoid peak-finding.
id.lim	Limits of identities to represent.
pos.lim	Limits of positions to represent (in bp, regardless of ‘pos.units’).

pos.units	Units in which the positions should be represented (powers of 1,000 base pairs).
mar	
pos.splines	Smoothing parameter for the splines in the position histogram. Zero (0) for no splines. If non-zero, requires the stats package.
id.splines	Smoothing parameter for the splines in the identity histogram. Zero (0) for no splines. If non-zero, requires the stats package.
in.lwd	Line width for the sequencing depth of in-group matches.
out.lwd	Line width for the sequencing depth of out-group matches.
id.lwd	Line width for the identity histogram.
in.col	Color associated to in-group matches.
out.col	Color associated to out-group matches.
id.col	Color for the identity histogram.
breaks.col	Color of the vertical lines indicating sequence breaks.
peaks.opts	Options passed to 'enve.recplot2.findPeaks', if 'peaks.col' is not NA.
...	Any other graphic parameters (currently ignored).

Value

Returns a list of 'enve.RecPlot2.Peak' objects (see 'enve.recplot2.findPeaks'). If 'peaks.col'=NA or 'layout' doesn't include 4, returns NA.

Author(s)

Luis M. Rodriguez-R [aut, cre]

plot.enve.TRIBS *plot enve TRIBS*

Description

Plot an 'enve.TRIBS' object.

Usage

```
## S3 method for class 'enve.TRIBS'
plot(x, new = TRUE, type = c("boxplot", "points"),
     col = "#00000044", pt.cex = 1/2, pt.pch = 19, pt.col = col,
     ln.col = col, ...)
```

Arguments

x	'enve.TRIBS' object to plot.
new	Should a new canvas be drawn?
type	Type of plot. The 'points' plot shows all the replicates, the 'boxplot' plot represents the values found by 'boxplot.stats' as areas, and plots the outliers as points.
col	Color of the areas and/or the points.
pt.cex	Size of the points.
pt.pch	Points character.
pt.col	Color of the points.
ln.col	Color of the lines.
...	Any additional parameters supported by 'plot'.

Author(s)

Luis M. Rodriguez-R [aut, cre]

plot.enve.TRIBStest *plot enve TRIBStest*

Description

Plots an 'enve.TRIBStest' object.

Usage

```
## S3 method for class 'enve.TRIBStest'
plot(x, type = c("overlap", "difference"), col = "#00000044",
      col1 = col, col2 = "#44001144", ylab = "Probability", xlim = range(attr(x,
      "dist.mids")), ylim = c(0, max(c(attr(x, "all.dist"),
      attr(x, "sel.dist")))), ...)
```

Arguments

x	'enve.TRIBStest' object to plot.
type	What to plot. 'overlap' generates a plot of the two contrasting empirical PDFs (to compare against each other), 'difference' produces a plot of the differences between the empirical PDFs (to compare against zero).
col	Main color of the plot if type='difference'.
col1	First color of the plot if type='overlap'.
col2	Second color of the plot if type='overlap'.
ylab	Y-axis label.
xlim	X-axis limits.
ylim	Y-axis limits.
...	Any other graphical arguments.

Author(s)

Luis M. Rodriguez-R [aut, cre]

summary.enve.TRIBS *summary enve TRIBS*

Description

Summary of an 'enve.TRIBS' object.

Usage

```
## S3 method for class 'enve.TRIBS'
summary(object, ...)
```

Arguments

object 'enve.TRIBS' object.
 ... No additional parameters are currently supported.

Author(s)

Luis M. Rodriguez-R [aut, cre]

summary.enve.TRIBStest *summary enve TRIBStest*

Description

Summary of an 'enve.TRIBStest' object.

Usage

```
## S3 method for class 'enve.TRIBStest'
summary(object, ...)
```

Arguments

object 'enve.TRIBStest' object.
 ... No additional parameters are currently supported.

Author(s)

Luis M. Rodriguez-R [aut, cre]

\$-methods

~~ Methods for Function \$ ~~

Description

~~ Methods for function \$ ~~

Methods

signature(x = "enve.RecPlot2")

signature(x = "enve.RecPlot2.Peak")

Index

*Topic **textasciitilde\textasciitilde**
other possible keyword(s)
\textasciitilde\textasciitilde
 \$-methods, 28

*Topic **datasets**
 phyla.counts, 23

*Topic **methods**
 \$-methods, 28

*Topic **package**
 enveomics.R-package, 2

\$, enve.RecPlot2-method (\$-methods), 28

\$, enve.RecPlot2.Peak-method
 (\$-methods), 28

\$-methods, 28

enve.__prune.iter, 22

enve.__prune.reduce, 22

enve.__tribs, 23

enve.barplot, 2

enve.cliopts, 4

enve.df2dist, 5

enve.df2dist.group, 5

enve.df2dist.list, 6

enve.prune.dist, 7

enve.recplot, 8

enve.RecPlot2 (enve.RecPlot2-class), 11

enve.recplot2, 10

enve.RecPlot2-class, 11

enve.recplot2.__counts, 15

enve.recplot2.__findPeak, 15

enve.recplot2.__findPeaks, 16

enve.recplot2.__peakHist, 17

enve.recplot2.__whichClosestPeak, 17

enve.recplot2.corePeak, 12

enve.recplot2.findPeaks, 12

enve.RecPlot2.Peak
 (enve.RecPlot2.Peak-class), 14

enve.RecPlot2.Peak-class, 14

enve.TRIBS (enve.TRIBS-class), 19

enve.tribs, 18

enve.TRIBS-class, 19

enve.TRIBS.merge, 20

enve.tribs.test, 20

enve.TRIBStest (enve.TRIBStest-class),
 21

enve.TRIBStest-class, 21

enveomics.R (enveomics.R-package), 2

enveomics.R-package, 2

phyla.counts, 23

plot.enve.RecPlot2, 24

plot.enve.TRIBS, 25

plot.enve.TRIBStest, 26

summary.enve.TRIBS, 27

summary.enve.TRIBStest, 27