Package 'LeyLabRMisc'

April 3, 2020

Type Package

Version 0.1.4

Author Nick Youngblut

Title Ley Lab misc R functions, rmd templates, etc.

Maintainer Nick Youngblut <nyoungb2@gmail.com>

Description L	Lab misc R functions, rmd templates, etc.	
License GPL-		
Encoding UT		
J		
LazyData true		
RoxygenNote	.0	
R topics d	umented:	
.well.	_index	2
.well	ndex	3
as.Nu		3
bash_		4
calcu	_rarefaction_curves	4
calc_	na_div	5
calc_	ı_div	5
calc_	oA	6
cat_fi		7
conda	0	7
df.dir		7
dfhea		8
dist_f	nat	8
estim	_rarified_richness	9
estim	richness_phy	9
extra	ltdt	10
fig_u		10
files_	ist	11
Freac		11
hello		12

2 .well384_index

																					28
١	write_table .		•									•		 		•					27
١	well2index																				26
ι	ınique_n													 							26
t	idy_pcoa																				25
S	summary_x .																				24
S	nakemakeInfo	О																			24
S	size_objects .																				23
S	send_email .													 							23
r	ow_sums													 							22
r	ow_means .													 							22
I	_ Robj_md5sum	ı .																			22
	ead_bracken																				21
	save_obj																				20
	Plot																				20
	pipelineInfo .																				19
	hyloseq2df.																				19
•	oath get label																				18
	o.dims																				18
	overlap																			•	17
	nake_dir																				17
	ist files																				
	tol_simplebar tol_symbol .																				16
	_																			•	14 15
																					14
	tol_externalsh	-																			
	tol_colorstrip																				

.well384_index

making 384-well plate index

Description

making 384-well plate index

Usage

.well384_index()

Value

named vector (Well -> location); column-wise location

.well96_index 3

 $.well96_index$

making a 96-well plate index

Description

making a 96-well plate index

Usage

```
.well96_index()
```

Value

named vector (Well -> location); column-wise location

as.Num

convert to numeric while avoiding factor conversion issues

Description

convert to numeric while avoiding factor conversion issues

Usage

```
as.Num(x)
```

Arguments

Х

an interable

Value

a numeric object

bash_job	bash	iob	using	conda	env
54511	Cusii	Ju	usung	contact	Citt

Description

The conda setup is assumed to be in your ~/.bashrc If print_output == TRUE: the stdout/stderr will be printed instead of returned Else: the stdout/stderr with be returned by the function stderr/stdout is printed unless print_output==FALSE

Usage

```
bash_job(cmd, conda_env, stdout = TRUE, stderr = TRUE, print_output = TRUE)
```

Arguments

cmd The bash command in a string format

conda_env The conda env to use

stdout Print the stdout from the command? stderr Print the stderr from the command?

quiet No printing

```
calculate_rarefaction_curves
```

Function for rarefaction analysis

Description

Running estimate_richness_phy() at multiple subsampling depths

Usage

```
calculate_rarefaction_curves(psdata, measures, depths, parallel = FALSE)
```

Arguments

psdata phyloseq object

measures Which diversity measures (see vegan package)

depths Which sequencing depths? Example: c(10, 100, 1000)

Value

A dataframe

calc_alpha_div 5

calc_alpha_div Calculate common alpha-diversity metrics

Description

Faith's Phylogenetic Diversity ("PD") can be calculated only if a tree is provided. The tree can have extra tips, but there must be tip labels for all taxa in the provided table.

Usage

```
calc_alpha_div(df, tree = NULL, index = c("nobs", "shannon", "PD"))
```

Arguments

df sample x taxon abundance table (usual format for vegan)

tree tree with tips matching taxa in the abundance table (only needed for PD)

index which of the indices to calculate? (nobs = no. of observations, shannon = Shan-

non Index, PD = Faith's PD)

Value

a data.frame of alpha diversity values (and sample names)

calc_beta_div	beta-diversity calculation	
---------------	----------------------------	--

Description

A wrapper around vegan::vegdist and rbiom (rbiom used for UniFrac calculations). For unifrac: "wunifrac" = weighted unifrac, "unifrac" = unweighted unifrac. The function returns a tidy dataframe of PCoA axes (PC1 & PC2), percent variance explained for each PC.

Usage

```
calc_beta_div(
   df,
   tree = NULL,
method = c("wunifrac", "unifrac", "manhattan", "euclidean", "canberra", "clark",
   "bray", "kulczynski", "jaccard", "gower", "altGower", "morisita", "horn", "mountford",
        "raup", "binomial", "chao", "cao", "mahalanobis"),
   threads = 1
)
```

6 calc_PCoA

Arguments

df sample x taxon dataframe. Colnames (taxa) must match the tree tip labels if the

tree is provided

tree phylogeny with tips matching the df colnames (only needed for wunifrac &

unifrac methods)

method distance method (vegdist distances; wunifrac=Weighted Unifrac; unifrac=Unweighted

Unifrac)

threads used for UniFrac calculations with rbiom

Details

Unifrac is calculated with the https://github.com/cmmr/rbiom package (requires bioconductor packages).

If the goal is PCoA, then see the "tidy_PCoA" function.

Value

data.frame

calc_PCoA Wrapper for cmdscale

Description

Simple wrapper for cmdscale to provide data.frame formatted table. If the distance matrices contain NAs, the samples containing NAs will be removed (with a warning).

Usage

```
calc_PCoA(dist_mtx, k = 2)
```

Arguments

dist_mtx distance matrix object

Value

data.frame

cat_file 7

cat_file

pretty printing of a text file via cat

Description

This is most useful for working with IRkernl in Jupyter notebooks

Usage

```
cat_file(file_name)
```

Arguments

file_name

the name of the file to print

condaInfo

"conda list" in R

Description

This is most useful for working with IRkernl in Jupyter notebooks

Usage

```
condaInfo(conda_env)
```

Arguments

conda_env

The name of the conda env to list

df.dims

Changing number of rows/columns shown when printing a data frame

Description

This is most useful for working with IRkernl in Jupyter notebooks

Usage

```
df.dims(nrows = 4, ncols = 20)
```

Arguments

nrows number of rows to print ncols number of columns to print

8 dist_format

dfhead

A simple dataframe summary

Description

A simple dataframe summary

Usage

```
dfhead(df, n = 3)
```

Arguments

df dataframe object

n Number of lines to print

Value

a dataframe object

dist_format

creating a string with distance & percent explained

Description

creating a string with distance & percent explained

Usage

```
dist_format(dist, PC1_perc_exp, PC2_perc_exp, label1 = 1, label2 = 2)
```

Arguments

dist str, distance metric

PC1_perc_exp float, percent variance explained for PC1
PC2_perc_exp float, percent variance explained for PC2

label1 First PC labellabel2 Seconda PC label

Value

```
str, formatted as "metric, <PC1_perc_exp>
```

estimate_rarified_richness

Helper Function for rarefaction analysis

Description

Helper Function for rarefaction analysis

Usage

```
estimate_rarified_richness(psdata, measures, depth)
```

Arguments

psdata phyloseq object

measures Which diversity measures depth The sampling depth

Value

molten alpha diversity object

 $\verb|estimate_richness_phy|| phyloseq::estimate_richness, but includes Faith's PD|$

Description

See physeq::estimate richness for full details

Usage

```
estimate_richness_phy(physeq, split = TRUE, measures = NULL)
```

Arguments

physeq Phyloseq object

split Splitting the OTU table

measures Which diversity measures (Faith's PD = "FaithPD)

Value

Dataframe can calculate faith's PD (using Picante, "FaithPD")

fig_uuid

extract_pltdt

Extract data from ggplot object

Description

The data is written to files

Usage

```
extract_pltdt(plot_object, output_path)
```

Arguments

plot_object A ggplot object

output_path Where to write the output

fig_uuid

create UUID for figure file name

Description

create UUID for figure file name

Usage

```
fig_uuid(full = FALSE)
```

Arguments

full

Full length uuid or trimmed to just 24 char?

Value

character object

files_to_list

			- • ·
fil	65	tο	list

convert a vector of file paths into a named list

Description

convert a vector of file paths into a named list

Usage

```
files_to_list(files, label_index = -1)
```

Arguments

files Vector of file paths (eg., by using "list_files()")

label_index Which item in the path to return? 1-indexing. If <1, samples selected from the

end.

Examples

```
files = c('/path/to/project/Sample1/table.txt', '/path/to/project/Sample2/table.txt') \\ files_to_list(files, -1) \\ files = c('/path/to/project/Sample1.txt', '/path/to/project/Sample2.txt') \\ files_to_list(files, 0) \\
```

Fread

Simple wrapper around data.table::fread

Description

Simple wrapper around data.table::fread

Usage

```
Fread(infile = NULL, cmd = NULL, sep = "\t", check.names = TRUE, ...)
```

Arguments

infile input file name

cmd command instead of input file (eg., "gunzip -c INFILE")

sep value delimiter

check.names format check column names
... passed to data.table::fread

Value

data.table

itol_boxplot

hello

Hello, World!

Description

Prints 'Hello, world!'.

Usage

hello()

Examples

hello()

itol_boxplot

create itol boxplot file

Description

https://itol.embl.de/help.cgi#boxplot

Usage

```
itol_boxplot(
   df,
   dataset_label,
   out_file,
   out_dir = NULL,
   key_color = "#ff0000"
)
```

Arguments

df	Dataframe, in which the rownames should correspond with the tree labels; the columns must specify: minimum,q1,median,q3,maximum,extreme_value1,extreme_value2
dataset_label	What to label the itol dataset
out_file	Name of the output file
out_dir	Where to write the output

key_color The color for the legend key

itol_colorstrip 13

itol_colorstrip	create itol colorstrip file
-----------------	-----------------------------

Description

https://itol.embl.de/help.cgi#strip

Usage

```
itol_colorstrip(df, dataset_label, out_file, out_dir = NULL, legend = NULL)
```

Arguments

df Dataframe, in which the rownames should correspond with the tree labels; the

plotting parameter should be column 1

dataset_label What to label the itol dataset
out_file Name of the output file
out_dir Where to write the output
legend Specify particular legend

itol_externalshape

create itol external shape file

Description

https://itol.embl.de/help.cgi#shapes

Usage

```
itol_externalshape(
   df,
   dataset_label,
   out_file,
   out_dir = NULL,
   legend = NULL,
   WIDTH = 200
)
```

Arguments

-I C	D C	1 1.1.1.71	11	24. 41 4 1.1141
df	Dataframe.	in which the rownames st	hould correspond	with the tree labels: other

columns should be values corresponding to symbol size

dataset_label What to label the itol dataset out_file Name of the output file out_dir Where to write the output legend Specify particular legend

14 itol_multibar

i+~1	heatmap
TOT	neatillab

create itol heatmap file

Description

https://itol.embl.de/help.cgi#heatmap

Usage

```
itol_heatmap(
   df,
   dataset_label,
   out_file,
   out_dir = NULL,
   tree = NULL,
   dist_method = "bray",
   color_scheme = c("color", "bw")
)
```

Arguments

df	Dataframe, in which the rownames should correspond with the tree labels; all columns should be numeric values for the heatmap
dataset_label	What to label the itol dataset
out_file	Name of the output file
out_dir	Where to write the output
tree	Tree object used for ordering the heatmap columns; if NULL, the dist_method will be used to create the tree
dist_method	vegan::vegdist method for creating the correlation dendrogram
color_scheme	Heatmap color scheme. color = blue-orange-yellow; bw=white-grey-black

itol_multibar

create itol multi-bar file

Description

https://itol.embl.de/help.cgi#multibar

itol_simplebar 15

Usage

```
itol_multibar(
   df,
   dataset_label,
   out_file,
   out_dir = NULL,
   legend = NULL,
   WIDTH = 200
)
```

Arguments

df Dataframe, in which the rownames should correspond with the tree labels
dataset_label What to label the itol dataset
out_file Name of the output file
out_dir Where to write the output
legend A list that includes shapes, colors, and labels

WIDTH Bar width

itol_simplebar

create itol simple-bar file

Description

https://itol.embl.de/help.cgi#bar

Usage

```
itol_simplebar(
   df,
   dataset_label,
   out_file,
   out_dir = NULL,
   legend = NULL,
   WIDTH = 200
)
```

Arguments

df Dataframe, the rownames should correspond with the tree labels

dataset_label What to label the itol dataset
out_file Name of the output file
out_dir Where to write the output
legend Specify particular legend

WIDTH Bar width

list_files

itol_symbol create itol symbol file

Description

https://itol.embl.de/help.cgi#symbols

Usage

```
itol_symbol(df, dataset_label, out_file, out_dir = NULL, MAXIMUM_SIZE = 50)
```

Arguments

df	Dataframe, in which the rownames should correspond with the tree internal node
	labels, and other columns should be: symbol,size,color,fill,position,(label)
	XXI 1 1 1 1 2 1 1 1

dataset_label What to label the itol dataset
out_file Name of the output file
out_dir Where to write the output
MAXIMUM_SIZE The max size of the symbols

list_files lis

list.files with full.names=TRUE & recursive=TRUE

Description

list.files with full.names=TRUE & recursive=TRUE

Usage

```
list_files(path, pattern = NULL, full.names = TRUE, recursive = TRUE, ...)
```

Arguments

path	a character vector of full	path names; the default corres	ponds to the working
------	----------------------------	--------------------------------	----------------------

directory,

pattern an optional regular expression. Only file names which match the regular expres-

sion will be returned.

full.names a logical value. If TRUE, the directory path is prepended to the file names to

give a relative file path. If FALSE, the file names (rather than paths) are returned

recursive logical. Should the listing recurse into directories?

Value

A character vector containing the names of the files in the specified directories

make_dir

make_dir

A helper function for creating a directory (recursively)

Description

A helper function for creating a directory (recursively)

Usage

```
make_dir(dir, quiet = FALSE)
```

Arguments

```
dir path for the new directory (will create recursively)
quite quite output
```

overlap

Determine counts of setdiff, intersect, & union of 2 vectors (or data.tables)

Description

The output is printed text of intersect, each-way setdiff, and union. Data.table compatible! Just make sure to provide sel_col_x and/or sel_col_y

Usage

```
overlap(
    x,
    y,
    sel_col_x = NULL,
    sel_col_y = NULL,
    to_return = c("counts", "diff_x", "diff_y", "diff_fuzzy")
)
```

Arguments

```
x vector1 or data.table. If data.table, sel_col_x must not be NULL
y vector2 or data.table. If data.table, sel_col_y must not be NULL
sel_col_x If x = data.table, which column to assess?
sel_col_y If y = data.table, which column to assess?
to_return "counts" = print overlap counts; "diff_x-or-y" = return setdiff; "diff_fuzzy" = return closest matches for those that differ (ordered best to worst)
```

path_get_label

	-1	٠		_
р	α	7	m	c
v	ч	_	111	J

Global change of plot size options

Description

This is most useful for working with IRkernl in Jupyter notebooks

Usage

```
p.dims(w = 5, h = 5, res = 200)
```

Arguments

w figure widthh figure height

res figure resolution (DPI)

path_get_label

splitting path and returning just one item in the vector

Description

This is useful for merging tables in which the individual table ID is within the file path.

Usage

```
path_get_label(file_path, index)
```

Arguments

file_path File path(s). If vector or list of paths provided, then a list will be returned

index Which item in the path to return? 1-indexing. If <1, samples selected from the

end. "O" will select the file name.

Value

string if 1 path, else list

phyloseq2df 19

mh.	lose	~っって
nnv	Lose	nzar

Convert a sub-object of a phyloseq object to a dataframe

Description

A helper function for converting OTU, taxonomy, and metadata to dataframes

Usage

```
phyloseq2df(physeq_obj, physeq_func, long = FALSE)
```

Arguments

physeq_obj

The phyloseq object

physeq_func

Which object do you want ('otu_table', 'tax_table', or 'sample_data')

long

Do you want the table in "long" format ("gathered")

Value

A tibble

pipelineInfo

pipeline sessionInfo

Description

sessionInfo for LeyLab snakemake pipelines

Usage

```
pipelineInfo(pipeline_path, head_n = 10)
```

Arguments

pipeline_path

The path to the pipeline directory

head_n

The number of lines to print from the readme

20 qsave_obj

Plot

plot figure and save the figure grob object to a file at the same time

Description

This is most useful for working with IRkernl in Jupyter notebooks

Usage

```
Plot(
   p,
   file = NULL,
   path = NULL,
   suffix = "",
   saveObj = TRUE,
   saveImg = FALSE,
   width = NA,
   height = NA,
   ...
)
```

Arguments

р	Plot object (ggplot2, base, etc)
file	File name to write
path	Path to write to
suffix	File name suffix (eg., '.png')
saveObj	Write the Robj to a file?
saveImg	Write the image to a file?
width	Figure width. If NA, uses global options
height	Figure height. If NA, uses global options

qsave_obj	Simple function for serializing a distance matrix or list of distance
	matrices

Description

Serializing done with the "qs" R package.

Usage

```
qsave_obj(x, file, msg = "Writing file to: ", threads = 1)
```

read_bracken 21

Arguments

x a distance matrix or list of distance matrices

file file name to save to

threads number of threads used for serializing

Value

the input distance matrix or list of distance matrices

read_bracken

Function for reading in a bracken taxonomy table

Description

The table will be converted to long form (sample ~ abundance). Only "_frac" or "_num" columns will be kept (see "keep_frac"). Taxonomy will be split into separate levels (see "tax_levs"). tidytable (w/ data.table) used to speed the process up.

Usage

```
read_bracken(
  infile,
  is_gzip = FALSE,
  n_lines = 0,
  keep_frac = TRUE,
  tax_levs = c("Domain", "Phylum", "Class", "Order", "Family", "Genus", "Species")
)
```

Arguments

infile	Path to bracken table file
is_gzip	Is the table file gzip'ed?

 $n_lines \qquad \qquad Number \ of \ lines \ to \ read. \ If < 1, \ all \ lines \ will \ be \ read.$

keep_frac If TRUE, keep all columns ending in "_frac"; otherwise, keep "_num" columns.

tax_levs Taxonomic levels to separate the taxonomy column into.

Value

data.table

row_sums

Robj_md5sum

Dump an R object as text to a temp file and get the md5sum of the file

Description

Dump an R object as text to a temp file and get the md5sum of the file

Usage

```
Robj_md5sum(Robj)
```

Arguments

Robj

Any R object

Value

md5sum

row_means

rowMeans that works inside a dplyr::mutate() call

Description

rowMeans that works inside a dplyr::mutate() call

Usage

```
row_means(..., na.rm = TRUE)
```

row_sums

rowSums that works inside a dplyr::mutate() call

Description

rowSums that works inside a dplyr::mutate() call

Usage

```
row_sums(..., na.rm = TRUE)
```

send_email 23

send_email

A helper function to send an email via the mail bash cmd

Description

A helper function to send an email via the mail bash cmd

Usage

```
send_email(
  body,
  subject = "R job complete",
  email = NULL,
  email_ext = "tuebingen.mpg.de"
)
```

Arguments

body The email body subject The email subject line

email The email address. If NULL, then username used

email_ext The part after the "at" symbol

Value

The output of the system() call

size_objects

Returns the sizes of R objects

Description

Returns the sizes of R objects

Usage

```
size_objects(Robj)
```

Arguments

Robj

Vector with the names of R objects as characters

Value

A list with the name of R objects as names and the formatted size of the objects

24 summary_x

conda info

anakamaka Infa	an alcom alco
snakemakeInfo	snakemake

Description

snakemake conda info

Usage

```
snakemakeInfo(config_file, pipeline_dir, conda_env)
```

Arguments

config_file The path to the config file

pipeline_dir The path to the pipeline_directory

conda_env The conda env that has snakemake installed

Value

The environment info

summary_	Y

Summary for numeric vectors that includes sd and stderr

Description

```
sd = standard deviation stderr = standard error of the mean (<math>sd(x) / sqrt(length(x)))
```

Usage

```
summary_x(x, label = NULL, sel_col = NULL, rnd = 3)
```

Arguments

x a numeric vector

label row name label for the output. If NULL, then the label will be the input object

label.

sel_col If "x" is data.table or data.frame, which column to assess?

rnd number of digits to round sd and stderr to

Value

a matrix

tidy_pcoa 25

tidy_pcoa

PCoA on a 'long' (tidy) tibble, and a long tibble is returned

Description

Perform PCoA in a "tidy" way. If multiple diversity metrics are provided (eg., "bray" and "jaccard"), all PCoA results will be combined into one data.frame.

Usage

```
tidy_pcoa(
    df,
    taxon_col,
    sample_col,
    abundance_col,
    dists = c("bray", "jaccard", "wunifrac", "unifrac"),
    tree = NULL,
    threads = 1,
    threads_unifrac = 1,
    k = 2,
    dist_mtx_file = NULL,
    pcoa_file = NULL
)
```

Arguments

df	data.frame or tibble
taxon_col	the column specifying taxa or OTUs (no quotes needed)
sample_col	the column specifying sample names (no quotes needed)
abundance_col	the column specifying the taxon abundances in each sample (no quotes needed)
dists	vector of beta-diversity distances ('wunifrac' = weighted UniFrac, 'unifrac' = unweighted Unifrac; see vegan::vegsist for others)
tree	phylogeny for UniFrac calculations. It can have more tips that what is in the data.frame
threads	number of parallel calculations of each distance metric (1 thread per distance)
threads_unifrac	
	number of threads to use for wunifrac & unifrac calculations
k	passed to cmdscale
dist_mtx_file	file name for saving the distance matrices (qs serialization; use ".qs" for the file extension) $\ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \$
pcoa_file	file name for saving the raw pcoa results

26 well2index

Details

Weighted/Unweighted UniFrac is calculated via the rbiom R package. All other beta-diversity metrics are calculated via the vegan R package.

Value

```
a tibble of PCoA info for all selected "dists"
```

unique_n

Pretty print number of unique elements in a vector

Description

The result will be cat'ed to the screen. tidytable compatable. Maje

Usage

```
unique_n(x, label = "items", sel_col = NULL)
```

Arguments

X	a vector or data.table. If data.table, sel_col must not be NULL
label	what to call the items in the vector (eg., "samples")
sel_col	If x is data.table or data.frame, which column to assess?

well2index

Convert between wellID and column-num

Description

Useful for converting between WellIDs (eg., "A2") and well position in a plate (eg., 9)

Usage

```
well2index(x, plate_type = "96-well")
```

Arguments

x A vector of well IDs plate_type Either 96-well or 384-well

Value

A vector of plate positions

write_table 27

write_table writing table convience function
--

Description

This is most useful for working with IRkernl in Jupyter notebooks

Usage

```
write_table(df, file, sep = "\t", quote = FALSE, row.names = FALSE, ...)
```

Arguments

df	Data.frame to write out
file	Output file path
sep	the field separator string. Values within each row of x are separated by this string
quote	a logical value (TRUE or FALSE) or a numeric vector. If TRUE, any character or factor columns will be surrounded by double quotes.
row.names	either a logical value indicating whether the row names of x are to be written along with x, or a character vector of row names to be written.
	Passed to write.table

Index

.well384_index, 2	p.dims, 18
.well96_index, 3	<pre>path_get_label, 18 phyloseq2df, 19</pre>
as.Num,3	pipelineInfo, 19 Plot, 20
bash_job, 4	
calc_alpha_div,5	qsave_obj, 20
calc_beta_div, 5	read_bracken, 21
calc_PCoA, 6	Robj_md5sum, 22
calculate_rarefaction_curves, 4	row_means, 22
cat_file, 7	row_sums, 22
condaInfo, 7	send_email, 23
df.dims,7	size_objects, 23
dfhead, 8	snakemakeInfo, 24
dist_format, 8	summary_x, 24
arst_rormat, o	
estimate_rarified_richness,9	tidy_pcoa, 25
estimate_richness_phy, 9	unique n 26
extract_pltdt, 10	unique_n, 26
firm and 10	well2index, 26
fig_uuid, 10	write_table, 27
files_to_list, 11	
Fread, 11	
hello, 12	
itol_boxplot, 12	
itol_colorstrip, 13	
itol_externalshape, 13	
itol_heatmap, 14	
itol_multibar, 14	
itol_simplebar, 15	
itol_symbol, 16	
list_files, 16	
make_dir, 17	
overlan 17	