# Package 'LeyLabRMisc'

# April 1, 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>

<b>Description</b> Ley Lab misc R functions, rmd templates, etc.
License GPL-3
Encoding UTF-8
LazyData true
RoxygenNote 7.1.0
R topics documented:
.well384 index
.well96_index
as.Num
bash_job
calculate_rarefaction_curves
calc_alpha_div
calc_beta_div
calc_PCoA
cat_file
condaInfo
df.dims
dfhead
dist_format
estimate_rarified_richness
estimate_richness_phy
extract_pltdt
fig_uuid
files_to_list
Fread
hallo 11

2 .well384\_index

	itol_boxplot	12
	itol_colorstrip	12
	itol_externalshape	13
	itol_heatmap	13
	itol_multibar	14
	itol_simplebar	15
	itol_symbol	15
	list_files	16
	make_dir	16
	overlap	17
	p.dims	17
	path_get_label	18
	phyloseq2df	18
	pipelineInfo	19
	Plot	19
	qsave_obj	20
	read_bracken	20
	Robj_md5sum	21
	row_means	21
	row_sums	22
	send_email	22
	size_objects	23
	snakemakeInfo	23
	summary_x	24
	tidy_pcoa	24
	unique_n	25
	well2index	26
	write_table	26
Index		27

.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	alc_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 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
---------------

## 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, method, tree, threads = 1)
```

#### **Arguments**

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

tree is provided

method distance method (vegdist or UniFrac)

tree phylogeny with tips matching the df colnames threads threads used for UniFrac calculations with rbiom

#### Value

data.frame

6 cat\_file

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

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 7

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

dfhead

A simple dataframe summary

# Description

A simple dataframe summary

```
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 label label2 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

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

estimate\_richness\_phy 9

#### **Arguments**

psdata phyloseq object

measures Which diversity measures depth The sampling depth

#### Value

molten alpha diversity object

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")

# 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

files\_to\_list

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

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 11

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

hello

Hello, World!

# Description

Prints 'Hello, world!'.

# Usage

hello()

# **Examples**

hello()

12 itol\_colorstrip

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

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 13

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**

df	Dataframe, in which the rownames should 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

itol\_heatmap

create itol heatmap file

# Description

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

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

14 itol\_multibar

## **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

## 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 15

1101	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

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)
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

16 make\_dir

		_		_	
lis	+	f	i	1	6.5

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 corresponds 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

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 17

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

# 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
у	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)

p.dims

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 heightres figure resolution (DPI)
```

phyloseq2df

			-
path	get	Tabe	T

splitting path and returning just one item in the vector

# Description

splitting path and returning just one item in the vector

## Usage

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

## **Arguments**

file\_path

File path

index

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

end.

## Value

string

phyloseq2df

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

 ${\tt physeq\_func}$ 

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 19

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

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

```
p 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?
```

20 read\_bracken

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)
```

## 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.

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

Robj\_md5sum 21

#### **Arguments**

infile Path to bracken table file is\_gzip Is the table file gzip'ed?

 $n_{lines}$  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

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

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

send\_email

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

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 23

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

snakemakeInfo

snakemake conda info

## **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

24 tidy\_pcoa

summary\_x

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

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.

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

25 unique\_n

#### **Arguments**

df data.frame or tibble

the column specifying taxa or OTUs (no quotes needed) taxon\_col sample\_col the column specifying sample names (no quotes needed)

the column specifying the taxon abundances in each sample (no quotes needed) abundance\_col dists

vector of beta-diversity distances ('wunifrac' = weighted UniFrac, 'unifrac' =

unweighted Unifrac; see vegan::vegsist for others)

phylogeny for UniFrac calculations. It can have more tips that what is in the tree

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

#### **Details**

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

## Value

data.frame

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**

a vector or data.table. If data.table, sel\_col must not be NULL Х

label what to call the items in the vector (eg., "samples") If x is data.table or data.frame, which column to assess? sel\_col

26 write\_table

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 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**

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