Package 'OTUtable'

October 12, 2022

Type Package

Title North Temperate Lakes - Microbial Observatory 16S Time Series
Data and Functions

Version 1.1.2

Date 2018-05-26

Author Alexandra Linz

Maintainer Alexandra Linz <amlinz16@gmail.com>

Description Analyses of OTU tables produced by 16S rRNA gene amplicon sequencing, as well as example data. It contains the data and scripts used in the paper Linz, et al. (2017) ``Bacterial community composition and dynamics spanning five years in freshwater bog lakes,'' <doi:10.1128/mSphere.00169-17>.

License GPL-3

LazyLoad yes

Depends R (>= 2.10)

NeedsCompilation no

Repository CRAN

Date/Publication 2018-05-26 15:20:55 UTC

R topics documented:

OTUtable-package   *OTU table analysis functions*

Description

Contains functions for the analysis of an OTU table generated from 16S rRNA amplicon sequencing. It also includes the data from the North Temperate Lakes-Microbial Observatory used in the paper Linz, et al. (2017) "Bacterial community composition and dynamics spanning five years in freshwater bog lakes." Please cite this paper if you use OTUtable. The data and code used in this paper are available at <https://github.com/McMahonLab/North_Temperate_Lakes-Microbial_Observatory>. Three data files are included: otu_table, taxonomy, and metadata. Access these by calling them with data(). There is also a fasta file associated with this dataset that is not included in this package - it can be found on our GitHub page in Data/16S_data. This package does not include functionality for fasta files; if you need this for analyses such as calculating UniFrac distance, please see the R package "phyloseq".

Details

Package:	OTUtable
Type:	Package
Version:	1.1.2
Date:	2018-05-26
License:	GPL-3

Functions include: bog_subset
chao1
clean_shared
clean_mothur_taxonomy
clean_TaxAss_taxonomy
combine_otus
extract_date

> filter_taxa
> grab_group
> make_do_matrix
> make_temp_matrix
> obs_richness
> pielou
> plot_column
> reduce_names
> remove_reps
> rotate
> shannon
> strat_metric
> year_subset
> zscore

Author(s)

Alexandra Linz <amlinz16@gmail.com>

References

Alexandra M. Linz, Benjamin C. Crary, Ashley Shade, Sarah Owens, Jack A. Gilbert, Rob Knight, Katherine D. McMahon. Bacterial Community Composition and Dynamics Spanning Five Years in Freshwater Bog Lakes. mSphere Jun 2017, 2 (3) e00169-17; DOI: 10.1128/mSphere.00169-17

---

bog_subset | *Subset OTU table by sampling site*

Description

Returns an OTU table containing only samples from the identified sampling site. This function can also be used on tables of higher level taxa generated by combine_otus(), or on tables that have already been processed by year_subset().

Usage

```
bog_subset(bog_id, table)
```

Arguments

bog_id
: The three letter code indicating the sampling site. The bog is represented by letters one and two; options are TB, SS, CB, NS, MA, HK, WS, and FB. The third letter indicates the layer; E for epilimnion and H for hypolimnion. The bog_id should be in quotes, and regular expressions can be used.

table
: A table containing the relative abundances of each taxon as rows and samples as columns. Sample names must be coded in the format bog, layer, date, and replicate (example: TBE07JUN08.R2 == Trout Bog Epilimnion, collected 07Jun08, replicate 2)

Value

Returns a relative abundance table containing samples from the specified sampling site in columns, with taxa in rows

Author(s)

Alexandra Linz <amlinz16@gmail.com>

Examples

```
data(otu_table)

Trout_Bog_Epilimnion <- bog_subset("TBE", otu_table)

Hells_Kitchen_Hypolimnion <- bog_subset("HKH", otu_table)

# Include both epilimnion and hypolimnion in a single table
Trout_Bog_both_layers <- bog_subset("TB.", otu_table)

# Include all meromictic hypolimnia
meromictic_hypolimnia <- bog_subset("HKH|MAH", otu_table)
```

---

chao1   *Chao1 Richness*

Description

Calculates Chao1 richness of a vector of relative abundance data. This alpha diversity metric takes into account the number of singletons and doubletons for a more accurate estimate than observed richness.

Usage

```
chao1(sample)
```

Arguments

sample   A vector of relative abundance data, typically a column in a matrix

Value

Returns a single number indicating the estimated richness in the tested sample based on the number of taxa appearing only once or twice

Note

Use apply functions to calculate Chao1 richness for all samples in a matrix

Author(s)

Alexandra Linz <amlinz16@gmail.com>

Examples

```
data(otu_table)
chao1_richness <- apply(otu_table, 2, chao1)
```

---

clean_mothur_taxonomy   *Clean mothur-format Taxonomy File*

Description

Reduces information in a mothur .taxonomy file by removing the second column with the number of reads per OTU. It will also check for and remove OTUs in the taxonomy file that are not in OTU table - this may be the case if rarefaction was performed after classification, as was used in the NTL-Microbial Observatory dataset. This function was formerly clean_taxonomy in v1.0.0.

Usage

```
clean_mothur_taxonomy(taxonomy_file, table, remove_bootstrap)
```

Arguments

taxonomy_file	A .taxonomy file output by mothur
table	An OTU table containing OTU numbers as row names
remove_bootstrap	
	TRUE or FALSE: if TRUE, removes bootstrap values from the classification strings

Value

Returns the taxonomy with OTUs as row names and seven columns containing each taxonomic level (Kingdom, Phylum, Class, Order, Lineage, Clade, and Tribe)

Author(s)

Alexandra Linz <amlinz16@gmail.com>

Examples

```
# Example path only: path <- "mothur_output/bogs.taxonomy"
# table <- clean_shared("mothur_output/bogs.shared", trim.names = T)
# taxonomy <- clean_mothur_taxonomy(path, table, remove_bootstrap = F)
```

clean_shared	*Reformat a shared file*

Description

Converts a mothur .shared file into a simplified OTU table. The columns indicating total reads for each OTU and the clustering level are removed, and the table is transposed so that OTUs are rows and samples are columns. The "trim.names" variable provides an option to shorten sample names to the first "." character - this is specific to the NTL-Microbial Observatory dataset. Manual curation of sample names took place after this step for the NTL-Microbial Observatory dataset in order to maintain consistency across all sample names.

Usage

```
clean_shared(shared_file, trim.names)
```

Arguments

shared_file	A .shared file output by mothur
trim.names	TRUE or FALSE - if TRUE, sample names will be trimmed to the first "." character.

Value

Returns an OTU table with samples as columns and OTUs as rows.

Author(s)

Alexandra Linz <amlinz16@gmail.com>

Examples

```
# Example path only: path <- "mothur_output/bogs.shared"

# otu_table <- clean_shared(path, trim.names = T)
# write.csv(otu_table, file = "bogs_otu_table.csv", quote = F, row.names = T)
```

---

clean_TaxAss_taxonomy   *Clean Taxonomy File Output by TaxAss Workflow*

Description

Formats a taxonomy file output by the McMahon Lab TaxAss 16S classification workflow (github.com/McMahonLab/TaxAss into the same format produced by clean_mothur_taxonomy(). It will also check for and remove OTUs in the taxonomy file that are not in OTU table - this may be the case if rarefaction was performed after classification, as was used in the NTL-Microbial Observatory dataset.

Usage

```
clean_TaxAss_taxonomy(taxonomy_file, table, remove_bootstrap)
```

Arguments

taxonomy_file   A .taxonomy file output by the TaxAss workflow

table   An OTU table containing OTU numbers as row names

remove_bootstrap

 TRUE or FALSE: if TRUE, removes bootstrap values from the classification strings

Value

Returns the taxonomy with OTUs as row names and seven columns containing each taxonomic level (Kingdom, Phylum, Class, Order, Lineage, Clade, and Tribe)

Author(s)

Alexandra Linz <amlinz16@gmail.com>

Examples

```
# Example path only: path <- "TaxAss_output/bogs.taxonomy"
# table <- clean_shared("mothur_output/bogs.shared", trim.names = T)
# taxonomy <- clean_TaxAss_taxonomy(path, table, remove_bootstrap = F)
```

---

combine_otus   *Combine OTUs based on identical taxonomic assignments*

Description

Sums the abundances of OTUs with the same taxonomy at a given level into a single vector for that taxonomy. This creates a new table of relative abundance data at a higher taxonomic level than OTU. This function only works with the OTU level as input, but can be used on any subset of the OTU table created by year_subset() or bog_subset(). The OTU table must have the same number of rows as the taxonomy file (do not remove rows with no reads before running combine_otus()) If bootstrap values were not removed by expand_taxa(), this command will likely create spurious groupings based on identical bootstrap values.

Usage

```
combine_otus(level, table, taxonomy)
```

Arguments

level	The desired level at which to combine OTUs; options are the column names from the taxonomy dataset
table	An OTU table containing the relative abundances of each OTU.
taxonomy	A taxonomy dataset in the form produced by expand_taxa().

Value

Returns a table of relative abundance data with each row representing all OTUs of a given taxonomic assignment summed together. Row names are now the full taxonomic assignment of each row. To keep only the the lowest taxonomic level in the row names, run the function reduce_names()

Author(s)

Alexandra Linz <amlinz16@gmail.com>

Examples

```
data(otu_table)
data(taxonomy)

example_table <- year_subset("05", otu_table)
example_table <- bog_subset("TBE", example_table)

phylum_table <- combine_otus("Phylum", example_table, taxonomy)
```

---

extract_date   *Extract sampling date from a vector of sample names*

Description

The date each sample was collected is encoded in the sample ID. Extract this into R date format using this command.

Usage

```
extract_date(sample_ids)
```

Arguments

sample_ids   A vector of sample names. Samples must be labeled using the bog, layer, date, and replicate system (MAH04JUL05.R1 = Mary Lake Hypolimnion, 04Jul05, replicate 1)

Value

Returns a vector of dates corresponding to each sample

Author(s)

Alexandra Linz <amlinz16@gmail.com>

Examples

```
samples <- c("TBE01JUN09.R1", "TBE05JUN09", "TBE10JUN09.R2")
extract_date(samples)

# Extract sample dates from the OTU table
data(otu_table)
x <- extract_date(colnames(otu_table))

# Extract sample dates from the metadata
data(metadata)
x <- extract_date(metadata$Sample_Name)
```

---

filter_taxa   *Filter Taxa Based on Abundance and Persistence*

Description

Returns a table containing only taxa that meet the imposed requirements of a minimum abundance and a minimum number of samples containing that taxon

Usage

```
filter_taxa(table, abundance, persistence)
```

Arguments

table
: A table containing the relative abundances of each OTU or taxon in the form produced by clean_shared(). Can be used on the output of grab_groups() or combine_otus()

abundance
: The minimum threshold for percentage of reads attributed to a taxon in at least one sample. Taxa at abundances greater than or equal this number will be retained.

persistence
: The minimum threshold for the percentage of samples in which a taxon has been observed. Taxa at abundances greater than or equal this number will be retained.

Value

Returns a table with all taxa that met the imposed thresholds

Note

Thanks Juliana Dias for suggesting this function!

Author(s)

Alexandra Linz <amlinz16@gmail.com>

Examples

```
# To make a table containing only OTUs with at least 0.1% abundance
# in at least one sample that were observed
# (at any abundance) in at least 50% of samples:
# library(OTUtable)
# data(otu_table)
# filtered_table <- filter_taxa(otu_table, abundance = 0.1, persistence = 50)

# To make a table containing only phyla with at least 10% abundance
# in any one sample and were observed
# at any abundance in at least 10% of samples:
```

```
# data(taxonomy)
# phylum_table <- combine_otus("Phylum", otu_table, taxonomy)
# filtered_phylum_table <- filter_taxa(phylum_table, abundance = 10, persistence = 10)
```

---

grab_group   *Subset OTU table by taxonomic assignment*

Description

Returns a table containing only taxa from a given phylogenetic group

Usage

```
grab_group(group, level, table, taxonomy)
```

Arguments

group	The phylogenetic classification of interest (can be a regular expression)
level	The phylogenetic level of the group of interest (must be a column name in the taxonomy file)
table	A table containing the relative abundances of each OTU in the form produced by clean_shared()
taxonomy	A taxonomy dataset in the form produced by expand_taxa()

Value

Returns a table with all taxa of a given taxonomic assignment

Note

This function must be run on the OTU level table. However, the output of this function can be run through combine_otus() to create a higher level table of results. Sometimes closely related groups were classified better in the Greengenes vs the freshwater database during classification of the NTL-Microbial Observatory dataset. In this case, it is necessary to search for the names generated by both datasets to get all closely related OTUs. For example, Methylophilaceae in Greengenes are named betIV in the freshwater database.

Author(s)

Alexandra Linz <amlinz16@gmail.com>

Examples

```
data(otu_table)
data(taxonomy)

acI <- grab_group("acI", "Clade", otu_table, taxonomy)
verruco <- grab_group("Verrucomicrobia", "Phylum", otu_table, taxonomy)

# Example where two search terms are needed due to classification with two databases
methylophilaceae <- grab_group("Methylophilaceae|betIV", "Clade", otu_table, taxonomy)
```

---

make_do_matrix   *Make matrix of dissolved oxygen data*

Description

Takes a given sample ID and converts the dissolved oxygen data in data(metadata) from long format into a matrix. This is useful for plotting using plot_column()

Usage

```
make_do_matrix(sampleID, field_data)
```

Arguments

sampleID   A regular expression used to select a group of samples

field_data   A dataset of DO profiles in long format. Column names must be the same as the
               metadata file provided with this package

Details

Also fills in NA values with the average of the depth above and below the missing value. If the value is at the bottom of the water column, the second deepest is substituted.

Value

Returns matrix of DO data with depth in rows and date in columns

Note

This is mainly used for generating contour plots. In general, long format is easier to work with. In the metadata file included in this package, each DO measurement is listed twice, once under the epilimnion sample name and again under the hypolimnion sample name.

Author(s)

Alexandra Linz <amlinz16@gmail.com>

Examples

```
data(metadata)

dissolved_oxygen <- make_do_matrix("TBE.....07", metadata)
```

---

make_temp_matrix *Make matrix of temperature data*

Description

Takes a given sample ID and converts temperature data of water profiles over time from long format into a matrix. This is most often useful for plotting using plot_column().

Usage

```
make_temp_matrix(sampleID, field_data)
```

Arguments

sampleID	A regular expression used to select a group of samples
field_data	A dataset of temperature profiles in long format. Column names must be the same as the metadata file provided with this package

Value

Returns matrix of temperature data with depth in rows and date in columns

Note

This is mainly used for generating contour plots. In general, long format is easier to work with. In the included metadata file, each temperature measurement is recorded twice, once as epilimnion and once as hypolimnion.

Author(s)

Alexandra Linz <amlinz16@gmail.com>

Examples

```
data(metadata)

temp <- make_temp_matrix("TBE.....07", metadata)
```

---

metadata	*Lake metadata for OTU table*

Description

A dataset containing temperature and oxygen profiles from the lakes in this study

Usage

```
data(taxonomy)
```

Format

A dataframe with 6 columns (measured variables) and 13,607 rows (depth profiles)

Details

Missing data indicated by NA Some sample dates and metadata dates may not match up exactly; if this presents an issue, please email and I will look at our written records for the right date Epilimnion and hypolimnion samples each have an identical depth profile entry; search for just one or the other

Author(s)

Alexandra Linz <amlinz16@gmail.com>

---

obs_richness	*Observed Richness*

Description

Calculates observed richness on a single column of relative abundance data.

Usage

```
obs_richness(sample)
```

Arguments

sample   A vector of relative abundance data, typically a single column in a matrix

Value

Returns a single number indicating the number of taxa in the tested sample

Note

Use apply functions to calculate richness for all samples in a matrix

Author(s)

Alexandra Linz <amlinz16@gmail.com>

Examples

```
data(otu_table)
richness <- apply(otu_table, 2, obs_richness)
```

---

otu_table	*OTU table generated from 8 bog lakes over 4 years*

Description

A dataset containing bacterial relative abundance data from the North Temperate Lakes Microbial Observatory Produced from mothur output using clean_shared()

Usage

```
data(otu_table)
```

Format

A dataframe with 1,387 columns (samples) and 6,208 rows (OTUs)

Details

Contains replicate samples Each column has been rarefied to 2500 Sample names encode sampling site ("TB"), epilimnion or hypolimnion ("E" or "H"), sampling date ("01JUN07") and replicate(".R2")

Author(s)

Alexandra Linz <amlinz16@gmail.com>

---

pielou	*Pielou's Evenness*

Description

Calculates Pielou's evenness for a single vector of relative abundance data

Usage

```
pielou(sample)
```

Arguments

sample   A vector of relative abundance data

Value

Returns a single value indicating the evenness of a community

Author(s)

Alexandra Linz <amlinz16@gmail.com>

Examples

```
data(otu_table)
even <- apply(otu_table, 2, pielou)
```

---

plot_column   *Plot DO or temperature data from a depth profile over time*

Description

Takes output from make_do_matrix or make_temp_matrix and plots using filled.contour()

Usage

```
plot_column(data_matrix, title)
```

Arguments

data_matrix   A matrix output by make_do_matrix() or make_temp_matrix()

title   The title you would like on the plot

Value

Plots a filled contour plot showing the water column over time

Note

Depends on the function rotate(). The functions make_do_matrix() and make_temp_matrix() fill in missing values with the average of the measurement at each depth above and below; however, if missing values are present in the matrix for plotting, these will appear as white space on the plot.

Author(s)

Alexandra Linz <amlinz16@gmail.com>

Examples

```
data(metadata)
temp <- make_temp_matrix("TBE.....07", metadata)
plot_column(temp, "Trout Bog 2007 Temperature")
```

---

reduce_names   *Shorten taxonomic assignment in table row names*

Description

Reduces the full string indicating taxonomy to the last classified level. Works on tables at levels higher than OTUs.

Usage

```
reduce_names(table)
```

Arguments

table   A table containing the relative abundances of each taxa produced by combine_otus()

Value

Returns the same table with shortened row names

Note

This function is often most useful for plotting, so that the full string does not appear on the plot

Author(s)

Alexandra Linz <amlinz16@gmail.com>

Examples

```
data(otu_table)
data(taxonomy)

# Create a small table for the example
# example <- year_subset("05", otu_table)
# example <- bog_subset("TBE", example)

# clade_table <- combine_otus("Clade", example, taxonomy)
# clade_table <- clade_table[which(rowSums(clade_table) > 0),]
# head(rownames(clade_table))
# reduced_clades <- reduce_names(clade_table)
# head(rownames(reduced_clades))
```

---

remove_reps   *Remove the second replicate of each sample, when it exists*

Description

Sometimes it is desirable to remove replicate samples (often for plotting). This command removes all samples marked as replicate 2. Please note that you should always check the similarity of replicates for your metric of interest before removing them for aesthetic purposes.

Usage

```
remove_reps(table)
```

Arguments

table   An OTU table containing the relative abundances of each OTU

Value

Returns an OTU table containing only one replicate for each sample

Author(s)

Alexandra Linz <amlinz16@gmail.com>

Examples

```
data(otu_table)
no_reps <- remove_reps(otu_table)
```

---

rotate   *Rotate a matrix*

Description

Rotates a matrix of data so that columns are reversed

Usage

```
rotate(data_matrix)
```

Arguments

data_matrix   Used in this package with matrix output by make_do_matrix or make_temp_matrix as part of the function plot_column(). Any matrix will work, though.

Details

Used to rotate the DO or temperature matrices so that depth 0 is at the top of a contour plot and the max depth is at the bottom.

Value

Returns a matrix that has been rotated so that it reads from bottom to top

Note

Used with make_do_matrix(), make_temp_matrix(), and plot_column(). plot_column() depends on this function.

Author(s)

An anonymous author on Stack Overflow Alexandra Linz <amlinz16@gmail.com>

Examples

```
data(metadata)
temp <- make_temp_matrix("TBE.....07", metadata)
r_temp <- rotate(temp)
```

---

shannon   *Shannon's Biodiversity Index*

Description

Calculates Shannon's Biodiversity Index on a single column of relative abundance data. This metric takes into account both richness and evenness.

Usage

```
shannon(sample)
```

Arguments

sample   A vector of relative abundance data, typically a single column in a matrix

Value

Returns a single number indicating the amount of biodiversity in the tested sample

Note

Use apply functions to calculate Shannon's index for all samples in a matrix

Author(s)

Alexandra Linz <amlinz16@gmail.com>

Examples

```
data(otu_table)
richness <- apply(otu_table, 2, shannon)
```

---

taxonomy   *Taxonomic assignments of OTUs*

Description

A dataset containing the taxonomy of each OTU in the otu_table Produced from mothur output using clean_taxonomy() Bootstrap values have been removed from this dataset, but are still available in as part of the Data folder in the McMahonLab/North_Temperate_Lakes-Microbial_Observatory GitHub repo

Usage

```
data(taxonomy)
```

Format

A dataframe with 7 columns (taxonomic levels) and 6,208 rows (OTUs)

Details

Classified using our Freshwater database, followed by Greengenes - for the full workflow, visit the McMahonLab Github 16STax-Ass repository Some OTUs are missing; these were removed by subsampling of the OTU table The presence of both blank (__) assignments and "unclassified" assignments are the result of the dual classification.

Author(s)

Alexandra Linz <amlinz16@gmail.com>

---

year_subset *Subset samples by a specific year*

Description

Takes the year value in the last two digits of the sample ID and allows selection of a single year of data. Can be performed on tables at higher taxonomic levels generated by combine_otus(), or on tables already subset by bog_subset().

Usage

```
year_subset(year_id, table)
```

Arguments

table   A table containing the relative abundances of each taxa

year_id   Two digit code indicating the last two digits of the year of interest (05, 07, 08, 09) surrounded by quotes. Regular expressions can be used.

Value

Returns an OTU table containing only samples from the specified year

Author(s)

Alexandra Linz <amlinz16@gmail.com>

Examples

```
data(otu_table)

seven <- year_subset("07", otu_table)

# Select two years at once
two_years <- year_subset("07|08", otu_table)
```

---

zscore *Z-score normalize relative abundance data*

Description

Normalizes taxa abundances in a table of relative abundance data using the z-score method. ((Abundance of one OTU in one sample) - (mean abundance for that OTU))/(standard deviation of that OTU)

Usage

```
zscore(table)
```

Arguments

table   A table of relative abundance data with taxa in rows and samples in columns

Value

Returns a table with relative abundance data replaced by z-scores

Note

There is debate on whether this method of normalization is valid for microbial communities, as their abundance distrubtions tend to be heavily skewed. I found it useful for plotting heatmaps and for input into network analysis.

Author(s)

Alexandra Linz <amlinz16@gmail.com>

Examples

```
data(otu_table)

# Create a small table for z-score normalization
example <- year_subset("05", otu_table)
example <- bog_subset("TBE", example)

# Remove OTUs that are not present in this subset
example <- example[which(rowSums(example) > 0), ]

z_otu_table <- zscore(example)
```

Index