

Package ‘qiimer’

August 29, 2016

License GPL (>= 2)

Maintainer Kyle Bittinger <kylebittinger@gmail.com>

Title Work with QIIME Output Files in R

Author Kyle Bittinger <kylebittinger@gmail.com>

Description Open QIIME output files in R, compute statistics, and create plots from the data.

Version 0.9.4

URL <http://www.r-project.org>

Date 2015-12-27

Imports grDevices, pheatmap, stats, utils

Suggests testthat, biom

RoxygenNote 5.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2015-12-28 00:01:50

R topics documented:

| | |
|-----------------------------------|---|
| biom_raw_data | 2 |
| biom_taxonomy | 3 |
| dist_get | 4 |
| dist_groups | 4 |
| dist_subset | 5 |
| make_otu_table | 6 |
| otu_heatmap | 6 |
| qiimer | 7 |
| rarefaction_stats | 7 |
| read_blast_table | 8 |
| read_qiime_distmat | 8 |
| read_qiime_mapping_file | 9 |
| read_qiime_otu_mapping | 9 |

| | |
|----------------------------------|----|
| read_qiime_otu_table | 10 |
| read_qiime_rarefaction | 10 |
| relmbeta | 11 |
| relmbeta_alpha | 11 |
| relmbeta_assignments | 12 |
| relmbeta_biom | 12 |
| relmbeta_counts | 13 |
| relmbeta_dist | 14 |
| saturated_rainbow | 14 |
| simplify_assignments | 15 |
| split_assignments | 16 |
| taxonomic_ranks | 16 |

Index 17

| | |
|---------------|---|
| biom_raw_data | <i>Extract raw data from a BIOM object.</i> |
|---------------|---|

Description

This function extracts the raw data from a BIOM object, using the correct row and column names in the result.

Usage

```
biom_raw_data(b)
```

Arguments

b A BIOM object.

Details

The BIOM object can be any list-like representation of the JSON source code in a BIOM-format file produced by QIIME. There are several options for creating BIOM objects from QIIME output files. The official library for BIOM files, `biom`, can create compatible objects via the `read_biom` function. Alternately, the `fromJSON` function from either `RJSONIO` or `rjson` may be used.

Value

For sparse biom objects, returns a 3-column data frame of row names, column names, and the data value. The first column is named using the first word in the BIOM object's type attribute (e.g. "OTU" for OTU tables). The second and third columns are named "SampleID" and "value", respectively. For dense biom objects, returns a matrix.

Examples

```
data(relmbeta_biom)
head(biom_raw_data(relmbeta_biom))
```

| | |
|---------------|--|
| biom_taxonomy | <i>Extract taxonomy info from a biom object.</i> |
|---------------|--|

Description

For BIOM objects representing OTU tables, this function will extract the taxonomy for each OTU. The taxonomy info is typically stored in an attribute of the row metadata named "taxonomy".

Usage

```
biom_taxonomy(b, attr = "taxonomy")
```

Arguments

| | |
|------|--|
| b | A BIOM object. |
| attr | The metadata attribute under which the taxonomy information can be found for each row item in the biom file. |

Details

This function may be used more generally to extract metadata from rows in a BIOM object. The `attr` argument may be adjusted to match the desired attribute name. If the metadata has a nested structure, a character vector may be supplied to the `attr` argument.

The BIOM object can be any list-like representation of the JSON source code in a BIOM-format file produced by QIIME. There are several options for creating BIOM objects from QIIME output files. The official library for BIOM files, `biom`, can create compatible objects via the `read_biom` function. Alternately, the `fromJSON` function from either `RJSONIO` or `rjson` may be used.

Value

A list of character vectors, one per row.

Examples

```
data(relmbeta_biom)
head(biom_taxonomy(relmbeta_biom))
```

| | |
|----------|---|
| dist_get | <i>Retrieve distances from a "dist" object.</i> |
|----------|---|

Description

Retrieve distances from a "dist" object.

Usage

```
dist_get(d, idx1, idx2)
```

Arguments

| | |
|------------|--|
| d | A distance matrix object of class "dist". |
| idx1, idx2 | Indices specifying the distances to extract. |

Value

A vector of distances.

Examples

```
data(relmbeta_dist)
dist_get(relmbeta_dist, "A1", "A2")
dist_get(relmbeta_dist, "A1", c("A2", "A3", "A4", "A5"))
dist_get(relmbeta_dist, c("A1", "A2", "A3"), c("B1", "B2", "B3"))
```

| | |
|-------------|--|
| dist_groups | <i>Create a data frame of distances between groups of items.</i> |
|-------------|--|

Description

Create a data frame of distances between groups of items.

Usage

```
dist_groups(d, g)
```

Arguments

| | |
|---|---|
| d | A distance matrix object of class "dist". |
| g | A factor representing the groups of objects in 'd'. |

Value

A data frame with 6 columns. "Item1" and "Item2" identify the items compared, using the label if available. Likewise, "Group1" and "Group2" identify the groups of the items. "Label" is a factor giving a convenient label for the type of comparison. Finally, "Distance" contains the distance of interest.

Examples

```
data(relmbeta_dist)
data(relmbeta)
head(dist_groups(relmbeta_dist, relmbeta$Diet))
```

| | |
|-------------|--|
| dist_subset | <i>Extract parts of a "dist" object.</i> |
|-------------|--|

Description

Extract parts of a "dist" object.

Usage

```
dist_subset(d, idx)
```

Arguments

| | |
|-----|--|
| d | A distance matrix object of class "dist". |
| idx | Indices specifying the subset of distances to extract. |

Value

A distance matrix.

Examples

```
data(relmbeta_dist)
dist_subset(relmbeta_dist, c("A1", "A2", "A3", "A4", "A5"))
```

| | |
|----------------|-----------------------------|
| make_otu_table | <i>Create an OTU table.</i> |
|----------------|-----------------------------|

Description

Create an OTU table.

Usage

```
make_otu_table(otus, sample_ids = NULL)
```

Arguments

| | |
|------------|--|
| otus | A list of QIIME-format sequence identifiers for each OTU. |
| sample_ids | An optional vector of sample IDs to include in the result. |

Value

A matrix of OTU counts by sample.

| | |
|-------------|--|
| otu_heatmap | <i>Create a heatmap of OTU counts.</i> |
|-------------|--|

Description

Create a heatmap of OTU counts.

Usage

```
otu_heatmap(otu_counts, assignments, threshold = 0, plot = T,
            color = saturated_rainbow(100), breaks = c(0, 1e-10, seq(0.001, 1,
            length.out = 100)), ...)
```

Arguments

| | |
|-------------|---|
| otu_counts | A matrix of OTU counts, one row per OTU and one column per sample. |
| assignments | A character vector of OTU assignments. Length should match number of rows in otu_counts. |
| threshold | Minimum number of OTU counts necessary for an assignment to be included in the heatmap. Assignments are filtered after calculating the proportions, so the threshold setting does not affect the display of the remaining OTUs. |
| plot | If true, display a plot. If false, just return the computed abundances. |
| color | Vector of colors to use in the heatmap. |
| breaks | Vector of color breaks, one element greater in length than 'colors'. |
| ... | Additional arguments are passed to the pheatmap function. |

Value

A heatmap plot of the proportions of assignments in each sample, and invisibly returns a matrix of the proportions in the plot.

See Also

[saturated_rainbow](#)

Examples

```
data(relmbeta_assignments)
data(relmbeta_counts)
a <- simplify_assignments(split_assignments(relmbeta_assignments))

## Not run:
otu_heatmap(relmbeta_counts, a, threshold=10)
otu_heatmap(
  relmbeta_counts, a, threshold=10,
  cluster_rows=FALSE, cluster_cols=FALSE,
  cellwidth=12, cellheight=12)

## End(Not run)

heatmap_data <- otu_heatmap(relmbeta_counts, a, threshold=10, plot=FALSE)
head(heatmap_data)
```

qiimer

qiimer: Read QIIME output files and create plots

Description

qiimer: Read QIIME output files and create plots

rarefaction_stats

Compute summary statistics for collated alpha diversity tables.

Description

Compute summary statistics for collated alpha diversity tables.

Usage

```
rarefaction_stats(rarefaction_table)
```

Arguments

rarefaction_table
A collated alpha diversity data frame.

Value

A data frame of summary statistics, with columns 'SampleID', 'sequences_per_sample', 'diversity.mean', and 'diversity.sd'.

Examples

```
data(relmbeta_alpha)
head(rarefaction_stats(relmbeta_alpha))
```

read_blast_table *Parse tabular output from BLAST.*

Description

Parse tabular output from BLAST.

Usage

```
read_blast_table(filepath)
```

Arguments

filepath Path to tabular BLAST output file.

Value

A data frame of BLAST results.

read_qiime_distmat *Read a QIIME distance matrix file.*

Description

Read a QIIME distance matrix file.

Usage

```
read_qiime_distmat(filepath)
```

Arguments

filepath Path to QIIME distance matrix file.

Value

A distance matrix.

`read_qiime_mapping_file`

Read a QIIME sample mapping file.

Description

Read a QIIME sample mapping file.

Usage

```
read_qiime_mapping_file(filepath)
```

Arguments

| | |
|----------|--|
| filepath | Path to sample mapping file. The file must conform to the QIIME standards, detailed at http://qiime.org/documentation/file_formats.html . |
|----------|--|

Value

A data frame of sample information. Because the SampleID column is so often used to extract data from distance matrices and OTU tables, it is returned as a character vector.

`read_qiime_otu_mapping`

Parse an OTU mapping file from QIIME.

Description

Parse an OTU mapping file from QIIME.

Usage

```
read_qiime_otu_mapping(filepath, prefix = "")
```

Arguments

| | |
|----------|---|
| filepath | Path to the OTU mapping file. |
| prefix | OTU names will be prefixed with this value. |

Value

A list of sequence identifiers for each OTU.

read_qiime_otu_table *Parse a QIIME OTU table file in "calssic" format.*

Description

Parse a QIIME OTU table file in "calssic" format.

Usage

```
read_qiime_otu_table(filepath, commented = TRUE, metadata = TRUE)
```

Arguments

| | |
|-----------|---|
| filepath | Path to OTU table file. |
| commented | TRUE if the header line is preceeded by an additional comment line, otherwise FALSE. This is usually the case for OTU tables generated with QIIME, so we default to TRUE. |
| metadata | TRUE if the OTU table contains a metadata column, otherwise FALSE. The metadata column usually contains taxonomic assignments, and must be located on the right-hand side of the table. |

Value

A list with four attributes: sample_ids, otu_ids, counts, and metadata, a data structure similar to that returned by the python function 'qiime.parse.parse_otu_table'. The sample_ids, otu_ids, and metadata attributes are character vectors. The counts attribute is a matrix with one column per sample_id and one row per otu_id.

read_qiime_rarefaction
Read a collated alpha diversity table from QIIME.

Description

Read a collated alpha diversity table from QIIME.

Usage

```
read_qiime_rarefaction(filepath)
```

Arguments

| | |
|----------|--|
| filepath | Path to collated alpha diversity file. |
|----------|--|

Value

A data frame with columns 'SampleID', 'sequences_per_sample', 'iteration', and 'diversity'.

relmbeta

Sample dataset from murine gut microbiome

Description

The relmbeta dataset is taken from a mouse study where wild-type and RELMbeta knockout mice were fed either a normal or high-fat diet. The diet was observed to have a pronounced effect on the gut microbiome composition. The genotype also had an effect, but less so.

Format

A data frame with 20 rows and 3 columns.

Details

The bacterial 16S rDNA gene was sequenced using 454 FLX technology, producing about 26k reads. The reads were processed via the standard QIIME workflow, `pick_de_novo_otus.py`, which clustered the reads into 776 operational taxonomic units (OTUs). QIIME version 1.7.0 was used for the analysis.

The OTU table was filtered to remove OTUs appearing in only one sample. Following this, a single rarefaction was performed at a level of 500 reads per sample. The unweighted UniFrac distance was then computed for each pair of samples.

The relmbeta data frame lists the sample IDs, genotypes, and dietary assignments for the mice.

Source

Hildebrandt et al. High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology* **137**, 1716 (2009).

relmbeta_alpha

Shannon diversity measurements from murine gut microbiome

Description

relmbeta_alpha is a data frame containing Shannon diversity (base 2) at a level of 10-500 sequences per sample.

Format

A data frame with 2200 rows and 4 columns.

Details

The [relmbeta](#) documentation provides an overview of the study.

Source

Hildebrandt et al. High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology* **137**, 1716 (2009).

See Also

[relmbeta](#)

relmbeta_assignments *Taxonomic assignments from murine gut microbiome*

Description

The character vector relmbeta_assignments contains taxonomic assignments for each OTU in the study.

Format

A character vector with 337 elements.

Details

The [relmbeta](#) documentation provides an overview of the study.

Source

Hildebrandt et al. High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology* **137**, 1716 (2009).

See Also

[relmbeta](#)

relmbeta_biom *BIOM format object from murine gut microbiome*

Description

The relmbeta_biom object is a representation of the BIOM file produced by QIIME. It was produced by loading the BIOM file with the function `RJSONIO::fromJSON`.

Format

A "biom" object with 227 rows and 20 columns.

Details

The [relbeta](#) documentation provides an overview of the study.

Source

Hildebrandt et al. High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology* **137**, 1716 (2009).

See Also

[relbeta](#)

| | |
|----------------|--|
| relbeta_counts | <i>OTU counts from murine gut microbiome</i> |
|----------------|--|

Description

The matrix `relbeta_counts` contains the number of reads observed in each OTU, after rarefaction to 500 reads per sample. OTUs are listed in the rows and samples are listed in the columns.

Format

An integer matrix with 337 rows and 20 columns.

Details

The [relbeta](#) documentation provides an overview of the study.

Source

Hildebrandt et al. High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology* **137**, 1716 (2009).

See Also

[relbeta](#)

| | |
|---------------|--|
| relmbeta_dist | <i>Unweighted UniFrac distances from murine gut microbiome</i> |
|---------------|--|

Description

relmbeta_dist is an object of class "dist", containing the unweighted UniFrac distances between samples.

Format

An object of class "dist".

Details

The [relmbeta](#) documentation provides an overview of the study.

Source

Hildebrandt et al. High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology* **137**, 1716 (2009).

See Also

[relmbeta](#)

| | |
|-------------------|-----------------------------------|
| saturated_rainbow | <i>Saturated rainbow palette.</i> |
|-------------------|-----------------------------------|

Description

This palette is specially designed for data consisting of counts. It is intended to show both presence/absence and relative proportion in the same plot. For data containing N counts in the largest sample, the saturated rainbow palette should be created with length N + 1.

Usage

```
saturated_rainbow(n, saturation_limit = 0.4)
```

Arguments

| | |
|------------------|--|
| n | Length of the palette |
| saturation_limit | The fraction of the total palette length over which the rainbow extends. Above this limit, the color will remain the same. |

Details

The first element of the palette is white, indicating zero counts. The second element is dark blue, indicating one or very few counts. As the proportion increases within a sample, the palette transitions from blue to green, yellow, orange, and finally red.

The function defines a saturation limit, above which the color remains bright red. The saturation limit is set to 40 items with the largest relative proportion in a sample. The default value seems to work well for a wide range of circumstances – it allows items that are strongly dominant in a sample to be identified across the plot. Ideally, the total number of red squares should be kept low, never more than one per sample.

Value

A vector of colors.

Examples

```
saturated_rainbow(5)
```

`simplify_assignments` *Reformat taxonomic assignments for presentation.*

Description

Reformat taxonomic assignments for presentation.

Usage

```
simplify_assignments(assignments_df, rank1 = "Phylum", rank2 = "Genus")
```

Arguments

`assignments_df` A data frame of taxonomic assignments.
`rank1` The rank of taxonomy to use as the first word in the label.
`rank2` The rank of taxonomy to use as the second word in the label.

Value

A character vector of reformatted assignment labels.

See Also

[split_assignments](#)

Examples

```
data(relmbeta_assignments)
a <- split_assignments(relmbeta_assignments)
head(simplify_assignments(a))
```

| | |
|-------------------|---|
| split_assignments | <i>Split taxonomic assignment strings</i> |
|-------------------|---|

Description

Split taxonomic assignment strings

Usage

```
split_assignments(assignments, ranks = taxonomic_ranks, split = "; ", ...)
```

Arguments

| | |
|-------------|--|
| assignments | Character vector of taxonomic assignments. |
| ranks | Character vector of taxonomic ranks, used as column names in the resultant data frame. |
| split | Pattern on which to split taxa in assignment strings. |
| ... | Additional parameters are passed to the <code>strsplit</code> function. |

Value

A data frame of taxonomic assignments.

See Also

[taxonomic_ranks](#)

Examples

```
data(relmbeta_assignments)
a <- split_assignments(relmbeta_assignments)
head(a)
```

| | |
|-----------------|----------------------------------|
| taxonomic_ranks | <i>Standard taxonomic ranks.</i> |
|-----------------|----------------------------------|

Description

Standard taxonomic ranks.

Usage

```
taxonomic_ranks
```

Format

An object of class character of length 7.

Index

*Topic **datasets**

- relmbeta, 11
- relmbeta_alpha, 11
- relmbeta_assignments, 12
- relmbeta_biom, 12
- relmbeta_counts, 13
- relmbeta_dist, 14
- taxonomic_ranks, 16

- biom_raw_data, 2
- biom_taxonomy, 3

- dist_get, 4
- dist_groups, 4
- dist_subset, 5

- make_otu_table, 6

- otu_heatmap, 6

- qiimer, 7
- qiimer-package (qiimer), 7

- rarefaction_stats, 7
- read_blast_table, 8
- read_qiime_distmat, 8
- read_qiime_mapping_file, 9
- read_qiime_otu_mapping, 9
- read_qiime_otu_table, 10
- read_qiime_rarefaction, 10
- relmbeta, 11, 11, 12-14
- relmbeta_alpha, 11
- relmbeta_assignments, 12
- relmbeta_biom, 12
- relmbeta_counts, 13
- relmbeta_dist, 14

- saturated_rainbow, 7, 14
- simplify_assignments, 15
- split_assignments, 15, 16

- taxonomic_ranks, 16, 16