

Package ‘metagenomeFeatures’

October 18, 2017

Title Exploration of marker-gene sequence taxonomic annotations

Version 1.8.1

Author Nathan D. Olson, Joseph Nathaniel Paulson, Hector Corrada Bravo

Maintainer Nathan D. Olson <nolson@umiacs.umd.edu>

Description metagenomeFeatures was developed for use in exploring the taxonomic annotations for a marker-gene metagenomic sequence dataset. The package can be used to explore the taxonomic composition of a marker-gene database or annotated sequences from a marker-gene metagenome experiment.

Depends R (>= 3.4), Biobase (>= 2.17.8)

License Artistic-2.0

LazyData true

Imports Biostrings (>= 2.36.4), ShortRead (>= 1.26.0), dplyr (>= 0.7.0), dbplyr (>= 1.0.0), stringr (>= 1.0.0), lazyeval (>= 0.1.10), RSQLite (>= 1.0.0), magrittr (>= 1.5), methods (>= 3.3.1), lattice (>= 0.20.33), metagenomeSeq (>= 1.14.2), ape (>= 3.5), purrr (>= 0.2.2)

Suggests knitr (>= 1.11), msd16s (>= 0.102.0), testthat (>= 0.10.0), rmarkdown (>= 1.3)

Collate 'utils.R' 'phylo-class.R' 'mgDb-class.R' 'mgFeatures-class.R' 'demoMgDb.R' 'mgDb-taxa_accessors.R' 'aggregate_taxa.R' 'mgQuery.R' 'vignette_data.R' 'mgDb_method_annotateMRExp.R' 'mgDb_method_select.R' 'mgDb_method_annotateFeatures.R' 'mockMgDb.R' 'mock_query_df.R' 'mock_mgF.R'

VignetteBuilder knitr

URL <https://github.com/HCBravoLab/metagenomeFeatures>

BugReports <https://github.com/HCBravoLab/metagenomeFeatures/issues>

biocViews Microbiome, Metagenomics, Annotation, Infrastructure, Sequencing, Software

NeedsCompilation no

RoxygenNote 6.0.1

R topics documented:

aggregate_taxa	2
annotateFeatures	3
annotateMRexp_fData	4
get_demoMgDb	4
get_mockMgDb	5
MgDb-class	5
mgdb_meta	6
mgDb_select	6
mgdb_seq	7
mgdb_taxa	8
mgdb_tree	8
mgFeatures-class	9
mgF_meta	9
mgF_seq	10
mgF_taxa	10
mgF_tree	11
mgQuery	11
mock_mgF	12
mock_query_df	12
show,MgDb-method	13
taxa_columns	13
taxa_keys	14
taxa_keytypes	14
taxa_levels	15
Index	16

aggregate_taxa	<i>Aggregates a MRexperiment object and returns either an aggregated MRexperiment or counts matrix to a user defined taxonomic level.</i>
----------------	---

Description

Using the featureData information in the MRexperiment-class object, aggregate_taxa aggregates the OTU count data (MRexperiment assayData slot) to a user defined taxonomic level (i.e. 'genus') using the defined aggfun function (default colSums). Possible aggfun alternatives include and column wise matrix calculations, e.g. colMeans, colMedians.

Usage

```
aggregate_taxa(obj, lvl, aggfun = colSums, out = "MRexperiment", ...)
```

Arguments

obj	A MRexperiment-class object or count matrix.
lvl	featureData column name from the MRexperiment object or if count matrix object a vector of labels.
aggfun	Matrix aggregation function, e.g. colSums.
out	Either 'MRexperiment' or 'matrix'
...	Additional parameters to pass to MRcount, e.g. norm, log, and sl.

Value

An aggregated count matrix or MRexperiment

Examples

```
data("mouseData", package = "metagenomeSeq")
aggregate_taxa(mouseData[1:100,], lvl="class", norm=TRUE, aggfun=colSums)
aggregate_taxa(mouseData, lvl="class", norm=TRUE, aggfun=colSums)
aggregate_taxa(mouseData, lvl='phylum', norm=FALSE, aggfun=colSums)
```

annotateFeatures	<i>Annotating metagenome data with taxonomic information</i>
------------------	--

Description

This method is used to create a [mgFeatures](#) class object

Usage

```
annotateFeatures(mgdb, ...)
```

S4 method for signature 'MgDb'

```
annotateFeatures(mgdb, query)
```

Arguments

mgdb	MgDb class object
...	additional arguments passed to select function
query	A data frame with experimental data to annotate with taxonomic information, must include column named "Key" with database ids. Or a vector of database Keys of entries to include in mgFeatures-class object.

Value

mgFeatures-class object

Examples

```
## MgDb with mock community ids
mockMgDb <- get_mockMgDb()
## generating mgFeatures object
data(mock_query_df)
mock_mgF <- annotateFeatures(mockMgDb, mock_query_df)
```

annotateMRexp_fData *Annotate MRexperiment object featureData slot using MgDb object*

Description

This method is used to define a MRexperiment object featureData slot with taxonomic information from a [MgDb-class](#) object using the MRexperiment object's Feature names. object.

Usage

```
annotateMRexp_fData(mgdb, MRobj, ...)
```

```
## S4 method for signature 'MgDb'  
annotateMRexp_fData(mgdb, MRobj)
```

Arguments

mgdb	MgDb class object
MRobj	MRexperiment class object
...	additional arguments passed to select function

Value

MRexperiment-class object

Examples

```
# see vignette
```

get_demoMgDb *Example MgDb-class object*

Description

Example [MgDb-class](#) object with 3211 entries from the Greengenes 13.8 OTU 99 database.

Usage

```
get_demoMgDb()
```

Value

MgDb-class object

Examples

```
get_demoMgDb()
```

get_mockMgDb	<i>Mock MgDb-class object</i>
--------------	-------------------------------

Description

Mock [MgDb-class](#) object with a subset of the GreenGenes 13.9 OTU 0.99 database including the ids for OTU from a mock community dataset from the study Bokulich et al. 2013, the OTU ids were extracted from a biom file downloaded from QIITA (<https://qiita.ucsd.edu>).

Usage

```
get_mockMgDb()
```

Details

Bokulich, Nicholas A., et al. "Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing." *Nature methods* 10.1 (2013): 57-59.

Value

MgDb-class object

Examples

```
get_mockMgDb()
```

MgDb-class	<i>Metagenome Database class</i>
------------	----------------------------------

Description

The MgDb-class object contains sequence and taxonomic data for a 16S rRNA taxonomic database, see the [greengenes13.5MgDb](#) package as an example database. The `get_demoMgDb` function exports a small subset of the database in [greengenes13.5MgDbmetagenomeFeatures](#) package as an example of a MgDb-class object.

Usage

```
# library(greengenes13.5MgDb)
```

Fields

taxa taxonomic information for database sequences
 seq database reference sequences
 tree reference phylogenetic tree
 taxa_file name of sqlite db
 tree_file name of phylogenetic tree file
 metadata associated metadata for the database

Note

Currently the only database with a MgDb package is the **Greengenes database** (version 13.5), additional packages are planned.

Examples

```
# example MgDb-class object, a small subset of the Greengenes 13.5 database.
get_demoMgDb()
```

mgdb_meta	<i>MgDb metadata slot accessor</i>
-----------	------------------------------------

Description

MgDb metadata slot accessor

Usage

```
mgdb_meta(mgdb)
```

Arguments

mgdb	MgDb class object
------	-------------------

Value

list

Examples

```
demoMgDb <- get_demoMgDb()
mgdb_meta(demoMgDb)
```

mgDb_select	<i>Querying MgDb objects</i>
-------------	------------------------------

Description

Function for querying **MgDb** class objects, user defines the taxonomic levels (keytype) and a vector of taxonomic names (keys) being selected. If specific database ids are being selected for use keytype="Keys". Additionally, users can specify whether they want only the taxonomic and sequence data, or both.

Usage

```
mgDb_select(mgdb, type, ...)
```

```
## S4 method for signature 'MgDb'
mgDb_select(mgdb, type, keys = NULL, keytype = NULL,
            columns = "all")
```

Arguments

mgdb	MgDb class object
type	either "taxa", "seq", "tree", "all" or a character vector of types. "taxa", "seq", and "tree" only query the reference taxonomy, sequences, and phylogenetic tree respectively. "all" queries the reference taxonomy, sequence, and phylogenetic tree.
...	additional arguments passed to select function
keys	specific taxonomic groups to select for
keytype	taxonomic level of keys
columns	keytypes in taxonomy database to return, all by default

Value

returned object depends on type: 'taxa' - dataframe with taxa information; 'seq' - DNAStrngSet with sequence data; 'tree' - phylogenetic tree of class phylo; 'all' - list with the dataframe, DNAStrngSet, and phylo.

Examples

```
demoMgDb <- get_demoMgDb()
# select taxa only
mgDb_select(demoMgDb, type = "taxa",
            keys = c("Vibrionaceae", "Enterobacteriaceae"),
            keytype = "Family")

# select seq only
mgDb_select(demoMgDb, type = "seq",
            keys = c("Vibrionaceae", "Enterobacteriaceae"),
            keytype = "Family")

# select all taxa, seq, and tree
mgDb_select(demoMgDb, type = "all",
            keys = c("Vibrionaceae", "Enterobacteriaceae"),
            keytype = "Family")
```

mgdb_seq

MgDb seq slot accessor

Description

MgDb seq slot accessor

Usage

```
mgdb_seq(mgdb)
```

Arguments

mgdb	MgDb class object
------	-------------------

Value

DNAStrngSet class object

Examples

```
demoMgDb <- get_demoMgDb()  
mgdb_seq(demoMgDb)
```

mgdb_taxa

MgDb taxa slot accessor

Description

MgDb taxa slot accessor

Usage

```
mgdb_taxa(mgdb)
```

Arguments

mgdb MgDb class object

Value

tbl_sql connection to sqlite table

Examples

```
demoMgDb <- get_demoMgDb()  
mgdb_taxa(demoMgDb)
```

mgdb_tree

MgDb tree slot accessor

Description

MgDb tree slot accessor

Usage

```
mgdb_tree(mgdb)
```

Arguments

mgdb MgDb class object

Value

phylo class object

Examples

```
demoMgDb <- get_demoMgDb()
mgdb_tree(demoMgDb)
```

mgFeatures-class	<i>mgFeature-class object</i>
------------------	-------------------------------

Description

Object contains taxonomic annotation and reference sequence data for classified OTUs. The class extends the [AnnotatedDataFrame](#) class with a slot with a subset of the reference database sequences and phylogenetic tree for taxonomically classified OTUs, along with an additional slot for metadata including information on the database source.

Value

mgFeature class object

Slots

metadata list
 refDbSeq DNAStrngSet
 refDbTree phyloOrNULL

Examples

```
data(mock_mgF)
```

mgF_meta	<i>mgFeatures metadata slot accessor</i>
----------	--

Description

mgFeatures metadata slot accessor

Usage

```
mgF_meta(mgF)
```

Arguments

mgF mgFeatures class object

Value

list

Examples

```
data(mock_mgF)
mgF_meta(mock_mgF)
```

mgF_seq	<i>mgFeatures refDbSeq slot accessor</i>
---------	--

Description

mgFeatures refDbSeq slot accessor

Usage

```
mgF_seq(mgF)
```

Arguments

mgF mgFeatures class object

Value

DNAStrngSet class object

Examples

```
data(mock_mgF)
mgF_seq(mock_mgF)
```

mgF_taxa	<i>mgFeatures taxa slot accessor</i>
----------	--------------------------------------

Description

mgFeatures taxa slot accessor

Usage

```
mgF_taxa(mgF)
```

Arguments

mgF mgFeatures class object

Value

AnnotatedDataFrame

Examples

```
data(mock_mgF)
mgF_taxa(mock_mgF)
```

mgF_tree	<i>mgFeatures refDbTree slot accessor</i>
----------	---

Description

mgFeatures refDbTree slot accessor

Usage

```
mgF_tree(mgF)
```

Arguments

mgF mgFeatures class object

Value

phylo class object

Examples

```
data(mock_mgF)
mgF_tree(mock_mgF)
```

mgQuery	<i>Example ShortRead 16S experiment dataset</i>
---------	---

Description

A dataset containing the top OTU centers from a 16S marker gene survey. Specifically, the OTU centers came from 'Healthy and moderate to severe diarrhea 16S expression data' - bioconductor dataset 'msd16s'.

Usage

```
mgQuery
```

Format

A metagenomeFeatures object with 500 sequences.

Source

<http://bioconductor.org/packages/release/data/experiment/html/msd16s.html>

Examples

```
data(mgQuery)
```

`mock_mgF`*Example mgFeatures class object*

Description

Example `mgFeatures-class` object generated using the `annotateFeatures MgDb-class` method. The dataset contains the OTU ids and Greengenes database version 13.5 ids. Specifically, the OTU centers came from a mock community dataset from the study Bokulich et al. 2013, the OTU ids were extracted from a biom file downloaded from QIITA (<https://qiita.ucsd.edu>).

Usage`mock_mgF`**Format**`mgFeatures`**Details**

Bokulich, Nicholas A., et al. "Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing." *Nature methods* 10.1 (2013): 57-59.

Examples`data(mock_mgF)`

`mock_query_df`*Example Query Data Frame*

Description

Example `query_df` for use in generating a `mgFeatures-class` object using the `annotateFeatures MgDb-class` method. The dataset contains the OTU ids and Greengenes database version 13.5 ids. Specifically, the OTU centers came from a mock community dataset from the study Bokulich et al. 2013, the OTU ids were extracted from a biom file downloaded from QIITA (<https://qiita.ucsd.edu>).

Usage`mock_query_df`**Format**`data.frame`**Details**

Bokulich, Nicholas A., et al. "Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing." *Nature methods* 10.1 (2013): 57-59.

Source

<http://bioconductor.org/packages/release/data/experiment/html/msd16s.html>

Examples

```
data(mock_query_df)
```

show,MgDb-method	<i>Display summary of MgDb-class object</i>
------------------	---

Description

Display summary of MgDb-class object

Usage

```
## S4 method for signature 'MgDb'
show(object)
```

Arguments

object MgDb-class object

Value

MgDb-class summary

taxa_columns	<i>Column names for MgDb taxonomy slot object</i>
--------------	---

Description

Column names for MgDb taxonomy slot object

Usage

```
taxa_columns(mgdb)

## S4 method for signature 'MgDb'
taxa_columns(mgdb)
```

Arguments

mgdb object of MgDB class

Value

character vector

Note

Same function as [taxa_keytypes](#).

Examples

```
demoMgDb <- get_demoMgDb()
taxa_columns(demoMgDb)
```

taxa_keys	<i>Taxonomy values for a given keytype</i>
-----------	--

Description

Taxonomy values for a given keytype

Usage

```
taxa_keys(mgdb, keytype)

## S4 method for signature 'MgDb'
taxa_keys(mgdb, keytype)
```

Arguments

mgdb	object of MgDB class
keytype	taxonomic classification level

Value

tbl_df

Examples

```
demoMgDb <- get_demoMgDb()
taxa_keys(demoMgDb, keytype = "Phylum")
```

taxa_keytypes	<i>Column names for MgDb taxonomy slot object</i>
---------------	---

Description

Column names for MgDb taxonomy slot object

Usage

```
taxa_keytypes(mgdb)

## S4 method for signature 'MgDb'
taxa_keytypes(mgdb)
```

Arguments

mgdb object of MgDB class

Value

tbl_df

Examples

```
demoMgDb <- get_demoMgDb()
taxa_keytypes(demoMgDb)
```

taxa_levels *Accessor function for taxonomic levels in MRExperiment featureData*

Description

Accessor function for taxonomic levels in MRExperiment featureData

Usage

```
taxa_levels(obj)
```

Arguments

obj an 'MRExperiment-class' object

Value

character vector with taxonomic levels

Examples

```
data("mouseData", package = "metagenomeSeq")
taxa_levels(mouseData)
```

Index

*Topic **datasets**

- mgQuery, [11](#)
- mock_mgF, [12](#)
- mock_query_df, [12](#)

- aggregate_taxa, [2](#)
- AnnotatedDataFrame, [9](#)
- annotateFeatures, [3](#), [12](#)
- annotateFeatures, MgDb-method
(annotateFeatures), [3](#)
- annotateMRexp_fData, [4](#)
- annotateMRexp_fData, MgDb-method
(annotateMRexp_fData), [4](#)

- get_demoMgDb, [4](#)
- get_mockMgDb, [5](#)

- MgDb, [6](#)
- MgDb (MgDb-class), [5](#)
- mgdb (MgDb-class), [5](#)
- MgDb-class, [4](#), [5](#), [5](#), [12](#)
- mgdb_meta, [6](#)
- mgDb_select, [6](#)
- mgDb_select, MgDb-method (mgDb_select), [6](#)
- mgdb_seq, [7](#)
- mgdb_taxa, [8](#)
- mgdb_tree, [8](#)
- mgF_meta, [9](#)
- mgF_seq, [10](#)
- mgF_taxa, [10](#)
- mgF_tree, [11](#)
- mgFeatures, [3](#)
- mgFeatures-class, [9](#), [12](#)
- mgQuery, [11](#)
- mock_mgF, [12](#)
- mock_query_df, [12](#)

- select, MgDb-method (mgDb_select), [6](#)
- show, MgDb-method, [13](#)

- taxa_columns, [13](#)
- taxa_columns, MgDb-method
(taxa_columns), [13](#)
- taxa_keys, [14](#)
- taxa_keys, MgDb-method (taxa_keys), [14](#)

- taxa_keytypes, [14](#), [14](#)
- taxa_keytypes, MgDb-method
(taxa_keytypes), [14](#)
- taxa_levels, [15](#)