

Package ‘metagenomeFeatures’

October 16, 2018

Title Exploration of marker-gene sequence taxonomic annotations

Version 2.0.0

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), S4Vectors (>= 0.14.7), 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), ape (>= 3.5), purrr (>= 0.2.2), DECIPHER (>= 2.4.0)

Suggests knitr (>= 1.11), testthat (>= 0.10.0), rmarkdown (>= 1.3), tidyverse (>= 1.2.1), devtools (>= 1.13.5), ggtree (>= 1.8.2)

Collate 'utils.R' 'phylo-class.R' 'DNAStringSetOrNull-class.R' 'mgDb-class.R' 'mgFeatures-class.R' 'gg13.8_85MgDb.R' 'mgDb-taxa_accessors.R' 'mgDb_method_select.R' 'mgDb_method_annotateFeatures.R' 'mock_query_df.R' 'mock_mgFR'

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

git_url <https://git.bioconductor.org/packages/metagenomeFeatures>

git_branch RELEASE_3_7

git_last_commit c295899

git_last_commit_date 2018-04-30

Date/Publication 2018-10-15

R topics documented:

annotateFeatures	2
get_gg13.8_85MgDb	3
MgDb-class	3
mgDb_	4
mgDb_select	4
mgFeatures	5
mgFeatures-class	6
mgF_	6
mock_mgF	7
mock_query_df	8
newMgDb	8
show,MgDb-method	9
taxa_	10

Index	11
--------------	-----------

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
gg85 <- get_gg13.8_85MgDb()
## generating mgFeatures object
data(mock_query_df)
mock_mgF <- annotateFeatures(gg85, mock_query_df)
```

get_gg13.8_85MgDb	<i>MgDb-class object with Greengenes Database Version 13.8 85% OTUs</i>
-------------------	---

Description

Example [MgDb-class](#) object with Greengenes Database Version 13.8 85% OTUs.

Usage

```
get_gg13.8_85MgDb()
```

Value

MgDb-class object

Examples

```
get_gg13.8_85MgDb()
```

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

Description

The MgDb-class object contains sequence, taxonomic data, and a phylogenetic tree (optional) for a 16S rRNA taxonomic database, see the **greengenes13.5MgDb** package as an example database. The `get_gg13.8_85MgDb()` function in **metagenomeFeatures** exports a small subset of the database in the **greengenes13.5MgDb** annotation package as an example MgDb-class object.

Value

MgDb-class object

Slots

seq database reference sequences
 tree reference phylogenetic tree
 taxa database taxonomy
 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, Greengenes 13.8 85% OTUs database.
gg85 <- get_gg13.8_85MgDb()
```

mgDb_ *MgDb-class accessors*

Description

Accessors for **MgDb**-class object slots. mgDb_seq - sequence slot, mgDb_taxa - taxa slot, mgDb_tree - phylogenetic tree slot, and mgDb_meta - metadata slot.

Usage

```
mgDb_tree(mgdb)
mgDb_seq(mgdb)
mgDb_taxa(mgdb)
mgDb_meta(mgdb)
```

Arguments

mgdb *MgDb-class object.*

Value

appropriate class object for the slot accessed

Examples

```
gg85 <- get_gg13.8_85MgDb()
mgDb_seq(gg85)
mgDb_taxa(gg85)
mgDb_tree(gg85)
mgDb_meta(gg85)
```

mgDb_select *Querying MgDb objects*

Description

Function for querying **MgDb-class** 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, DNAS-trngSet, and phylo.

Examples

```
gg85 <- get_gg13.8_85MgDb()
# select taxa only
mgDb_select(gg85, type = "taxa",
            keys = c("Vibrionaceae", "Enterobacteriaceae"),
            keytype = "Family")

# select seq only
mgDb_select(gg85, type = "seq",
            keys = c("Vibrionaceae", "Enterobacteriaceae"),
            keytype = "Family")

# select all taxa, seq, and tree
mgDb_select(gg85, type = "all",
            keys = c("Vibrionaceae", "Enterobacteriaceae"),
            keytype = "Family")
```

mgFeatures

mgFeatures-class constructor

Description

mgFeatures-class constructor

Usage

```
mgFeatures(taxa, tree = NULL, seq = NULL, metadata)
```

Arguments

taxa	a DataFrame-class or object that can be coerced into a DataFrame
tree	a phylo-class object with phylogenetic tree
seq	DNAStrngSet-object with feature sequences
metadata	a list

Value

mgFeatures-class object

Examples

```
mgFeatures(taxa = data.frame(), metadata = list())
```

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

Description

Object contains taxonomic annotation and reference sequence data for classified OTUs. The class extends the [DataFrame](#) 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 DNAStrngSetOrNull
 refDbTree phyloOrNull

Examples

```
data(mock_mgF)
```

mgF_	<i>mgFeatures accessors</i>
------	-----------------------------

Description

Accessors for [mgFeatures](#)-class object slots. mgF_seq - refDbSeq slot, mgF_taxa - taxa slot, mgF_tree - phylogenetic tree slot, and mgF_meta - metadata slot.

Usage

```
mgF_tree(mgF)
```

```
mgF_seq(mgF)
```

```
mgF_taxa(mgF)
```

```
mgF_meta(mgF)
```

Arguments

mgF mgFeatures-class object.

Value

appropriate class object for the slot accessed

Examples

```
data(mock_mgF)
mgF_seq(mock_mgF)
mgF_taxa(mock_mgF)
mgF_tree(mock_mgF)
mgF_meta(mock_mgF)
```

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.

Source

<https://qiita.ucsd.edu>

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 `annotateFeaturesMgDb-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

<https://qiita.ucsd.edu>

Examples

```
data(mock_query_df)
```

newMgDb

MgDb

Description

MgDb

Usage

```
newMgDb(db_file, tree, metadata)
```

Arguments

db_file	SQLite filename with database taxonomy and sequence data
tree	newick filename with database tree data
metadata	list with database metadata

Value

MbDb class object

Examples

```
metadata_file <- system.file("extdata", 'gg13.8_85_metadata.RData',
  package = "metagenomeFeatures")
load(metadata_file)

gg_db_file <- system.file("extdata", 'gg13.8_85.sqlite',
  package = "metagenomeFeatures")

gg_tree_file <- system.file("extdata", "gg13.8_85.tre",
  package = "metagenomeFeatures")

## Creating a new MgDb class object with gg13.8_85 data
newMgDb(db_file = gg_db_file,
  tree = gg_tree_file,
  metadata = metadata)
```

show,MgDb-method

Display summary of MgDb-class object

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_ *MgDb-class Taxa slot helper functions*

Description

Helper functions for for [MgDb](#)-class taxa slot. taxa_columns - taxa slot column names, taxa_keytypes - taxa slot keytypes (values used with taxa_columns and mgDb_select functions), and taxa_keys - database values for a specific keytype.

Column names for MgDb taxonomy slot object

Usage

```
taxa_keys(mgdb, keytype)

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

taxa_columns(mgdb)

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

taxa_keytypes(mgdb)

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

Arguments

mgdb MgDb-class object.
keytype character string specifying keys to return

Value

appropriate class object for the slot accessed

Examples

```
gg85 <- get_gg13.8_85MgDb()
taxa_columns(gg85)
taxa_keytypes(gg85)
taxa_keys(gg85, keytype = "Phylum")
```

Index

*Topic **datasets**

mock_mgF, 7

mock_query_df, 8

annotateFeatures, 2, 7, 8

annotateFeatures, MgDb-method
(annotateFeatures), 2

DataFrame, 6

get_gg13.8_85MgDb, 3

MgDb, 4, 10

mgdb (MgDb-class), 3

MgDb-class, 3, 3, 7, 8

mgDb_, 4

mgDb_meta (mgDb_), 4

mgDb_select, 4

mgDb_select, MgDb-method (mgDb_select), 4

mgDb_seq (mgDb_), 4

mgDb_taxa (mgDb_), 4

mgDb_tree (mgDb_), 4

mgF_, 6

mgF_meta (mgF_), 6

mgF_seq (mgF_), 6

mgF_taxa (mgF_), 6

mgF_tree (mgF_), 6

mgFeatures, 2, 5, 6

mgFeatures-class, 6, 7, 8

mock_mgF, 7

mock_query_df, 8

newMgDb, 8

select, MgDb-method (mgDb_select), 4

show, MgDb-method, 9

taxa_, 10

taxa_columns (taxa_), 10

taxa_columns, MgDb-method (taxa_), 10

taxa_keys (taxa_), 10

taxa_keys, MgDb-method (taxa_), 10

taxa_keytypes (taxa_), 10

taxa_keytypes, MgDb-method (taxa_), 10