Package ‘metagenomeFeatures’

April 16, 2019

Title Exploration of marker-gene sequence taxonomic annotations

Version 2.2.3

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 (>= 3.5), 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), DECIPHER (>= 2.4.0)

Suggests knitr (>= 1.11), testthat (>= 0.10.0), rmarkdown (>= 1.3), devtools (>= 1.13.5), ggtree(>= 1.8.2), BiocStyle (>= 2.8.2), phyloseq (>= 1.24.2), forcats (>= 0.3.0), ggplot2 (>= 3.0.0)

Collate 'utils.R' 'phylo-class.R' 'qiita_study_94_gg_ids.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_mgF.R'

VignetteBuilder knitr

URL `https://github.com/HCBravoLab/metagenomeFeatures`

BugReports `https://github.com/HCBravoLab/metagenomeFeatures/issues`

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

NeedsCompilation no

RoxygenNote 6.1.1

Encoding UTF-8

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

git_branch RELEASE_3_8

git_last_commit b33bd6b

git_last_commit_date 2019-02-25

Date/Publication 2019-04-15

R topics documented:

`annotateFeatures` *Annotating metagenome data with taxonomic information*

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 databse 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` *MgDb-class object with Greengenes Database Version 13.8 85% OTUs*

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-accessor` *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-class` *Metagenome Database class*

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-methods` *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

Examples

```
gg85 <- get_gg13.8_85MgDb()
show(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 databse to return, all by default

Value

returned object depends on type: 'taxa' - dataframe with taxa information; 'seq' - DNAStringSet with seqeunce data; 'tree' - phylogenetic tree of class phylo; 'all' - list with the dataframe, DNAStringSet, 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-accessors` *mgFeatures accessors*

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)
```

`mgFeatures-class` *Class mgFeature*

Description

Class defines 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.

Usage

```
mgFeatures(taxa = data.frame(), 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`	DNAStringSet-object with feature sequences
`metadata`	a list

Value

mgFeature class object

mgFeatures-class object

Examples

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

`mgFeatures-methods` *mgFeatures-class subset method*

Description

mgFeatures-class subset method

Usage

```
## S4 method for signature 'mgFeatures'
x[i, j, ..., drop = FALSE]
```

Arguments

`x`	Object to extract elements
`i, j`	element indices to extract or replace
`...`	other parameters to subset function
`drop`	default to FALSE

`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 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

`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)
```

`qiita_study_94_gg_ids` *QIITA Dataset For this vignette we are using 16S rRNA data from Rousk et al. 2010, a soil microbiome study, https://qiita.ucsd.edu/study/description/94. A BIOM and qiime mapping file for the study can be obtained from QIITA. A vector of Greengenes for the study cluster centers is included in this package for use in this vignette.*

Description

data(qiita_study_94_gg_ids)

Usage

`qiita_study_94_gg_ids`

Format

An object of class `character` of length 2305.

`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.

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