

Package ‘metagenomeFeatures’

March 7, 2016

Title Exploration of marker-gene sequence taxonomic annotations

Version 1.1.0

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 (>= 3.2), Biobase (>= 2.17.8)

License Artistic-2.0

LazyData true

Imports Biostrings (>= 2.36.4), ShortRead (>= 1.26.0), dplyr (>= 0.4.3), stringr (>= 1.0.0), lazyeval (>= 0.1.10), RSQLite (>= 1.0.0), magrittr (>= 1.5), methods, lattice

Collate 'metagenomeAnnotation-class.R' 'mgDb-class.R' 'demoMgDb.R' 'mgDb-taxa_accessors.R' 'mgQuery.R' 'msd16s_metagenomeAnnotation.R' 'msd16s_query_df.R' 'utils.R'

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

VignetteBuilder knitr

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

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

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

NeedsCompilation no

R topics documented:

annotate	2
get_demoMgDb	3

metagenomeAnnotation-class	3
MgDb-class	4
mgQuery	5
msd16s_metagenomeAnnotation	5
msd16s_query_df	6
select	6
show,MgDb-method	7
split_by	8
taxa_columns	8
taxa_keys	9
taxa_keytypes	10

Index 11

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

Description

This method is used to create a [metagenomeAnnotation](#) class object with user supplied taxonomic assignments and [MgDb-class](#) object. As input users can provide a vector with database ids, a data.frame with database ids as well as count data for different samples as columns along with a column of database ids named Keys, additionally a [DNAStringSet](#) object can be passed with experimental sequence data. If experimental sequence data are provided, database ids must be passed as a data.frame and include a column SeqIDs with sequence names as well as database ids.

Usage

```
annotate(mgdb, ...)

## S4 method for signature 'MgDb'
annotate(mgdb, db_keys = NULL, query_df = NULL,
         query_seq = NULL, mapping = "user provided ids")
```

Arguments

mgdb	MgDb class object
...	additional arguments passed to select function
db_keys	(Optional) vector of database Keys of entries to include in metagenomeAnnotation class object
query_df	(Optional) data frame with experimental data to annotate with taxonomic information, must include column named "Key" with database ids.
query_seq	(Optional) DNAStringSet object sequences
mapping	(Optional) method used to map sequences to database, default "user provided", use for documenting methods used to perform the taxonomic assignment.

Value

metagenomeAnnotation-class object

Note

Must include either db_keys or query_df as argument.

Examples

```
# see vignette
```

get_demoMgDb	<i>Example MgDb-class object</i>
--------------	----------------------------------

Description

Example [MgDb-class](#) object with 249 entries from the Greengenes 13.5 database.

Usage

```
get_demoMgDb()
```

Value

MgDb-class object

Examples

```
get_demoMgDb()
```

metagenomeAnnotation-class	<i>metagenomeAnnotation-class object</i>
----------------------------	--

Description

Object contains taxonomic annotation and sequence data for an experiment . The class extends the [AnnotatedDataFrame](#) class with a slot for metadata including information on the database source and methods used to perform the taxonomic assignment. Additionally, the user can include experiment sequence data in the object for use in downstream analysis.

Value

metagenomeAnnotation class object

Slots

```
metadata list
experimentSeqData DNASTringSet
```

Examples

```
data(msd16s_metagenomeAnnotation)
```

MgDb-class

Metagenome Database class

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
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()
```

`mgQuery`*Example ShortRead 16S experiment dataset*

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)`

`msd16s_metagenomeAnnotation`*Example metagenomeAnnotation object*

Description

This example [metagenomeAnnotation-class](#) object includes [msd16s](#) 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`msd16s_metagenomeAnnotation`**Format**

A metagenomeAnnotation class object.

Source

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

Examples

```
data(msd16s_metagenomeAnnotation)
```

msd16s_query_df	<i>Example Query Data Frame</i>
-----------------	---------------------------------

Description

Example query_df for use in generating a [metagenomeAnnotation-class](#) object using the [annotate MgDb-class](#) method. The dataset contains the msd16s OTU ids and Greengenes database version 13.5 ids. Specifically, the OTU centers came from 'Healthy and moderate to severe diarrhea 16S expression data' - bioconductor dataset 'msd16s'.

Usage

```
msd16s_query_df
```

Format

```
data.frame
```

Source

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

Examples

```
data(msd15s_query_df)
```

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

```
select(mgdb, type, ...)

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

Arguments

mgdb	MgDb class object
type	either "taxa", "seq", or "both". "taxa" and "seq" only queries the taxonomy and sequences databases respectively. "both" queries both the taxonomy and sequence database.
...	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

generates database, function does not return anything

Examples

```
demoMgDb <- get_demoMgDb()
# select taxa only
select(demoMgDb, type = "taxa",
       keys = c("Vibrio", "Salmonella"),
       keytype = "Genus")

# select seq only
select(demoMgDb, type = "seq",
       keys = c("Vibrio", "Salmonella"),
       keytype = "Genus")

# select both taxa and seq
select(demoMgDb, type = "both",
       keys = c("Vibrio", "Salmonella"),
       keytype = "Genus")
```

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
--------	-------------------

split_by	<i>Split metagenomeAnnotation Object by Taxa</i>
----------	--

Description

Function takes a [metagenomeAnnotation-class](#) object and splits it based on a user provided taxonomic level, returning a list of [metagenomeAnnotation-class](#) objects for each unique taxa at that level.

Usage

```
split_by(mgAnno, taxa_level)

## S4 method for signature 'metagenomeAnnotation'
split_by(mgAnno, taxa_level)
```

Arguments

mgAnno	metagenomeAnnotation class object
taxa_level	taxonomic level used to split the metagenomeAnnotation object at

Value

list of metagenomeAnnotation objects

Methods (by class)

- metagenomeAnnotation:

Examples

```
data(msd16s_metagenomeAnnotation)
split_by(msd16s_metagenomeAnnotation, "Phylum")
```

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

tbl_df

Methods (by class)

- MgDb:

Note

Same function as [taxa_keytypes](#).

See Also

Other MgDb_accessors: [taxa_keys](#), [taxa_keys, MgDb-method](#); [taxa_keytypes](#), [taxa_keytypes, MgDb-method](#)

Examples

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

taxa_keys

Taxonomy values for a given keytype

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

Methods (by class)

- MgDb:

See Also

Other MgDb_accessors: [taxa_columns](#), [taxa_columns](#), [MgDb-method](#); [taxa_keytypes](#), [taxa_keytypes](#), [MgDb-method](#)

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

Methods (by class)

- MgDb:

See Also

Other MgDb_accessors: [taxa_columns](#), [taxa_columns](#), [MgDb-method](#); [taxa_keys](#), [taxa_keys](#), [MgDb-method](#)

Examples

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

Index

*Topic **datasets**

- mgQuery, [5](#)
- msd16s_metagenomeAnnotation, [5](#)
- msd16s_query_df, [6](#)

annotate, [2](#), [6](#)

annotate, MgDb-method (annotate), [2](#)

AnnotatedDataFrame, [3](#)

DNAStrngSet, [2](#)

get_demoMgDb, [3](#)

metagenomeAnnotation, [2](#)

metagenomeAnnotation-class, [3](#), [5](#), [6](#), [8](#)

MgDb, [6](#)

MgDb (MgDb-class), [4](#)

mgdb (MgDb-class), [4](#)

MgDb-class, [2](#), [3](#), [4](#), [6](#)

mgQuery, [5](#)

msd16s, [5](#)

msd16s_metagenomeAnnotation, [5](#)

msd16s_query_df, [6](#)

select, [6](#)

select, MgDb-method (select), [6](#)

show, MgDb-method, [7](#)

split_by, [8](#)

split_by, metagenomeAnnotation-method
(split_by), [8](#)

taxa_columns, [8](#), [10](#)

taxa_columns, MgDb-method
(taxa_columns), [8](#)

taxa_keys, [9](#), [9](#), [10](#)

taxa_keys, MgDb-method (taxa_keys), [9](#)

taxa_keytypes, [9](#), [10](#), [10](#)

taxa_keytypes, MgDb-method
(taxa_keytypes), [10](#)