Package ‘metagenomeFeatures’

January 30, 2017

Title Exploration of marker-gene sequence taxonomic annotations

Version 1.4.0

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 (>= 3.3), 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 (>= 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), markdown

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 5.0.1
aggregate_taxa

Aggregates a MRexperiment object and returns either an aggregated MRexperiment or counts matrix to a user defined taxonomic level.

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.
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_key)

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.

Value

mgFeatures-class object

Note

Must include either db_keys or query_df as argument.

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**

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

## S4 method for signature 'MgDb'

```r
annotateMRexp_fData(mgdb, MRobj)
```

**Arguments**

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

**Value**

MRexperiment-class object

**Examples**

```r
# 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**

```r
get_demoMgDb()
```

**Value**

MgDb-class object

**Examples**

```r
get_demoMgDb()
```
get_mockMgDb

Mock MgDb-class object

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

Value
MgDb-class object

Examples
get_mockMgDb()

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
tree reference phylogenetic tree
taxa_file name of sqlite db
tree_file name of phylogenetic tree file
metadata associated metadata for the database
**mgDb_select**

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

**Examples**

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

---

**mgdb_meta**

*MgDb metadata slot accessor*

**Description**
MgDb metadata slot accessor

**Usage**

```r
mgdb_meta(mgdb)
```

**Arguments**

- `mgdb` MgDb class object

**Value**

list

**Examples**

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

---

**mgDb_select**

*Querying MgDb objects*

**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**

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

### S4 method for signature 'MgDb'

```r
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, for 'taxa' - dataframe with taxa information, 'seq' a DNAStringSet with sequence data, 'tree' a phylogenetic tree of class phylo, 'all' a list with the dataframe, DNAStringSet, 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

DNAStringSet 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()
mgd_tree(demoMgDb)

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 DNAStringSet
refDbTree phyloOrNULL

Examples

data(mock_mgF)

mgF_meta

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

**Description**
mgFeatures refDbSeq slot accessor

**Usage**
mgF_seq(mgF)

**Arguments**
- mgF: mgFeatures class object

**Value**
DNAStringSet class object

**Examples**
```r
data(mock_mgF)
mgF_seq(mock_mgF)
```

### mgF_taxa

**Description**
mgFeatures taxa slot accessor

**Usage**
mgF_taxa(mgF)

**Arguments**
- mgF: mgFeatures class object

**Value**
AnnotatedDataFrame

**Examples**
```r
data(mock_mgF)
mgF_taxa(mock_mgF)
```
**mgF_tree**

**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**  
**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)
```
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**


**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**

show,MgDb-method

**Source**

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

**Examples**

data(mock_query_df)

demoMgDb <- get_demoMgDb()
taxa_columns(demoMgDb)

---

**show,MgDb-method**

*Display summary of MgDb-class object*

**Description**

Display summary of MgDb-class object

**Usage**

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

**Arguments**

- `object` : MgDb-class object

---

**taxa_columns**

*Column names for MgDb taxonomy slot object*

**Description**

Column names for MgDb taxonomy slot object

**Usage**

```r
taxa_columns(mgdb)
```

**Arguments**

- `mgdb` : object of MgDB class

**Note**

Same function as `taxa_keytypes`.

**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)
```

**Arguments**

- `mgdb`: object of MgDB class
- `keytype`: taxonomic classification level

**Value**

`tbl_df`

**Examples**

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

---

**taxa_keytypes**

*Column names for MgDb taxonomy slot object*

**Description**

Column names for MgDb taxonomy slot object

**Usage**

```
taxa_keytypes(mgdb)
```

**Arguments**

- `mgdb`: object of MgDB class

**Value**

`tbl_df`

**Examples**

```r
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

# not run
# 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, 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_keys, 14
taxa_keytypes, 13, 14
taxa_levels, 15