

# Package ‘metagenomeFeatures’

November 3, 2020

**Title** Exploration of marker-gene sequence taxonomic annotations

**Version** 2.10.0

**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 (>= 3.5), Biobase (>= 2.17.8)

**License** Artistic-2.0

**LazyData** true

**Imports** Biostrings (>= 2.36.4), S4Vectors (>= 0.23.18), 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** 7.0.2

**Encoding** UTF-8

**git\_url** <https://git.bioconductor.org/packages/metagenomeFeatures>

**git\_branch** RELEASE\_3\_12

**git\_last\_commit** 880eaad

**git\_last\_commit\_date** 2020-10-27

**Date/Publication** 2020-11-02

## R topics documented:

annotateFeatures . . . . .	2
get_gg13.8_85MgDb . . . . .	3
MgDb-accessor . . . . .	3
MgDb-class . . . . .	4
MgDb-methods . . . . .	5
mgDb_select . . . . .	5
mgFeatures-accessors . . . . .	6
mgFeatures-class . . . . .	7
mgFeatures-methods . . . . .	8
mock_mgF . . . . .	8
mock_query_df . . . . .	9
newMgDb . . . . .	9
qiita_study_94_gg_ids . . . . .	10
taxa_ . . . . .	10
<b>Index</b>	<b>12</b>

---

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-accessor	<i>MgDb-class accessors</i>
---------------	-----------------------------

---

**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 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-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	<i>Class mgFeature</i>
------------------	------------------------

---

**Description**

Class defines taxonomic annotation and reference sequence data for classified OTUs. The class indirectly 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	DNAStrngSet-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	<i>Example Query Data Frame</i>
---------------	---------------------------------

---

**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	<i>MgDb</i>
---------	-------------

---

**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 Green-genes 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

## \* datasets

- mock\_mgF, 8
- mock\_query\_df, 9
- qiita\_study\_94\_gg\_ids, 10
- [,mgFeatures-method (mgFeatures-methods), 8

- annotateFeatures, 2, 8, 9
- annotateFeatures,MgDb-method (annotateFeatures), 2

- DataFrame, 7

- get\_gg13.8\_85MgDb, 3

- MgDb, 3, 10

- mgdb (MgDb-class), 4
- MgDb-accessor, 3
- MgDb-class, 3, 4, 8, 9
- MgDb-methods, 5
- mgDb\_meta (MgDb-accessor), 3
- mgDb\_select, 5
- mgDb\_select, MgDb-method (mgDb\_select), 5
- mgDb\_seq (MgDb-accessor), 3
- mgDb\_taxa (MgDb-accessor), 3
- mgDb\_tree (MgDb-accessor), 3
- mgF\_meta (mgFeatures-accessors), 6
- mgF\_seq (mgFeatures-accessors), 6
- mgF\_taxa (mgFeatures-accessors), 6
- mgF\_tree (mgFeatures-accessors), 6
- mgFeatures, 2, 6
- mgFeatures (mgFeatures-class), 7
- mgFeatures-accessors, 6
- mgFeatures-class, 7, 8, 9
- mgFeatures-methods, 8
- mock\_mgF, 8
- mock\_query\_df, 9

- newMgDb, 9

- qiita\_study\_94\_gg\_ids, 10

- select, MgDb-method (mgDb\_select), 5
- show, MgDb-method (MgDb-methods), 5

- 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