

Package ‘biom’

July 2, 2014

Version 0.3.12

Date 2014-02-23

Maintainer Paul J. McMurdie <mcmurdie@stanford.edu>

License GPL-2

Title An interface package (beta) for the BIOM file format.

Type Package

Author Paul J. McMurdie <mcmurdie@stanford.edu> and the
biom-format team <http://biom-format.org/>

BugReports <https://github.com/joey711/biom/issues>

URL <https://github.com/joey711/biom/>, <http://biom-format.org/>

Description This is an R package for interfacing with the BIOM format. This package includes basic tools for reading biom-format files, accessing and subsetting data tables from a biom object (which is more complex than a single table), as well as limited support for writing a biom-object back to a biom-format file. The design of this API is intended to match the python API and other tools included with the biom-format project, but with a decidedly “R flavor” that should be familiar to R users. This includes S4 classes and methods, as well as extensions of common core functions/methods.

Imports plyr (>= 1.8), RJSONIO (>= 1.0.3), Matrix (>= 1.0.12)

Depends R (>= 3.0.0), methods

Suggests testthat (>= 0.7.1), knitr (>= 1.3)

VignetteBuilder knitr

Collate 'allClasses.R' 'allPackage.R' 'IO-methods.R' 'BIOM-class.R' 'validity-methods.R'

NeedsCompilation no

Repository CRAN

Date/Publication 2014-02-24 01:20:23

R topics documented:

biom-package	2
biom	3
biom-class	4
biom_data	5
biom_shape	6
colnames	7
header	7
matrix_element_type	8
ncol	9
nrow	9
observation_metadata	10
read_biom	11
rownames	12
sample_metadata	13
show	14
write_biom	15
Index	16

 biom-package

This is an R package for interfacing with the biom format.

Description

This is an R package for interfacing with the biom-format. It includes basic utilities for reading and writing BIOM format using R, and native R methods for interacting with biom-data that maps to functionality in other languages that support biom-format, like python.

Author(s)

Paul J. McMurdie II <mcmurdie@stanford.edu>

References

<http://www.biom-format.org/>

<https://github.com/joey711/biom>

<http://cran.r-project.org/web/packages/biom/index.html>

`biom`*Build and return an instance of the biom-class.*

Description

This is for instantiating a biom object within R ([biom-class](#)), and assumes relevant data is already available in R. This is different than reading a biom file into R. If you are instead interested in importing a biom file into R, you should use the [read_biom](#) function. This function is made available (exported) so that advanced-users/developers can easily represent analogous data in this structure if needed. However, most users are expected to instead rely on the [read_biom](#) function for data import, followed by accessor functions that extract R-friendly subsets of the data stored in the biom-format derived list.

Usage

```
biom(x)
```

Arguments

`x` (REQUIRED). A named list conforming to conventions arising from the [fromJSON](#) function reading a biom-format file with default settings. See [read_biom](#) for more details about data import and [biom-class](#) for more details about accessor functions that extract R-friendly subsets of the data and metadata stored in `x`.

Details

`biom()` is a constructor method. This is the main method suggested for constructing an experiment-level ([biom-class](#)) object from its component data.

Value

An instance of the [biom-class](#).

See Also

The [read_biom](#) import function.
Accessor functions like [header](#).

Examples

```
#
# import with default parameters, specify a file
biom_file = system.file("extdata", "rich_sparse_otu_table.biom", package = "biom")
x = read_biom(biom_file)
show(x)
print(x)
header(x)
biom_data(x)
```

```
biom_shape(x)
nrow(x)
ncol(x)
observation_metadata(x)
sample_metadata(x)
```

biom-class

The biom format data class.

Description

This class inherits from the [list-class](#), with validity checks specific to the definition to the biom-format. Effectively this means the list must have certain index names, some elements of which must have a specific structure or value. For further details see [the biom-format definition](#). Importantly, this means other special properties of lists, like operations with `$` and single- or double-square-braces are also supported; as-is the `apply`-family function that can operate on lists. Note that some features of the biom-format can be essentially empty, represented by the string "null" in the file. These fields are returned as `NULL` when accessed by an accessor function.

See Also

The constructor, [biom](#)

Accessor functions:

[header](#), [biom_shape](#), [nrow](#), [ncol](#), [matrix_element_type](#), [biom_data](#), [observation_metadata](#), [sample_metadata](#)

Examples

```
biom_file = system.file("extdata", "rich_sparse_otu_table.biom", package = "biom")
x = read_biom(biom_file)
header(x)
biom_shape(x)
nrow(x)
ncol(x)
rownames(x)
colnames(x)
matrix_element_type(x)
biom_data(x)
observation_metadata(x)
sample_metadata(x)
```

biom_data	<i>Access main data observation matrix data from biom-class.</i>
-----------	--

Description

Retrieve and organize main data from [biom-class](#), represented as a matrix with index names.

Usage

```
biom_data(x, rows, columns, parallel=FALSE)
```

Arguments

x	(Required). An instance of the biom-class .
rows	(Optional). The subset of row indices described in the returned object. For large datasets, specifying the row subset here, rather than after creating the whole matrix first, can improve speed/efficiency. Can be vector of index numbers (numeric-class) or index names (character-class).
columns	(Optional). The subset of column indices described in the returned object. For large datasets, specifying the column subset here, rather than after creating the whole matrix first, can improve speed/efficiency. Can be vector of index numbers (numeric-class) or index names (character-class).
parallel	(Optional). Logical. Whether to perform the accession parsing using a parallel-computing backend supported by the plyr-package via the foreach-package . Note: At the moment, the header accessor does not need nor does it support parallel-computed parsing.

Value

A matrix containing the main observation data, with index names. The type of data (numeric or character) will depend on the results of [matrix_element_type\(x\)](#). The class of the matrix returned will depend on the sparsity of the data, and whether it has numeric or character data. For now, only numeric data can be stored in a [Matrix-class](#), which will be stored sparsely, if possible. Character data will be returned as a vanilla [matrix-class](#).

Examples

```
min_dense_file = system.file("extdata", "min_dense_otu_table.biom", package = "biom")
min_sparse_file = system.file("extdata", "min_sparse_otu_table.biom", package = "biom")
rich_dense_file = system.file("extdata", "rich_dense_otu_table.biom", package = "biom")
rich_sparse_file = system.file("extdata", "rich_sparse_otu_table.biom", package = "biom")
min_dense_file = system.file("extdata", "min_dense_otu_table.biom", package = "biom")
rich_dense_char = system.file("extdata", "rich_dense_char.biom", package = "biom")
rich_sparse_char = system.file("extdata", "rich_sparse_char.biom", package = "biom")
# Read the biom-format files
x1 = read_biom(min_dense_file)
x2 = read_biom(min_sparse_file)
```

```
x3 = read_biom(rich_dense_file)
x4 = read_biom(rich_sparse_file)
x5 = read_biom(rich_dense_char)
x6 = read_biom(rich_sparse_char)
# Extract the data matrices
biom_data(x1)
biom_data(x2)
biom_data(x3)
biom_data(x4)
biom_data(x5)
biom_data(x6)
```

biom_shape

The matrix dimensions of a [biom-class](#) object.

Description

The matrix dimensions of a [biom-class](#) object.

Usage

```
biom_shape(x)
```

Arguments

x (Required). An instance of the [biom-class](#).

Value

A length two [integer-class](#) vector indicating the `nrow` and `ncol` of the main data matrix stored in x.

See Also

[biom-class](#)

Examples

```
## # import with default parameters, specify a file
biom_file = system.file("extdata", "rich_sparse_otu_table.biom", package = "biom")
(x = read_biom(biom_file) )
biom_shape(x)
```

colnames	<i>Method extensions to colnames for biom-class objects.</i>
----------	--

Description

See the general documentation of [colnames](#) method for expected behavior.

Arguments

x (Required). An instance of the [biom-class](#).

Value

The number of columns in x. A length 1 [integer-class](#).

See Also

[nrow](#)
[colnames](#)
[biom_shape](#)

Examples

```
### import with default parameters, specify a file
biom_file = system.file("extdata", "rich_sparse_otu_table.biom", package = "biom")
(x = read_biom(biom_file) )
colnames(x)
```

header	<i>Extract the header from a biom-class object as a list.</i>
--------	---

Description

Extract the header from a [biom-class](#) object as a list.

Usage

```
header(x)
```

Arguments

x (Required). An instance of the [biom-class](#).

Value

A list containing the header data. That is, all the required elements that are not the main data or index metadata.

Examples

```
biom_file = system.file("extdata", "rich_sparse_otu_table.biom", package = "biom")
x = read_biom(biom_file)
header(x)
```

matrix_element_type Access class of data in the matrix elements of a [biom-class](#) object

Description

Access class of data in the matrix elements of a [biom-class](#) object

Usage

```
matrix_element_type(x)
```

Arguments

x (Required). An instance of the [biom-class](#).

Value

A [character-class](#) string indicating the class of the data stored in the main observation matrix of x, with expected values "int", "float", "unicode".

See Also

[biom-class](#)

Examples

```
### import with default parameters, specify a file
biom_file = system.file("extdata", "rich_sparse_otu_table.biom", package = "biom")
(x = read_biom(biom_file) )
matrix_element_type(x)
```

ncol *Method extensions to ncol for biom-class objects.*

Description

See the general documentation of [ncol](#) method for expected behavior.

Usage

```
ncol(x)
```

Arguments

x (Required). An instance of the [biom-class](#).

Value

The number of columns in x. A length 1 [integer-class](#).

See Also

[nrow](#)
[ncol](#)
[biom_shape](#)

Examples

```
### import with default parameters, specify a file
biom_file = system.file("extdata", "rich_sparse_otu_table.biom", package = "biom")
(x = read_biom(biom_file) )
ncol(x)
```

nrow *Method extensions to nrow for biom-class objects.*

Description

See the general documentation of [nrow](#) method for expected behavior.

Usage

```
nrow(x)
```

Arguments

x (Required). An instance of the [biom-class](#).

Value

The number of rows in `x`. A length 1 [integer-class](#).

See Also

[ncol](#)
[nrow](#)
[biom_shape](#)

Examples

```
### import with default parameters, specify a file
biom_file = system.file("extdata", "rich_sparse_otu_table.biom", package = "biom")
(x = read_biom(biom_file) )
nrow(x)
```

`observation_metadata` *Access observation (row) meta data from [biom-class](#).*

Description

Retrieve and organize meta data from [biom-class](#), represented as a [data.frame](#) (if possible) or a list, with proper index names.

Usage

```
observation_metadata(x, rows, parallel=FALSE)
```

Arguments

<code>x</code>	(Required). An instance of the biom-class .
<code>rows</code>	(Optional). The subset of row indices described in the returned object. For large datasets, specifying the row subset here, – rather than first creating the complete data object – can improve speed/efficiency. This parameter can be vector of index numbers (numeric-class) or index names (character-class).
<code>parallel</code>	(Optional). Logical. Whether to perform the accession parsing using a parallel-computing backend supported by the plyr-package via the foreach-package .

Value

A [data.frame](#) or [list](#) containing the meta data, with index names. The precise form of the object returned depends on the metadata stored in `x`. A `data.frame` is created if possible.

Examples

```

min_dense_file = system.file("extdata", "min_dense_otu_table.biom", package = "biom")
min_sparse_file = system.file("extdata", "min_sparse_otu_table.biom", package = "biom")
rich_dense_file = system.file("extdata", "rich_dense_otu_table.biom", package = "biom")
rich_sparse_file = system.file("extdata", "rich_sparse_otu_table.biom", package = "biom")
min_dense_file = system.file("extdata", "min_dense_otu_table.biom", package = "biom")
rich_dense_char = system.file("extdata", "rich_dense_char.biom", package = "biom")
rich_sparse_char = system.file("extdata", "rich_sparse_char.biom", package = "biom")
# Read the biom-format files
x1 = read_biom(min_dense_file)
x2 = read_biom(min_sparse_file)
x3 = read_biom(rich_dense_file)
x4 = read_biom(rich_sparse_file)
x5 = read_biom(rich_dense_char)
x6 = read_biom(rich_sparse_char)
# Extract metadata
observation_metadata(x1)
observation_metadata(x2)
observation_metadata(x3)
observation_metadata(x3, 2:4)
observation_metadata(x3, 2)
observation_metadata(x3, c("GG_OTU_3", "GG_OTU_4", "whoops"))
observation_metadata(x4)
observation_metadata(x5)
observation_metadata(x6)

```

read_biom

Read a biom-format file, returning a biom-class.

Description

Import the data from a biom-format file into R, represented as an instance of the [biom-class](#); essentially a [list](#) with special constraints that map to [the biom-format definition](#).

Usage

```
read_biom(biom_file)
```

Arguments

biom_file (Required). A character string indicating the file location of the biom formatted file. This is a JSON formatted file specific to biological datasets. The format is formally defined at [the biom-format definition](#)

Details

The BIOM file format (canonically pronounced biome) is designed to be a general-use format for representing biological sample by observation contingency tables. BIOM is a recognized standard for the [Earth Microbiome Project](#) and is a [Genomics Standards Consortium](#) candidate project. Please see [the biom-format home page](#) for more details.

It is tempting to include an argument identifying the biom-format version number of the data file being imported. However, the biom-format version number is a required field in the biom-format definition. Rather than duplicate this formal specification and allow the possibility of a conflict, the version number of the biom format will be referred to only by the "format" field in the biom formatted data, or its representation in R.

Value

An instance of the biom-class.

References

<http://biom-format.org/>

See Also

The [biom](#) constructor function.
 Accessor functions like [header](#).

Examples

```
# # # import with default parameters, specify a file
biom_file <- system.file("extdata", "rich_sparse_otu_table.biom", package = "biom")
biom_file
read_biom(biom_file)
biom_file <- system.file("extdata", "min_sparse_otu_table.biom", package = "biom")
biom_file
read_biom(biom_file)
## The previous examples use system.file() because of constraints in specifying a fixed
## path within a reproducible example in a package.
## In practice, however, you can simply provide "hard-link"
## character string path to your file:
# mybiomfile <- "path/to/my/biomfile.biom"
# read_biom(mybiomfile)
```

rownames

Method extensions to rownames for biom-class objects.

Description

See the general documentation of [rownames](#) method for expected behavior.

Arguments

x (Required). An instance of the [biom-class](#).

Value

The number of columns in x. A length 1 [integer-class](#).

See Also

[nrow](#)

[rownames](#)

[biom_shape](#)

Examples

```
# # # import with default parameters, specify a file
biom_file = system.file("extdata", "rich_sparse_otu_table.biom", package = "biom")
(x = read_biom(biom_file) )
rownames(x)
```

sample_metadata	<i>Access meta data from biom-class.</i>
-----------------	--

Description

Retrieve and organize meta data from [biom-class](#), represented as a [data.frame](#) (if possible, or a list) with proper index names.

Usage

```
sample_metadata(x, columns, parallel=FALSE)
```

Arguments

x	(Required). An instance of the biom-class .
columns	(Optional). The subset of column indices described in the returned object. For large datasets, specifying the column subset here, rather than after creating the whole matrix first, can improve speed/efficiency. Can be vector of index numbers (numeric-class) or index names (character-class).
parallel	(Optional). Logical. Whether to perform the accession parsing using a parallel-computing backend supported by the plyr-package via the foreach-package .

Value

A [data.frame](#) or [list](#) containing the meta data, with index names. The precise form of the object returned depends on the metadata stored in x. A [data.frame](#) is created if possible.

Examples

```
min_dense_file = system.file("extdata", "min_dense_otu_table.biom", package = "biom")
min_sparse_file = system.file("extdata", "min_sparse_otu_table.biom", package = "biom")
rich_dense_file = system.file("extdata", "rich_dense_otu_table.biom", package = "biom")
rich_sparse_file = system.file("extdata", "rich_sparse_otu_table.biom", package = "biom")
min_dense_file = system.file("extdata", "min_dense_otu_table.biom", package = "biom")
rich_dense_char = system.file("extdata", "rich_dense_char.biom", package = "biom")
rich_sparse_char = system.file("extdata", "rich_sparse_char.biom", package = "biom")
# Read the biom-format files
x1 = read_biom(min_dense_file)
x2 = read_biom(min_sparse_file)
x3 = read_biom(rich_dense_file)
x4 = read_biom(rich_sparse_file)
x5 = read_biom(rich_dense_char)
x6 = read_biom(rich_sparse_char)
# Extract metadata
sample_metadata(x1)
sample_metadata(x2)
sample_metadata(x3)
sample_metadata(x3, 1:4)
sample_metadata(x4)
sample_metadata(x5)
sample_metadata(x6)
```

show

Method extensions to show for biom objects.

Description

See the general documentation of [show](#) method for expected behavior.

See Also

[show](#)

Examples

```
## # import with default parameters, specify a file
biom_file = system.file("extdata", "rich_sparse_otu_table.biom", package = "biom")
(x = read_biom(biom_file) )
show(x)
```

write_biom	<i>Write a biom-format file, returning a biom-class.</i>
------------	--

Description

Write a biom-format file, returning a biom-class.

Usage

```
write_biom(x, biom_file)
```

Arguments

x	(Required). A biom object that is going to be written to file as a proper biom formatted file, adhering to the biom-format definition .
biom_file	(Required). A character string indicating the file location of the biom formatted file. This is a JSON formatted file specific to biological datasets. The format is formally defined at the biom-format definition

Value

Nothing. The first argument, x, is written to a file.

References

<http://biom-format.org/>

See Also

Accessor functions like [header](#). [the biom-format home page](#)

Examples

```
biom_file <- system.file("extdata", "rich_sparse_otu_table.biom", package = "biom")
x = read_biom(biom_file)
outfile = tempfile()
write_biom(x, outfile)
y = read_biom(outfile)
identical(x, y)
```

Index

*Topic **package**

biom-package, 2

biom, 3, 4, 12

biom, list-method (biom), 3

biom-class, 4, 5–10, 12, 13

biom-package, 2

biom_data, 4, 5

biom_data, biom, ANY, character-method
(biom_data), 5

biom_data, biom, character, ANY-method
(biom_data), 5

biom_data, biom, missing, missing-method
(biom_data), 5

biom_data, biom, missing, numeric-method
(biom_data), 5

biom_data, biom, numeric, missing-method
(biom_data), 5

biom_data, biom, numeric, numeric-method
(biom_data), 5

biom_shape, 4, 6, 7, 9, 10, 13

biom_shape, biom-method (biom_shape), 6

colnames, 7, 7

colnames, biom-method (colnames), 7

data.frame, 10, 13

fromJSON, 3

header, 3, 4, 7, 12, 15

header, biom-method (header), 7

list, 10, 11, 13

matrix_element_type, 4, 5, 8

matrix_element_type, biom-method
(matrix_element_type), 8

ncol, 4, 6, 9, 9, 10

ncol, biom-method (ncol), 9

nrow, 4, 6, 7, 9, 9, 10, 13

nrow, biom-method (nrow), 9

NULL, 4

observation_metadata, 4, 10

observation_metadata, biom, character-method
(observation_metadata), 10

observation_metadata, biom, missing-method
(observation_metadata), 10

observation_metadata, biom, numeric-method
(observation_metadata), 10

read_biom, 3, 11

rownames, 12, 12, 13

rownames, biom-method (rownames), 12

sample_metadata, 4, 13

sample_metadata, biom, character-method
(sample_metadata), 13

sample_metadata, biom, missing-method
(sample_metadata), 13

sample_metadata, biom, numeric-method
(sample_metadata), 13

show, 14, 14

show, biom-method (show), 14

write_biom, 15