Package ‘MicrobiomeBenchmarkData’

May 21, 2024

Title Datasets for benchmarking in microbiome research

Version 1.6.0

Description The MicrobiomeBenchmarkData package provides functionality to access microbiome datasets suitable for benchmarking. These datasets have some biological truth, which allows to have expected results for comparison. The datasets come from various published sources and are provided as TreeSummarizedExperiment objects. Currently, only datasets suitable for benchmarking differential abundance methods are available.

License Artistic-2.0

LazyData false

Depends R (&gt;= 4.2), SummarizedExperiment, TreeSummarizedExperiment

Imports BiocFileCache, tools, S4Vectors, ape, utils

Suggests rmarkdown, knitr, BiocStyle, testthat (&gt;= 3.0.0), mia, ggplot2, tidyR, dplyr, magrittr, tibble, purrr

biocViews ExperimentData, MicrobiomeData, ReproducibleResearch, SequencingData

BugReports https://github.com/waldronlab/MicrobiomeBenchmarkData/issues


BiocType ExperimentData

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.1

VignetteBuilder knitr

Config/testthat/edition 3

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

git_branch RELEASE_3_19

git_last_commit 9b47f6c

1
Description

[assemblyTreeSummarizedExperiment] assembles a TreeSummarizedDataset taking as input the name of the dataset and the URL. This is a helper function for the getBenchmarkData function.

Usage

[assemblyTreeSummarizedExperiment](x)

Arguments

dat_name A character string with the name of the dataset.
dat_url A character string with the URL from Zenodo.

Value

A TreeSummarizedExperiment
**.getCache**

*Get cache*

**Description**

.getCache creates or loads a cache to store files downloaded through the MicrobiomeBenchmarkData package.

**Usage**

.getCache()

**Value**

A BiocFileCache object.

---

**.getResourcePath**

*Get resource path*

**Description**

.getResource downloads the count matrix and store it in the cache.

**Usage**

.getResourcePath(resource, suffix)

**Arguments**

- resource_name: A character string with the name of the dataset.
- resource_url: A character string with the URL from Zenodo.

**Value**

A character string containing the path to the count matrix in the cache.
.getSampleMetadata  Get sample metadata

Description
.getSampleMetadata returns sampleMetadata.

Usage
.getSampleMetadata()

Value
A data frame with sample metadata.

getBenchmarkData  Get dataset

Description
getBenchmarkData imports datasets as TreeSummarizedExperiment objects.

Usage
getBenchmarkData(x, dryrun = TRUE)

Arguments

x A character vector with the name(s) of the dataset(s). If empty and dryrun = TRUE, it returns a message with the names of the available datasets. If empty and dryrun = FALSE, it returns a list of TreeSummarizedExperiments with all of the datasets.

dryrun If TRUE, only returns a message and invisibly returns the names of the datasets as a character vector. If FALSE, it returns the TreeSummarizedExperiment datasets indicated in the argument 'x'.

Value
A list of TreeSummarizedExperiments when dryrun = FALSE. A data frame with the datasets characteristics when dryrun = TRUE.
MicrobiomeBenchmarkData

Examples

## Example 1
```r
datasets_names <- getBenchmarkData()
datasets_names
```

## Example 2
```r
dataset <- getBenchmarkData(
    "HMP_2012_16S_gingival_V35_subset", dryrun = FALSE
)
dataset[[1]]
```

MicrobiomeBenchmarkData

Description

The MicrobiomeBenchmarkData provide functions for accessing various microbiome datasets with biological ground truth.

removeCache

Remove cache

Description

removeCache removes all files saved in the cache.

Usage

```r
removeCache(ask = interactive())
```

Arguments

- **ask**: If TRUE, a prompt will appear asking the user to confirm removal of cache. Default value is given by the interactive function.

Value

NULL The cache and all of its contents are removed.

Examples

## Remove cache
```r
removeCache()
```
Description

A data frame with the combined metadata of all of the samples in the datasets provided through the MicrobiomeBenchmarkData package.

Usage

data("sampleMetadata", package = "MicrobiomeBenchmarkData")

Format

A data.frame.
Index

* datasets
  sampleMetadata, 6
* internal
  .assembleTreeSummarizedExperiment, 2
  .getCache, 3
  .getResourcePath, 3
  .getSampleMetadata, 4
  .assembleTreeSummarizedExperiment, 2
  .getCache, 3
  .getResourcePath, 3
  .getSampleMetadata, 4

getBenchmarkData, 2, 4

MicrobiomeBenchmarkData, 5

removeCache, 5

sampleMetadata, 6