Package 'BiocFileCache'

April 11, 2018

Title Manage Files Across Sessions

Version 1.2.3

Description This package creates a persistent on-disk cache of files that the user can add, update, and retrieve. It is useful for managing resources (such as custom Txdb objects) that are costly or difficult to create, web resources, and data files used across sessions.

Depends R (>= 3.4.0), dbplyr (>= 1.0.0)

Imports methods, stats, utils, dplyr, RSQLite, DBI, rappdirs, httr

BugReports https://github.com/Bioconductor/BiocFileCache/issues

DevelopmentURL https://github.com/Bioconductor/BiocFileCache

License Artistic-2.0

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

biocViews DataImport

VignetteBuilder knitr

Suggests testthat, knitr, BiocStyle, rmarkdown, rtracklayer

NeedsCompilation no

Author Lori Shepherd [aut, cre],
Martin Morgan [aut]

Maintainer Lori Shepherd <lori.shepherd@roswellpark.org>

R topics documented:

BiocFileCache-class 2

Index **10**

BiocFileCache-class *BiocFileCache class*

Description

This class represents the location of files stored on disk. Use the return value to add and retrieve files that persist across sessions.

Usage

```
BiocFileCache(cache = user_cache_dir(appname = "BiocFileCache"))

## S4 method for signature 'BiocFileCacheBase'
bfccache(x)

## S4 method for signature 'missing'
bfccache(x)

## S4 method for signature 'BiocFileCacheBase'
length(x)

bfcrid(x)

## S4 method for signature 'missing'
bfcrid(x)

## S4 method for signature 'BiocFileCacheReadOnly'
bfcrid(x)

## S4 method for signature 'BiocFileCache'
bfcrid(x)

## S4 method for signature 'BiocFileCache,character,missing'
x[i, j, ..., drop = TRUE]

## S4 method for signature 'BiocFileCacheReadOnly,character,missing'
x[i, j, ...,
  drop = TRUE]

## S4 method for signature 'BiocFileCache,missing,missing'
x[i, j, ..., drop = TRUE]

## S4 method for signature 'BiocFileCacheReadOnly,missing,missing'
x[i, j, ..., drop = TRUE]

## S4 method for signature 'BiocFileCacheBase,character,missing'
x[[i, j]]

## S4 replacement method for signature 'BiocFileCache,character,missing,character'
x[[i, j,
  ...]] <- value
```

```
## S4 method for signature 'missing'
bfcnew(x, rname, rtype = c("relative", "local"),
  ext = NA_character_)

## S4 method for signature 'BiocFileCache'
bfcnew(x, rname, rtype = c("relative", "local"),
  ext = NA_character_)

## S4 method for signature 'missing'
bfcadd(x, rname, fpath = rname, rtype = c("auto",
  "relative", "local", "web"), action = c("copy", "move", "asis"),
  proxy = "", ...)

## S4 method for signature 'BiocFileCache'
bfcadd(x, rname, fpath = rname, rtype = c("auto",
  "relative", "local", "web"), action = c("copy", "move", "asis"),
  proxy = "", ...)

## S4 method for signature 'missing'
bfcinfo(x, rids)

## S4 method for signature 'BiocFileCacheBase'
bfcinfo(x, rids)

## S4 method for signature 'tbl_bfc'
bfcrid(x)

## S4 method for signature 'missing'
bfcpath(x, rid)

## S4 method for signature 'BiocFileCacheBase'
bfcpath(x, rid)

## S4 method for signature 'missing'
bfcrpath(x, rnames, ..., rids)

## S4 method for signature 'BiocFileCacheBase'
bfcrpath(x, rnames, ..., rids)

## S4 method for signature 'missing'
bfcupdate(x, rids, value, ...)

## S4 method for signature 'BiocFileCache'
bfcupdate(x, rids, rname = NULL, rpath = NULL,
  fpath = NULL, proxy = "")

bfcmeta(x, name, ...) <- value

## S4 replacement method for signature 'BiocFileCacheBase'
bfcmeta(x, name, ...) <- value
```

```
## S4 method for signature 'missing'
bfcmetaremove(x, name, ...)

## S4 method for signature 'BiocFileCacheBase'
bfcmetaremove(x, name, ...)

## S4 method for signature 'missing'
bfcmetalist(x)

## S4 method for signature 'BiocFileCacheBase'
bfcmetalist(x)

## S4 method for signature 'missing'
bfcmeta(x, name, ...)

## S4 method for signature 'BiocFileCacheBase'
bfcmeta(x, name, ...)

## S4 method for signature 'missing'
bfcquerycols(x)

## S4 method for signature 'BiocFileCacheBase'
bfcquerycols(x)

## S4 method for signature 'missing'
bfcquery(x, query, field = c("rname", "rpath", "fpath"),
  ...)

## S4 method for signature 'BiocFileCacheBase'
bfcquery(x, query, field = c("rname", "rpath",
  "fpath"), ...)

## S4 method for signature 'missing'
bfccount(x)

## S4 method for signature 'BiocFileCacheBase'
bfccount(x)

## S4 method for signature 'tbl_bfc'
bfccount(x)

## S4 method for signature 'missing'
bfcneedsupdate(x, rids)

## S4 method for signature 'BiocFileCacheBase'
bfcneedsupdate(x, rids)

## S4 method for signature 'missing'
bfcdownload(x, rid, proxy = "")

## S4 method for signature 'BiocFileCache'
bfcdownload(x, rid, proxy = "")
```

```
## S4 method for signature 'missing'
bfcremove(x, rids)

## S4 method for signature 'BiocFileCache'
bfcremove(x, rids)

## S4 method for signature 'missing'
bfcsync(x, verbose = TRUE)

## S4 method for signature 'BiocFileCache'
bfcsync(x, verbose = TRUE)

## S4 method for signature 'missing'
cleanbfc(x, days = 120, ask = TRUE)

## S4 method for signature 'BiocFileCache'
cleanbfc(x, days = 120, ask = TRUE)

## S4 method for signature 'missing'
removebfc(x, ask = TRUE)

## S4 method for signature 'BiocFileCache'
removebfc(x, ask = TRUE)

## S4 method for signature 'BiocFileCacheBase'
show(object)
```

Arguments

cache	character(1) On-disk location (directory path) of cache. For default location see user_cache_dir.
x	A BiocFileCache instance or, if missing, the result of BiocFileCache().
i	character() 'rid' identifiers.
j	Ignored.
...	For 'bfcadd': For action="copy", additional arguments passed to file.copy. For 'bfcrpaths': Additional arguments passed to 'bfcadd'. For 'bfcquery': Additional arguments passed to grepl.
drop	Ignored.
value	character(1) Replacement file path.
rname	character(1) Name of object in file cache. For 'bfcupdate' a character vector of replacement rnames.
rtype	character(1) 'local', 'relative', or 'web' indicating if the resource is a local file, a relative path in the cache, or a web resource. For bfcnew: local or relative are only options. For bfcadd, the default 'auto' creates relative or web paths, based on the path prefix.
ext	character(1) A file extension to add to the local copy of the file (e.g., 'sqlite', 'txt', 'tar.gz').
fpath	For bfcadd(), character(1) path to current file location or remote web resource. If none is given, the rname is assumed to also be the path location. For bfcupdate() character() vector of replacement web resources.

action	character(1) How to handle the file: create a copy of fpath in the cache directory; move the file to the cache directory; or asis leave the file in current location but save the path in the cache. If 'rtype == "relative"', action can not be "asis".
proxy	character(1) (Optional) proxy server.
rids	character() Vector of rids.
rid	character(1) Unique resource id.
rnames	character() list of rnames to search and match on.
rpath	character() vector of replacement rpaths.
name	character(1) name of metadata table.
query	character() Regular expression pattern(s) to match in resource. It will match the pattern against fields, using & logic across query element. By default, case sensitive.
field	character() column names in resource to query, using || logic across multiple field elements. By default, matches pattern agains rname, rpath, and fpath. If exact matching, may only be a single value.
verbose	logical(1) If descriptive message and list of issues should be included as output.
days	integer(1) Number of days between accessDate and currentDate; if exceeded entry will be deleted.
ask	logical(1) Prompt if really want to remove cache and files.
object	A BiocFileCache instance.

Details

The package defines 'BiocFileCache', 'BiocFileCacheBase' and 'BiocFileCacheReadOnly' classes.

Slots unique to 'BiocFileCache' and related classes:

- 'cache': character(1), Describes the On-disk location (directory path) of the cache
- 'rid': character(), List of the unique rids in the cache.

The cache creates an RSQLite database to keep track of local and remote resources. Each item located in the database will have the following information:

- 'rid': resource id. Autogenerated. This is a unique identifier automatically generated when a resource is added to the cache
- 'rname': resource name. This is given by the user when a resource is added to the cache. It does not have to be unique and can be updated at anytime. We recommend descriptive key words and identifers.
- 'create_time': The date and time a resource is added to the cache.
- 'access_time': The date and time a resource is utilized within the cache. The access time is updated when the resource is updated or accessed
- 'rpath': resource path. This is the path to the local (on-disk) file
- 'rtype': resource type. Either "relative", "local", or "web", indicating if the resource has a remote origin
- 'fpath': If rtype is "web", this is the link to the remote resource. It will be utilized to download or update the remote data

- 'last_modified_time': For a remote resource, the last_modified (if available) information for the local copy of the data. This information is checked against the remote resource to determine if the local copy is stale and needs to be updated

All functions have a quick implementation where if the BiocFileCache object is not passed as an argument, the function uses default 'BiocFileCache()' for implementation. e.g 'bfcinfo()' can be used instead of 'bfcinfo(BiocFileCache())'. The only function this is not available for is 'bfcmeta()<-'; The BiocFileCache object must be defined as a varaible and passed as an argument. See vignette("BiocFileCache") for more details.

Value

For 'BiocFileCache': a `BiocFileCache` instance.

For 'bfccache': character(1) location of the directory containing the cache.

For 'length': integer(1) Number of objects in the file cache.

For '[': A subset of the BiocFileCache object.

For '[[': named character(1) rpath for the given resource in the cache.

For '[[<-': Updated BiocFileCache, invisibly.

For 'bfcnew': named character(1), the path to save your object / file. The name of the return value is the unique rid for the resource.

For 'bfcadd': named character(1), the path to save your object / file. The name of the character is the unique rid for the resource.

For 'bfcinfo': A `bfc_tbl` of current resources in the database.

For 'bfcpath': the file path location to load and original source information for web resources.

For 'bfcrpath': The local file path location to load.

For 'bfcupdate': an updated `BiocFileCache` object, invisibly.

For 'bfcmeta': updated BiocFileCache, invisibly

For 'bfcmetaremove': updated BiocFileCache, invisibly

For 'bfcmetalist': returns a character() of all metadata tables currently in the database. If no metadata tables are available returns character(0)

For 'bfcmeta': returns a data.frame representation of database table

For 'bfcquerycols': character() all columns in all database tables available for query.

For 'bfcquery': A `bfc_tbl` of current resources in the database whose `field` contained query. If multiple values are given, the resource must contain all of the patterns. A tbl with zero rows is returned when no resources match the query.

For 'bfccount': integer(1) Number of objects in the cache or query.

For 'bfcneedsupdate': named logical vector if resource needs to be updated. The name is the resource 'rid'. `TRUE`: fpath `modified` time of web resource more recent than in BiocFileCache; `FALSE`: fpath `modified` time of web resource not more recent than in BiocFileCache; `NA`: web resource modified time could not be determined.

For 'bfcdownload': character(1) path to downloaded resource in cache.

For 'bfcremove': updated BiocFileCache object, invisibly

For 'bfcsync': logical(1) indicating whether the cache is in sync (`TRUE`) or not. 'verbose' is TRUE by default, so descriptive messages will also be included.

For 'cleanbfc': updated BiocFileCache, invisibly.

For 'removebfc': TRUE if successfully removed.

Methods (by generic)

- bfccache: Get the location of the on-disk cache.
- length: Get the number of objects in the file cache.
- bfcrid: Get the rids of the object.
- [: Subset a BiocFileCache object.
- [[: Get a file path for select resources from the cache.
- [[<-: Set the file path of selected resources from the cache.
- bfcnew: Add a resource to the database
- bfcadd: Add an existing resource to the database
- bfcinfo: list resources in database
- bfcrid: Get the rids of the object
- bfcpath: display paths of resource
- bfcrpath: display rpath of resource. If 'rnames' is in the cache the path is returned, if it is not it will try to add it to the cache with 'bfcadd'
- bfcupdate: Update a resource in the cache
- bfcmeta<-: add meta data table in database
- bfcmetaremove: remove meta data table in database
- bfcmetalist: retrieve list of metadata table
- bfcmeta: retrieve metadata table
- bfcquerycols: Get all the possible columns to query
- bfcquery: query resource
- bfccount: Get the number of objects in the file cache or query.
- bfcneedsupdate: check if a resource needs to be updated
- bfcdownload: Redownload resource to location in cache
- bfcremove: Remove a resource to the database. If the local file is located in bfccache(x), the file will also be deleted. This will not delete information in any metadata table.
- bfcsync: sync cache and resource.
- cleanbfc: Remove old/unused files in BiocFileCache. If file to be removed is not in the bfccache location it will not be deleted.
- removebfc: Completely remove the BiocFileCache
- show: Display a BiocFileCache instance.

Examples

```
# bfc <- BiocFileCache()            # global cache
# bfc
bfc0 <- BiocFileCache(tempfile())         # temporary catch for examples
bfccache(bfc0)
length(bfc0)
path <- bfcnew(bfc0, "NewResource")
path
fl1 <- tempfile(); file.create(fl1)
bfcadd(bfc0, "Test1", fl1)                 # copy
fl2 <- tempfile(); file.create(fl2)
bfcadd(bfc0, "Test2", fl2, action="move")         # move
```

```
fl3 <- tempfile(); file.create(fl3)
add3 <- bfcadd(bfc0, "Test3", fl3, rtype="local", action="asis")  # reference
rid3 <- names(add3)

bfc0
file.exists(fl1)                                  # TRUE
file.exists(fl2)                                  # FALSE
file.exists(fl3)                                  # TRUE

# add a remote resource
url <- "http://httpbin.org/get"
bfcadd(bfc0, "TestWeb", fpath=url)
bfcinfo(bfc0)
bfcpath(bfc0, rid3)
bfcrpath(bfc0, rids = rid3)
bfcupdate(bfc0, rid3, rpath=fl3, rname="NewRname")
bfc0[[rid3]] = fl1
bfcupdate(bfc0, "BFC5", fpath="http://google.com")
meta = data.frame(list(rid = paste("BFC",seq_len(bfccount(bfc0)), sep=""),
                       num=seq(bfccount(bfc0),1,-1),
                       data=c(paste("Letter",
                       letters[seq_len(bfccount(bfc0))]))),
                  stringsAsFactors=FALSE)
bfcmeta(bfc0, name="resourcedata") <- meta
## Not run: bfcmetaremove(bfc0, "resourcedata")
bfcmetalist(bfc0)
tbl = bfcmeta(bfc0, "resourcedata")
tbl
bfcquerycols(bfc0)
bfcquery(bfc0, "Test")
bfcquery(bfc0, "^Test1$", field="rname")
bfccount(bfc0)
bfccount(bfcquery(bfc0, "test"))
bfcneedsupdate(bfc0, "BFC5")
bfcdownload(bfc0, "BFC5")
bfcremove(bfc0, rid3)
bfcinfo(bfc0)
bfcsync(bfc0)
bfcremove(bfc0, "BFC1")
bfcsync(bfc0, FALSE)
## Not run: cleanbfc(bfc, ask=FALSE)
## Not run: removebfc(bfc, ask=FALSE)
```

Index

[,BiocFileCache,character,missing-method (BiocFileCache-class), 2
[,BiocFileCache,missing,missing-method (BiocFileCache-class), 2
[,BiocFileCacheReadOnly,character,missing-method (BiocFileCache-class), 2
[,BiocFileCacheReadOnly,missing,missing-method (BiocFileCache-class), 2
[[,BiocFileCacheBase,character,missing-method (BiocFileCache-class), 2
[[<-,BiocFileCache,character,missing,character-method (BiocFileCache-class), 2

bfcadd (BiocFileCache-class), 2
bfcadd,BiocFileCache-method (BiocFileCache-class), 2
bfcadd,missing-method (BiocFileCache-class), 2
bfccache (BiocFileCache-class), 2
bfccache,BiocFileCacheBase-method (BiocFileCache-class), 2
bfccache,missing-method (BiocFileCache-class), 2
bfccount (BiocFileCache-class), 2
bfccount,BiocFileCacheBase-method (BiocFileCache-class), 2
bfccount,missing-method (BiocFileCache-class), 2
bfccount,tbl_bfc-method (BiocFileCache-class), 2
bfcdownload (BiocFileCache-class), 2
bfcdownload,BiocFileCache-method (BiocFileCache-class), 2
bfcdownload,missing-method (BiocFileCache-class), 2
bfcinfo (BiocFileCache-class), 2
bfcinfo,BiocFileCacheBase-method (BiocFileCache-class), 2
bfcinfo,missing-method (BiocFileCache-class), 2
bfcmeta (BiocFileCache-class), 2
bfcmeta,BiocFileCacheBase-method (BiocFileCache-class), 2
bfcmeta,missing-method (BiocFileCache-class), 2
bfcmeta<- (BiocFileCache-class), 2
bfcmeta<-,BiocFileCacheBase-method (BiocFileCache-class), 2
bfcmetalist (BiocFileCache-class), 2
bfcmetalist,BiocFileCacheBase-method (BiocFileCache-class), 2
bfcmetalist,missing-method (BiocFileCache-class), 2
bfcmetaremove (BiocFileCache-class), 2
bfcmetaremove,BiocFileCacheBase-method (BiocFileCache-class), 2
bfcmetaremove,missing-method (BiocFileCache-class), 2
bfcneedsupdate (BiocFileCache-class), 2
bfcneedsupdate,BiocFileCacheBase-method (BiocFileCache-class), 2
bfcneedsupdate,missing-method (BiocFileCache-class), 2
bfcnew (BiocFileCache-class), 2
bfcnew,BiocFileCache-method (BiocFileCache-class), 2
bfcnew,missing-method (BiocFileCache-class), 2
bfcpath (BiocFileCache-class), 2
bfcpath,BiocFileCacheBase-method (BiocFileCache-class), 2
bfcpath,missing-method (BiocFileCache-class), 2
bfcquery (BiocFileCache-class), 2
bfcquery,BiocFileCacheBase-method (BiocFileCache-class), 2
bfcquery,missing-method (BiocFileCache-class), 2
bfcquerycols (BiocFileCache-class), 2
bfcquerycols,BiocFileCacheBase-method (BiocFileCache-class), 2
bfcquerycols,missing-method (BiocFileCache-class), 2
bfcremove (BiocFileCache-class), 2
bfcremove,BiocFileCache-method (BiocFileCache-class), 2

bfcremove,missing-method
(BiocFileCache-class), 2
bfcrid (BiocFileCache-class), 2
bfcrid,BiocFileCache-method
(BiocFileCache-class), 2
bfcrid,BiocFileCacheReadOnly-method
(BiocFileCache-class), 2
bfcrid,missing-method
(BiocFileCache-class), 2
bfcrid,tbl_bfc-method
(BiocFileCache-class), 2
bfcrpath (BiocFileCache-class), 2
bfcrpath,BiocFileCacheBase-method
(BiocFileCache-class), 2
bfcrpath,missing-method
(BiocFileCache-class), 2
bfcsync (BiocFileCache-class), 2
bfcsync,BiocFileCache-method
(BiocFileCache-class), 2
bfcsync,missing-method
(BiocFileCache-class), 2
bfcupdate (BiocFileCache-class), 2
bfcupdate,BiocFileCache-method
(BiocFileCache-class), 2
bfcupdate,missing-method
(BiocFileCache-class), 2
BiocFileCache (BiocFileCache-class), 2
BiocFileCache-class, 2

cleanbfc (BiocFileCache-class), 2
cleanbfc,BiocFileCache-method
(BiocFileCache-class), 2
cleanbfc,missing-method
(BiocFileCache-class), 2

length,BiocFileCacheBase-method
(BiocFileCache-class), 2

removebfc (BiocFileCache-class), 2
removebfc,BiocFileCache-method
(BiocFileCache-class), 2
removebfc,missing-method
(BiocFileCache-class), 2

show,BiocFileCacheBase-method
(BiocFileCache-class), 2

user_cache_dir, 5