Package 'KEGGREST'

May 15, 2024

Version 1.44.0

Title Client-side REST access to the Kyoto Encyclopedia of Genes and Genomes (KEGG)

Depends R (>= 3.5.0)

Imports methods, httr, png, Biostrings

Suggests RUnit, BiocGenerics, knitr, markdown

Description A package that provides a client interface to the Kyoto Encyclopedia of Genes and Genomes (KEGG) REST API. Only for academic use by academic users belonging to academic institutions (see <https://www.kegg.jp/kegg/rest/>). Note that KEGGREST is based on KEGGSOAP by J. Zhang, R. Gentleman, and Marc Carlson, and KEGG (python package) by Aurelien Mazurie.

URL https://bioconductor.org/packages/KEGGREST

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

License Artistic-2.0

VignetteBuilder knitr

biocViews Annotation, Pathways, ThirdPartyClient, KEGG

RoxygenNote 7.1.1

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

git_branch RELEASE_3_19

git_last_commit 6d110a0

git_last_commit_date 2024-04-30

Repository Bioconductor 3.19

Date/Publication 2024-05-15

Author Dan Tenenbaum [aut],
Bioconductor Package Maintainer [aut, cre],
Martin Morgan [ctb],
Kozo Nishida [ctb],
Marcel Ramos [ctb],
Kristina Riemer [ctb],

Lori Shepherd [ctb],
Jeremy Volkening [ctb]

Maintainer Bioconductor Package Maintainer <maintainer@bioconductor.org>

Contents

---

keggCompounds *Get list of compounds IDs for pathway*

Description

Get list of compounds IDs for pathway.

Usage

```
keggCompounds(pathwayID)
```

Arguments

pathwayID	A KEGG pathway identifier with the prefix map and 5 digit number.

Value

A list of KEGG compound identifiers

Author(s)

Dan Tenenbaum, Kristina Riemer

References

https://www.genome.jp/kegg/pathway.html

Examples

```
keggCompounds("map00361")
```

keggConv *Convert KEGG identifiers to/from outside identifiers*

Description

Convert KEGG identifiers to/from outside identifiers.

Usage

```
keggConv(target, source, querySize = 100)
```

Arguments

target	A KEGG organism code (), T number, or one of the external databases ncbi-gi, ncbi-geneid, ncbi-proteinid, uniprot, or (for chemical substance identifiers) drug, compound, or glycan, pubchem, or chebi.
source	Same as target, but may also be a list of KEGG identifers representing internal or external names.
querySize	Empirically, KEGG limits queries to 100 source identifiers per query. This argument enables larger queries by dividing source into sub-queries of no more than querySize identifiers.

Value

A named character vector.

Author(s)

Dan Tenenbaum

References

https://www.kegg.jp/kegg/docs/keggapi.html

Examples

```
## conversion from NCBI GeneID to KEGG ID for E. coli genes
head(keggConv("eco", "ncbi-geneid"))
head(keggConv("ncbi-geneid", "eco")) ## opposite direction

## conversion from KEGG ID to NCBI GI
head(keggConv("ncbi-proteinid", c("hsa:10458", "ece:Z5100")))

## conversion from NCBI GI to KEGG ID when the organism code is not known:
head(keggConv("genes", "ncbi-geneid:3113320"))
```

keggFind *Finds entries with matching query keywords or other query data in a given database*

Description

Finds entries with matching query keywords or other query data in a given database.

Usage

```
keggFind(database, query, option = c("formula", "exact_mass",
    "mol_weight"))
```

Arguments

database	Either the name of a single KEGG database (list available via `listDatabases()`, a "T number" genome identifier, or a KEGG organism code (lists of both available via `keggList("organism")`).
query	One or more keywords, or a range of integers representing molecular weights. If `query` includes identifiers not known to KEGG, the results will not contain any information about those identifiers.
option	`Optional`. If `database` is `compound` or `drug`, `option` can be `formula`, `exact_mass`, or `weight`. Chemical formula search is a partial match irrespective of the order of atoms given. The exact mass (or molecular weight) is checked by rounding off to the same decimal place as the query data.

Value

A named character vector.

Author(s)

Dan Tenenbaum

References

https://www.kegg.jp/kegg/docs/keggapi.html

Examples

```
res <-
    keggFind("genes", c("shiga", "toxin")) ## for keywords "shiga" and "toxin"
length(res)
head(res)
res <- keggFind("genes", "shiga toxin")    ## for keywords "shiga toxin"
length(res)
head(res)
keggFind("compound", "C7H10O5", "formula") ## for chemical formula "C7H10O5"
```

```
res <- keggFind("compound", "O5C7", "formula") ## for chemical formula
                                               ## containing "O5" and "C7"
length(res)
head(res)
keggFind("compound", 174.05, "exact_mass") ## for 174.045
                                           ## =< exact mass < 174.055
res <- keggFind("compound", 300:310, "mol_weight") ## for 300 =<
                                                   ## molecular weight =< 310
length(res)
head(res)
```

---

keggGet *Retrieves given database entries*

Description

Retrieves given database entries.

Usage

```
keggGet(dbentries, option = c("aaseq", "ntseq", "mol", "kcf",
    "image", "kgml"))
```

Arguments

dbentries	One or more (up to a maximum of 10) KEGG identifiers.
option	Optional. Option governing the format of the output. aaseq is an amino acid sequence, ntseq is a nucleotide sequence. image returns an object which can be written to a PNG file, kgml returns a KGML document.

Details

Retrieves all entries from the KEGG database for a set of KEGG identifers.

keggGet() can only return 10 result sets at once (this limitation is on the server side). If you supply more than 10 inputs to keggGet(), KEGGREST will warn that only the first 10 results will be returned.

Value

A list wrapping a KEGG flat file. If option is aaseq, an AAStringSet object. If option is ntseq, a DNAStringSet object. If option is image, an object which can be written to a PNG file. If option is kgml, a KGML document.

Author(s)

Dan Tenenbaum

References

https://www.kegg.jp/kegg/docs/keggapi.html

Examples

```
res <- keggGet(c("cpd:C01290", "gl:G00092")) ## retrieves a compound entry
                                    ## and a glycan entry
str(res)
res <- keggGet(c("C01290", "G00092")) ## same as above, without prefixes
str(res)
res <- keggGet(c("hsa:10458", "ece:Z5100")) ## retrieves a human gene entry
                                    ## and an E.coli O157 gene entry
str(res)
res <- keggGet(c("hsa:10458", "ece:Z5100"), "aaseq") ## retrieves amino
                                    ## acid sequences of a human gene and an
                                    ## E.coli O157 gene
png <- keggGet("hsa05130", "image") ## retrieves the image file of a
                                    ## pathway map
t <- tempfile()
library(png)
writePNG(png, t)
res <- keggGet("hsa05130", "kgml")
str(res)
```

---

keggInfo *Displays the current statistics of a given database*

Description

Displays statistics of a given database, such as number of entries, version, release date, and source.

Usage

```
keggInfo(database)
```

Arguments

database	Either a KEGG database (list available via `listDatabases()`), a KEGG organism code (list available by calling `keggList()`) with the `organism` argument), or a T number (list available by calling `keggList()` with the `genome` argument.)

Value

A character vector containing statistics about `database`.

Author(s)

Dan Tenenbaum

References

https://www.kegg.jp/kegg/docs/keggapi.html

Examples

```
res <- keggInfo("kegg") ## displays the current statistics of the KEGG database
cat(res)
res <- keggInfo("pathway") ## displays the number pathway entries including both
                    ## the reference and organism-specific pathways
cat(res)
res <- keggInfo("hsa") ## displays the number of gene entries for the
                    ## KEGG organism Homo sapiens
cat(res)
```

`keggLink` *Find related entries by using database cross-references.*

Description

Find related entries by using database cross-references.

Usage

```
keggLink(target, source)
```

Arguments

`target`	Either the name of a single KEGG database (list available via `listDatabases()`, a "T number" genome identifier, or a KEGG organism code (lists of both available via `keggList("organism")`).
`source`	The same as `target`, but may also be one or more KEGG identifiers.

Details

Many of the old KEGGSOAP functions whose names started with 'get', such as `get.pathways.by.genes` and `get.pathways.by.reactions`, are replaced by using `keggLink` (see examples).

Value

A named character vector.

Author(s)

Dan Tenenbaum

References

https://www.kegg.jp/kegg/docs/keggapi.html

Examples

```
res <- keggLink("pathway", "hsa") ## KEGG pathways linked from each of
          ## the human genes equivalent to 'get.genes.by.pathway' in KEGGSOAP
length(res)
head(res)
res <- keggLink("hsa", "pathway") ## human genes linked from each of the
          ## KEGG pathways equivalent to 'get.pathways.by.genes' in KEGGSOAP
keggLink("pathway", c("hsa:10458", "ece:Z5100")) ## KEGG pathways
          ## linked from a human gene and an E. coli O157 gene
res <- keggLink("hsa:126") ## LinkDB search shows all KEGG
          ## resources related to hsa:126
head(res)
```

`keggList` *Returns a list of entry identifiers and associated definition for a given database or a given set of database entries.*

Description

Returns a list of entry identifiers and associated definition for a given database or a given set of database entries.

Usage

```
keggList(database, organism)
```

Arguments

`database`	Either a KEGG database (list available via `listDatabases()`), a KEGG organism code (list available via `keggList()` with the `organism` argument, a T number (list available via `keggList()` with the `genome` argument), or a character vector of KEGG identifiers.
`organism`	`Optional`. A KEGG organism identifier (list available via `keggList()` with the `organism` argument).

Value

A named character vector containing entry identifiers and associated definition.

Author(s)

Dan Tenenbaum

References

https://www.kegg.jp/kegg/docs/keggapi.html

Examples

```
res <- keggList("pathway") ## returns the list of reference pathways
length(res)
head(res)
res <- keggList("pathway", "hsa") ## returns the list of human pathways
length(res)
head(res)
res <- keggList("organism") ## returns the list of KEGG organisms with
                    ## taxonomic classification
nrow(res)
head(res)
res <- keggList("hsa")  ## returns the entire list of human genes
length(res)
head(res)
## keggList("T01001") ## same as above
keggList(c("hsa:10458", "ece:Z5100")) ## returns the list of a human gene
                                      ## and an E.coli O157 gene
keggList(c("cpd:C01290","gl:G00092")) ## returns the list of a compound entry
                                      ## and a glycan entry
keggList(c("C01290+G00092")) ## same as above (prefixes are not necessary)
```

`listDatabases` *Lists the KEGG databases which may be searched.*

Description

Lists the KEGG databases which may be searched. In most cases, you can also use a KEGG organism name or T number (genome identifier) as a database name.

Usage

```
listDatabases()
```

Value

A character vector of database names.

Author(s)

Dan Tenenbaum

References

https://www.kegg.jp/kegg/docs/keggapi.html

See Also

keggList

Examples

```
listDatabases()
res <- keggList("organism") ## list all organisms
nrow(res)
head(res)
res <- keggList("hsa") ## list all human genes
length(res)
head(res)
## keggList("T01001") ## list all human genes
res <- keggList("genome") ## list all genome identifiers
length(res)
head(res)
```

---

mark.pathway.by.objects

Client-side interface to obtain an url for a KEGG pathway diagram with a given set of genes marked

---

Description

Given a KEGG pathway id and a set of KEGG gene ids, the functions return the URL of a KEGG pathway diagram with the elements corresponding to the genes marked by red or specified color

Usage

```
mark.pathway.by.objects(pathway.id, object.id.list)
color.pathway.by.objects(pathway.id, object.id.list,
                                    fg.color.list, bg.color.list)
```

Arguments

`pathway.id`	`pathway.id` a character string for a KEGG pathway id. KEGG pathway ids consist of the string path followed by a colon, a three-letter code for the organism of concern, and then a number (e. g. "path:eco00020"). The three-letter organism code consists of the first letter of the genus name and the first two letters of the species name of the scientific name of the organism of concern
`object.id.list`	`object.id.list` a vector of character strings for KEGG gene ids. KEGG gene ids normally consist of three letters followed by a column and then several numeric numbers. The three letters are from the first letter of the genus name and the first two letters of the species name of the scientific name of the organism of concern (e. g. hsa:111 for Homo Sapiens)
`fg.color.list`	`fg.color.list` a vector of two character strings to indicate the color for the text and border, respectively, of the objects in a pathway diagram. The strings can either be a color code linke #ff0000 or letter link yellow
`bg.color.list`	`bg.color.list` a vector of character strings of the same length of `object.id.list` to indicate the background color of the objects in a pathway diagram. The strings can either be a color code like #ff0000 or letter like yellow

Details

This function only returns the URL of the KEGG pathway diagram. Use the function browseURL to view the diagram.

These functions are not part of the KEGG REST API; they are provided because they existed in KEGGSOAP and an alternative implementation was possible.

Value

This function returns a character string for the url

Author(s)

Jianhua Zhang

References

https://www.kegg.jp/kegg/docs/keggapi.html

See Also

browseURL

Examples

```
 url <- mark.pathway.by.objects(
    "path:eco00260", c("eco:b0002", "eco:c00263")
)
if(interactive()){
    browseURL(url)
}
url <- color.pathway.by.objects(
    "path:eco00260", c("eco:b0002", "eco:c00263"),
    c("#ff0000", "#00ff00"),
    c("#ffff00", "yellow")
)
```

Index