Package ‘STRINGdb’

May 6, 2024

Type Package

Title STRINGdb - Protein-Protein Interaction Networks and Functional Enrichment Analysis

Version 2.16.0

Date 2023-03-05

Author Andrea Franceschini <andrea.franceschini@isb-sib.ch>

Maintainer Damian Szklarczyk <damian.szklarczyk@imls.uzh.ch>

Description The STRINGdb package provides a R interface to the STRING protein-protein interactions database (https://string-db.org).

License GPL-2

Depends R (>= 2.14.0)

Imports png, sqldf, plyr, igraph, httr, methods, RColorBrewer, gplots, hash, plotrix

Suggests RUnit, BiocGenerics

biocViews Network

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

git_branch RELEASE_3_19

git_last_commit 22e5c74

git_last_commit_date 2024-04-30

Repository Bioconductor 3.19

Date/Publication 2024-05-05

Contents

<table>
<thead>
<tr>
<th>Function</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>add_diff_exp_color</td>
<td>2</td>
</tr>
<tr>
<td>add_proteins_description</td>
<td>3</td>
</tr>
<tr>
<td>coeffOfvar</td>
<td>4</td>
</tr>
<tr>
<td>delColDf</td>
<td>4</td>
</tr>
<tr>
<td>diff_exp_example1</td>
<td>5</td>
</tr>
<tr>
<td>downloadAbsentFile</td>
<td>5</td>
</tr>
</tbody>
</table>
Description

Take in input a dataframe containing a logFC column that reports the logarithm of the difference in expression level. Add a "color" column to the data frame such that strongly downregulated genes are colored in green and strong upregulated genes are in red. When the down or up-regulation is instead weak the intensity of the color gets weaker as well, accordingly.

Usage

```r
## S4 method for signature 'STRINGdb'
add_diff_exp_color(screen, logFcColStr="logFC")
```
### `add_proteins_description`

**Description**

Add description columns to the proteins that are present in the data frame given in input. The data frame must contain a column named "STRING_id".

**Usage**

```r
## S4 method for signature 'STRINGdb'
add_proteins_description(screen)
```

**Arguments**

- `screen` Dataframe containing the results of the experiment (e.g. the analyzed results of a microarray or RNAseq experiment)
- `logFcColStr` name of the column that contains the logFC of the expression

**Value**

vector containing the colors

**Author(s)**

Andrea Franceschini
### coeffOfvar

**Description**

coefficient of variation

**Usage**

coeffOfvar(x)

**Arguments**

- `x`: input number

**Details**

coefficient of variation

**Value**

coefficient of variation

**Author(s)**

Andrea Franceschini

---

### delColDf

**Description**

delete a column in the data frame

**Usage**

delColDf(df, colName)

**Arguments**

- `df`: data frame
- `colName`: name of the column to be deleted

**Value**

data frame
**diff_exp_example1**

**Author(s)**

Andrea Franceschini

---

**diff_exp_example1**  
example of microarray data (data processed from GEO GSE9008)

---

**Description**

example of microarray data (data processed from GEO GSE9008)

**Usage**

data(diff_exp_example1)

**Format**

Data frames with 20861 observations on the following 3 variables.

- **gene**: a character vector
- **pvalue**: a numeric vector
- **logFC**: a numeric vector

**Source**


---

**downloadAbsentFile**  
downloadAbsentFile

---

**Description**

download a file only if it is not present.

**Usage**

downloadAbsentFile(urlStr, oD = tempdir())

**Arguments**

- **urlStr**: url from which to download the file
- **oD**: directory where to store the file

**Author(s)**

Andrea Franceschini
**Description**

download a STRING file only if it is not present or if it is corrupted.

**Usage**

```r
downloadAbsentFileSTRING(urlStr, oD = tempdir())
```

**Arguments**

- `urlStr`: url from which to download the file
- `oD`: directory where to store the file

**Author(s)**

Andrea Franceschini

---

**Description**

Loads and returns the STRING alias table.

**Usage**

```r
## S4 method for signature 'STRINGdb'
get_aliases()
```

**Value**

a data frame containing the STRING alias table

**Author(s)**

Andrea Franceschini
get_annotations

Description
Loads and returns STRING annotations (i.e. GO annotations, KEGG pathways, domain databases).
The annotations are stored in the “annotations” variable.

Usage
```r
## S4 method for signature 'STRINGdb'
get_annotations()
```

Value
a data frame containing the annotations to the STRING proteins (e.g. GeneOntology, KEGG pathways, InterPro domains)

Author(s)
Andrea Franceschini

get_annotations_desc

Description
Returns a data frame with the description of every STRING annotation term (it downloads and caches the information the first time that is called).

Usage
```r
## S4 method for signature 'STRINGdb'
get_annotaions_desc()
```

Value
data frame with the description of every STRING annotation term.

Author(s)
Andrea Franceschini
get_bioc_graph

Description

Returns the interaction graph as an object of the graph package in Bioconductor.

Usage

```r
## S4 method for signature 'STRINGdb'
get_bioc_graph()
```

Value

interaction graph as an object of the graph package in Bioconductor.

Author(s)

Andrea Franceschini

get_clusters

Description

Returns a list of clusters of interacting proteins. See the iGraph (http://igraph.sourceforge.net/) documentation for additional information on the algorithms.

Usage

```r
## S4 method for signature 'STRINGdb'
get_clusters(string_ids, algorithm="fastgreedy")
```

Arguments

- `string_ids`: a vector of STRING identifiers.
- `algorithm`: algorithm to use for the clustering. You can choose between "fastgreedy", "walktrap", "spinglass" and "edge.betweenness").

Value

list of clusters of interacting proteins.

Author(s)

Andrea Franceschini
get_enrichment

Description

Returns the enrichment in pathways of the vector of STRING proteins that is given in input.

Usage

```r
## S4 method for signature 'STRINGdb'
get_enrichment(string_ids, category = "Process", methodMT = "fdr", iea = TRUE, minScore=NULL)
```

Arguments

- `string_ids`: a vector of STRING identifiers.
- `category`: category for which to compute the enrichment (i.e. "Process", "Component", "Function", "KEGG", "Pfam", "InterPro"). The default category is "Process".
- `methodMT`: method to be used for the multiple testing correction. (i.e. "fdr", "bonferroni"). The default is "fdr".
- `iea`: specify whether you also want to use electronic inference annotations
- `minScore`: with Tissue and Disease categories is possible to filter the annotations having an annotation score higher than this threshold (from 0 to 5)

Value

Data frame containing the enrichment in pathways of the vector of STRING proteins that is given in input.

Author(s)

Andrea Franceschini

get_graph

Description

Return an igraph object with the STRING network (for information about igraph visit http://igraph.sourceforge.net)

Usage

```r
## S4 method for signature 'STRINGdb'
get_graph()
```
Description

Returns the list of closest homologs (as measured by bitscore) of the given input identifiers in all STRING species or single target species.

Usage

```r
## S4 method for signature 'STRINGdb'
get_homologs_besthits(string_ids, target_species_id=NULL)
```

Arguments

- `string_ids` 
  a vector of STRING identifiers.
- `target_species_id` 
  NCBI taxonomy identifier of the species to query for homologs (the species must be present in the STRING database)

Value

Data frame containing the best blast hits x species of the given input identifiers.

Author(s)

Andrea Franceschini
**get_interactions**

Description
- Shows the interactions in between the proteins that are given in input.

Usage

```r
## S4 method for signature 'STRINGdb'
get_interactions(string_ids)
```

Arguments
- `string_ids`: a vector of STRING identifiers

Value
- Data frame containing the interactions in between the input proteins.

Author(s)
- Andrea Franceschini

**get_link**

Description
- Returns a short link to the network page of our STRING website that shows the protein interactions between the given identifiers.

Usage

```r
## S4 method for signature 'STRINGdb'
get_link(string_ids, required_score=NULL, network_flavor="evidence", payload_id = NULL)
```

Arguments
- `string_ids`: a vector of STRING identifiers.
- `required_score`: minimum STRING combined score of the interactions (if left NULL we get the combined score of the object, which is 400 by default).
- `network_flavor`: specify the flavor of the network ("evidence", "confidence" or "actions". default "evidence").
- `payload_id`: an identifier of payload data on the STRING server (see method post_payload for additional informations)
Value

short link to the network page of our STRING website that shows the protein interactions between the input identifiers.

Author(s)

Andrea Franceschini

---

get_neighbors
get_neighbors

Description

Get the neighborhoods of a protein (or of a vector of proteins) that is given in input.

Usage

```r
## S4 method for signature 'STRINGdb'
get_neighbors(string_ids)
```

Arguments

- `string_ids`: a vector of STRING identifiers

Value

vector containing the neighborhoods of a protein (or of a vector of proteins) that is given in input.

Author(s)

Andrea Franceschini

---

get_paralogs
get_paralogs

Description

Returns the list of paralogs of the given input in their species.

Usage

```r
## S4 method for signature 'STRINGdb'
get_paralogs(string_ids)
```

Arguments

- `string_ids`: a vector of STRING identifiers.
get_png

Value

Data frame containing the best blast hits x species of the given input identifiers.

Author(s)

Andrea Franceschini

Description

Returns a png image of a STRING protein network with the given identifiers.

Usage

```r
## S4 method for signature 'STRINGdb'
get_png(string_ids, required_score=NULL, network_flavor="evidence", file=NULL, payload_id=NULL)
```

Arguments

- `string_ids`: a vector of STRING identifiers.
- `required_score`: minimum STRING combined score of the interactions (if left NULL we get the combined score of the object, which is 400 by default).
- `network_flavor`: specify the flavor of the network ("evidence", "confidence" or "actions". default "evidence").
- `file`: file where to save the image
- `payload_id`: identifier of the payload

Value

Returns a png image of a STRING protein network with the given identifiers.

Author(s)

Andrea Franceschini
get_ppi_enrichment

Description

Returns a pvalue representing the enrichment in interactions of the list of proteins (i.e. the probability to obtain such a number of interactions by chance).

Usage

```r
## S4 method for signature 'STRINGdb'
get_ppi_enrichment(string_ids)
```

Arguments

- `string_ids`: a vector of STRING identifiers

Value

Returns a pvalue representing the enrichment in interactions of the list of proteins (i.e. the probability to obtain such a number of interactions by chance).

Author(s)

Andrea Franceschini

get_proteins

Description

Returns the STRING proteins data frame. (it downloads and caches the information the first time that is called).

Usage

```r
## S4 method for signature 'STRINGdb'
get_proteins()
```

Value

STRING proteins data frame.

Author(s)

Andrea Franceschini
get_subnetwork

Description
Returns the subgraph generated by the given input proteins.

Usage
```r
## S4 method for signature 'STRINGdb'
get_subnetwork(string_ids)
```

Arguments
- `string_ids`: a vector of STRING identifiers

Value
Returns the subgraph (i.e. an iGraph object) generated by the given input proteins.

Author(s)
Andrea Franceschini

get_summary

Description
Returns a summary of the STRING sub-network containing the identifiers provided in input.

Usage
```r
## S4 method for signature 'STRINGdb'
get_summary(string_ids)
```

Arguments
- `string_ids`: a vector of STRING identifiers

Value
Returns a summary (i.e. a text description) of the STRING sub-network containing the identifiers provided in input.

Author(s)
Andrea Franceschini
### get_term_proteins

**Description**

Returns the proteins annotated to belong to a given term.

**Usage**

```r
## S4 method for signature 'STRINGdb'
get_term_proteins(term_ids, string_ids=NULL, enableIEA=TRUE)
```

**Arguments**

- `term_ids`: vector of terms
- `string_ids`: a vector of STRING identifiers. If the variable is set, the method returns only the proteins that are present in this vector.
- `enableIEA`: whether to consider also Electronic Inferred Annotations

**Value**

Returns the proteins annotated to belong to a given term.

**Author(s)**

Andrea Franceschini

### interactions_example

**Example of a protein-protein interactions sorted data frame**

**Description**

Example of a sorted list of protein-protein interactions, resulting from our cooccurrence algorithm (SVD_Phy)

**Usage**

```r
data(interactions_example)
```

**Format**

Data frames with 20861 observations on the following 3 variables.

- **proteinA**: a character vector
- **proteinB**: a character vector
- **score**: a numeric vector
Description

Downloads and returns the STRING network (the network is set also in the graph variable of the STRING_db object).

It makes use of the variables: "backgroundV" vector containing STRING identifiers to be used as background (i.e. the STRING network loaded will contain only the proteins that are present also in this vector) "score_threshold" STRING combined score threshold (the network loaded contains only interactions having a combined score greater than this threshold)

Usage

```r
## S4 method for signature 'STRINGdb'
load()
```

Value

STRING network (i.e. an iGraph object. For info look to http://igraph.sourceforge.net)

Author(s)

Andrea Franceschini

Description

Force download and loading of all the files (so that you can later store the object on the hard disk if you like). It makes use of the variables: "backgroundV" vector containing STRING identifiers to be used as background (i.e. the STRING network loaded will contain only the proteins that are present also in this vector) "score_threshold" STRING combined score threshold (the network loaded contains only interactions having a combined score greater than this threshold)

Usage

```r
## S4 method for signature 'STRINGdb'
load_all()
```

Author(s)

Andrea Franceschini
Description

Maps the gene identifiers of the input dataframe to STRING identifiers. It returns the input dataframe with the "STRING_id" additional column.

Usage

```r
# S4 method for signature 'STRINGdb'
map(my_data_frame, my_data_frame_id_col_names, takeFirst=TRUE, removeUnmappedRows=FALSE, quiet=FALSE)
```

Arguments

- **my_data_frame**: data frame provided as input.
- **my_data_frame_id_col_names**: vector containing the names of the columns of "my_data_frame" that have to be used for the mapping.
- **takeFirst**: boolean indicating what to do in case of multiple STRING proteins that map to the same name. If TRUE, only the first of those is taken. Otherwise all of them are used. (default TRUE)
- **removeUnmappedRows**: remove the rows that cannot be mapped to STRING (by default those lines are left and their STRING_id is set to NA).
- **quiet**: Setting this variable to TRUE we can avoid printing the warning relative to the unmapped values.

Value

Returns the dataframe that is given in input with the "STRING_id" additional column.

Author(s)

Andrea Franceschini

Description

Maps the gene identifiers of the input vector to STRING identifiers (using a take first approach). It returns a vector with the STRING identifiers of the mapped proteins.
## S4 method for signature 'STRINGdb'

*mp(proteinAliases)*

### Arguments

*protein_aliases*

vector of protein aliases that we want to convert to STRING identifiers

### Value

It returns a vector with the STRING identifiers of the mapped proteins.

### Author(s)

Andrea Franceschini

---

### multi_map_df

#### Description

mapping function (it add the possibility to map using more than one column of the data frame)

#### Usage

*multi_map_df(dfToMap, dfMap, strColsFrom, strColFromDfMap, strColToDfMap, caseSensitive=FALSE)*

#### Arguments

*dfToMap*

input data frame (that contains the columns that need to be mapped)

*dfMap*

data frame containing the mapping data

*strColsFrom*

sorted vector containing the names of the columns to be used in the input data frame for the mapping (the order of the elements in the vector defines the priority for the mapping)

*strColFromDfMap*

name of the column in the mapping data frame to be used as source for the mapping

*strColToDfMap*

name of the column in the mapping data frame to be used as target for the mapping

*caseSensitive*

specify whether the mapping should be case sensitive

#### Value

data frame with an additional column containing the result of the mapping
Author(s)

Andrea Franceschini

---

plot_network

---

Description

Plots an image of the STRING network with the given proteins.

Usage

```r
## S4 method for signature 'STRINGdb'
plot_network(string_ids, payload_id=NULL, required_score=NULL, add_link=TRUE, add_summary=TRUE)
```

Arguments

- `string_ids`: a vector of STRING identifiers
- `payload_id`: an identifier of payload data on the STRING server (see method `post_payload` for additional informations)
- `required_score`: a threshold on the score that overrides the default `score_threshold`, that we use only for the picture
- `add_link`: parameter to specify whether you want to generate and add a short link to the relative page in STRING. As default this option is active but we suggest to de-activate it in case one is generating many images (e.g. in a loop). Deactivating this option avoids to generate and store a lot of short-urls on our server.
- `add_summary`: parameter to specify whether you want to add a summary text to the picture. This summary includes a p-value and the number of proteins/interactions.

Author(s)

Andrea Franceschini

---

post_payload

---

Description

Posts the input to STRING and returns an identifier that you can use to access the payload when you enter in our website.

Usage

```r
## S4 method for signature 'STRINGdb'
post_payload(stringIds, colors=NULL, comments=NULL, links=NULL, iframe_urls=NULL, logo_imgF=NULL, legend_imgF=NULL)
```
Arguments

stringIds  vector of STRING identifiers.

colors vector containing the colors to use for a every STRING identifier (the order of the elements must match those in the string_ids vector).

comments vector containing the comments to use for every STRING identifier (the order of the elements must match those in the string_ids vector).

links vector containing the links to use for every STRING identifier (the order of the elements must match those in the string_ids vector).

iframe_urls vector containing the urls of the iframes to use for every STRING identifier (the order of the elements must match those in the string_ids vector).

logo_imgF path to a file containing the logo image to be display in the STRING website.

legend_imgF path to a file containing a legend image to be display in the STRING website.

Value

identifier of the payload.

Author(s)

Andrea Franceschini

Description

With this method it is possible to remove the interactions that are composed by a pair of homologous/similar proteins, having a similarity bitscore between each other higher than a threshold.

Usage

## S4 method for signature 'STRINGdb'
remove_homologous_interactions(interactions_dataframe, bitscore_threshold = 60)

Arguments

interactions_dataframe

a data frame containing the sorted interactions to be benchmarked. The data frame should have the following column names: proteinA, proteinB, score.

bitscore_threshold

filter out pairs of homologous proteins, having a similarity bitscore higher than this parameter.
**renameColDf**

**Description**

Rename a column of a data frame

**Usage**

```r
renameColDf(df, colOldName, colNewName)
```

**Arguments**

- `df`: input data frame
- `colOldName`: column name to be changed
- `colNewName`: new column name

**Value**

data frame with the column name changed

**Author(s)**

Andrea Franceschini

---

**set_background**

**Description**

With this method you can specify a vector of proteins to be used as background. The network is reloaded and only the proteins that are present in the background vector are inserted in the graph. Besides, the background is taken in consideration for all the enrichment statistics.

**Usage**

```r
## S4 method for signature 'STRINGdb'
set_background(background_vector)
```
Arguments

background_vector

vector of STRING protein identifiers

Author(s)

Andrea Franceschini

Description

The R package STRINGdb provides a convenient interface to the STRING protein-protein interactions database for the R/bioconductor users. Please look at the manual/vignette to get additional information and examples on how to use the package. STRING is a database of known and predicted protein-protein interactions. It contains information from numerous sources, including experimental repositories, computational prediction methods and public text collections. Each interaction is associated with a combined confidence score that integrates the various evidences. STRING is regularly updated, the latest version 9.05 contains information on 5 millions proteins from more than 1100 species. The STRING web interface is freely accessible at: http://string-db.org/

Extends

All reference classes extend and inherit methods from "envRefClass".

Fields

annotations: Object of class data.frame
annotations_description: Object of class data.frame
graph: Object of class igraph
proteins: Object of class data.frame
speciesList: Object of class data.frame
species: Object of class numeric
version: Object of class character
input_directory: Object of class character
backgroundV: Object of class vector
score_threshold: Object of class numeric
Methods

set_background(background_vector): ~~
post_payload(stringIds, colors, comments, links, iframe_urls, logo_imgF, legend_imgF):
  ~~
plot_network(string_ids, payload_id, required_score): ~~
plot_ppi_enrichment(string_ids, file, sliceWindow, edgeWindow, windowExtendedReferenceThreshold, minVal):
  ~~
map(my_data_frame, my_data_frame_id_col_names, takeFirst, removeUnmappedRows, quiet):
  ~~
load(): ~~
get_term_proteins(term_ids, string_ids, enableIEA): ~~
get_summary(string_ids): ~~
get_subnetwork(string_ids): ~~
get_ppi_enrichment_full(string_ids, sliceWindow, edgeWindow, windowExtendedReferenceThreshold, growing):
  ~~
get_ppi_enrichment(string_ids): ~~
get_proteins(): ~~
get_png(string_ids, required_score, network_flavor, file, payload_id): ~~
get_neighbors(string_ids): ~~
get_link(string_ids, required_score, network_flavor, payload_id): ~~
get_interactions(string_ids): ~~
get_homologs_besthits(string_ids, symbets, target_species_id, bitscore_threshold):
  ~~
get_homologs(string_ids, target_species_id, bitscore_threshold): ~~
get_graph(): ~~
get_enrichment(string_ids, category, methodMT, iea): ~~
get_clusters(string_ids, algorithm): ~~
get_annotations_desc(): ~~
get_annotations(): ~~
load_all(): ~~
initialize(...): ~~
add_proteins_description(screen): ~~
add_diff_exp_color(screen, logFcColStr): ~~
show(): ~~

Author(s)

Andrea Franceschini
References


See Also

http://stitch-db.org

Examples

showClass("STRINGdb")
Index

* classes
  STRINGdb-class, 23
* datasets
  add_differential_expression, 5
  add_interactions, 16
  add_diff_exp_color, 2
  add_proteins_description, 3
  coeff_of_var, 4
  delColDf, 4
  diff_exp_example1, 5
  download_absent_file, 5
  download_absent_file_STRING, 6
  envRefClass, 23
  get_aliases, 6
  get_annotations, 7
  get_annotations_desc, 7
  get_bioc_graph, 8
  get_clusters, 8
  get_enrichment, 9
  get_graph, 9
  get_homologs_besthits, 10
  get_interactions, 11
  get_link, 11
  get_neighbors, 12
  get_paralogs, 12
  get_png, 13
  get_ppi_enrichment, 14
  get_proteins, 14
  get_subnetwork, 15
  get_summary, 15
  get_term_proteins, 16
  interactions_example, 16
load, 17
load_all, 17

map, 18
mp, 18
multi_map_df, 19
plot_network, 20
post_payload, 20
remove_homologous_interactions, 21
renameColDf, 22
set_background, 22
STRINGdb (STRINGdb-class), 23
STRINGdb-class, 23