Package: STRINGdb (via r-universe)

November 4, 2026

Type Package

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

Version 2.24.3

Date 2026-11-04

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

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

Description The STRINGdb package provides an R interface to STRING, a protein-protein interaction database and functional enrichment analysis tool (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

Config/pak/sysreqs libglpk-dev libpng-dev libxml2-dev libssl-dev

Repository Bioconductor 3.23

Date/Publication 2026-11-04 13:32:02 UTC

RemoteUrl https://github.com/bioc/STRINGdb

RemoteRef RELEASE_3_23

RemoteSha a43eee52e7737ed496801543435145f91767314e

Contents

add_diff_exp_color 2
add_proteins_description 3
coeffOfvar 4
delColDf 4
diff_exp_example1 5

downloadAbsentFile 5
downloadAbsentFileSTRING 6
get_aliases 6
get_annotations 7
get_annotations_desc 7
get_bioc_graph 8
get_clusters 8
get_enrichment 9
get_enrichment_figure 10
get_geneset_description 11
get_graph 12
get_homologs_besthits 12
get_interaction_partners 13
get_interactions 14
get_link 14
get_neighbors 16
get_paralogs 16
get_png 17
get_ppi_enrichment 19
get_proteins 19
get_subnetwork 20
get_summary 20
get_term_proteins 21
interactions_example 22
load 22
load_all 23
map 23
mp 24
multi_map_df 25
plot_network 25
post_payload 26
remove_homologous_interactions 27
renameColDf 28
set_background 28
STRINGdb-class 29

Index **32**

`add_diff_exp_color` *add_diff_exp_color*

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

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

Arguments

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

Value

vector containing the colors

Author(s)

Andrea Franceschini

---

add_proteins_description
add_proteins_description

Description

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

Usage

```
## 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)

Value

returns the same dataframe given in input with an additional columns containing a description of the proteins.

Author(s)

Andrea Franceschini

`coeffOfvar` *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` *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

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

Whyte L, Huang YY, Torres K, Mehta RG. Molecular mechanisms of resveratrol action in lung cancer cells using dual protein and microarray analyses. Cancer Res 2007.

`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

`downloadAbsentFileSTRING` *downloadAbsentFileSTRING*

Description

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

Usage

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

Arguments

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

Author(s)

Andrea Franceschini

`get_aliases` *get_aliases*

Description

Loads and returns STRING aliases. Depending on `takeFirst`, this returns either all alias mappings or a single preferred mapping for ambiguous aliases.

Usage

```
## S4 method for signature 'STRINGdb'
get_aliases(takeFirst=TRUE, usePreferredSources=TRUE)
```

Arguments

`takeFirst`	boolean indicating whether ambiguous aliases should be collapsed to a single STRING identifier. If `FALSE`, all alias mappings are returned. If `TRUE`, one mapping is returned for each ambiguous alias.
`usePreferredSources`	boolean indicating whether preferred alias sources should be used to disambiguate aliases when `takeFirst=TRUE`. This parameter is ignored when `takeFirst=FALSE`.

Value

a data frame containing STRING aliases. With `takeFirst=FALSE`, all alias mappings are returned. With `takeFirst=TRUE`, ambiguous aliases are collapsed to one STRING identifier.

Author(s)

Andrea Franceschini

`get_annotations` *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

```
## 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` *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

```
## S4 method for signature 'STRINGdb'
get_annotations_desc()
```

Value

data frame with the description of every STRING annotation term.

Author(s)

Andrea Franceschini

get_bioc_graph *get_bioc_graph*

Description

Returns the interaction graph as an object of the graph package in Bioconductor. For `network_type="regulatory"`, the returned `graphNEL` object preserves edge direction.

Usage

```
## 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 *get_clusters*

Description

Returns a list of clusters of interacting proteins. Clustering is not available for the directed `network_type="regulatory"` network. See the iGraph (http://igraph.sourceforge.net/) documentation for additional information on the algorithms.

Usage

```
## 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 *get_enrichment*

Description

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

Usage

```
## S4 method for signature 'STRINGdb'
get_enrichment(string_ids, category = "All", methodMT = NULL, iea = NULL, minScore = NULL, max_fdr = NUL
```

Arguments

string_ids	a vector of STRING identifiers.
category	category for which to return the enrichment (i.e. "Process", "Component", "Function", "KEGG", "Pfam", "InterPro"). The default "All" returns all categories.
methodMT	deprecated: only the FDR correction is available. A value is ignored, with a warning.
iea	deprecated: a value is ignored, with a warning.
minScore	deprecated: a value is ignored, with a warning.
max_fdr	optional filter: the highest FDR of a term, above 0 and at most 1.
min_gene_count	optional filter: the lowest number of input proteins annotated with a term, a whole number of at least 1.
min_strength	optional filter: the lowest strength of a term, log10(observed / expected).
min_signal	optional filter: the lowest signal of a term, STRING's balance of strength and FDR.
merge_similarity	optional: merges the terms whose protein sets overlap at least this much (above 0, at most 1; 1 merges identical sets only) into the most significant of them.

Details

STRING applies the optional filters like the filters of the enrichment table on its website. Without them (the default NULL), no term is filtered. The returned table has no strength and signal columns.

The filters need STRING version 12.5 or later. An older version, such as version="12.0", would ignore them, so they are not used there, with a warning.

Value

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

Author(s)

Andrea Franceschini

get_enrichment_figure *get_enrichment_figure*

Description

Returns a STRING enrichment figure for the given identifiers and enrichment category.

Usage

```
## S4 method for signature 'STRINGdb'
get_enrichment_figure(string_ids, category = "Process", file = NULL, output_format = "image", group_by_
```

Arguments

string_ids	a vector of STRING identifiers.
category	enrichment category to visualize. Supported categories are "Process", "Function", "Component", "Keyword", "KEGG", "RCTM", "HPO", "MPO", "DPO", "WPO", "ZPO", "FYPO", "Pfam", "SMART", "InterPro", "PMID", "NetworkNeighborAL", "COMPARTMENTS", "TISSUES", "DISEASES", and "WikiPathways".
file	file where to save the image output.
output_format	STRING image output format: "image", "highres_image" or "svg".
group_by_similarity	threshold for visually grouping related terms on the plot, from 0.1 to 1.
color_palette	color palette used to represent FDR values. Supported values are "mint_blue", "lime_emerald", "green_blue", "peach_purple", "straw_navy", and "yellow_pink".
number_of_term_shown	maximum number of terms displayed on the plot. STRING shows at most 100.
x_axis	variable used on the x-axis and, without sort_by, for ranking the terms: "signal", "strength", "FDR", or "gene_count".
caller_identity	caller identifier sent to STRING.
graph_type	the kind of plot: "dotplot" (default) or "barplot".
sort_by	variable that decides which terms are shown and in which order: "signal", "strength", "FDR", or "gene_count". The default NULL uses the x_axis variable.
max_fdr	optional filter: the highest FDR of a term, above 0 and at most 1.
min_gene_count	optional filter: the lowest number of input proteins annotated with a term, a whole number of at least 1.
min_strength	optional filter: the lowest strength of a term, log10(observed / expected).
min_signal	optional filter: the lowest signal of a term, STRING's balance of strength and FDR.
merge_similarity	optional: merges the terms whose protein sets overlap at least this much (above 0, at most 1; 1 merges identical sets only) into the most significant of them. Unlike group_by_similarity, which only groups related terms on the plot, merged terms are not shown.

Details

STRING applies the optional filters like the filters of the enrichment table on its website; without them (the default `NULL`), no term is filtered. When STRING refuses the request, the method stops with STRING's explanation.

`graph_type`, `sort_by` and the filters need STRING version 12.5 or later. An older version, such as `version="12.0"`, would ignore them, so they are not used there, with a warning.

Value

For `output_format="image"` and `output_format="highres_image"`, returns a PNG image array. For `output_format="svg"`, returns the SVG markup as text.

Author(s)

Damian Szklarczyk

`get_geneset_description` *get_geneset_description*

Description

Returns concise primary, secondary, and tertiary descriptions of a protein set using the STRING gene-set description API.

Usage

```
## S4 method for signature 'STRINGdb'
get_geneset_description(string_ids, caller_identity="STRINGdb-package")
```

Arguments

`string_ids`	a vector of STRING identifiers.
`caller_identity`	caller identifier sent to STRING.

Value

A one-row data frame containing primary, secondary, and tertiary descriptions when they are available.

Author(s)

Damian Szklarczyk

get_graph *get_graph*

Description

Return an igraph object with the STRING network (for information about iGraph visit http://igraph.sourceforge.net). For `network_type="regulatory"`, the graph is directed from regulator to target.

Usage

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

Value

igraph object with the STRING network

Author(s)

Andrea Franceschini

References

Csardi G, Nepusz T: The igraph software package for complex network research, InterJournal, Complex Systems 1695. 2006. http://igraph.sf.net

See Also

In order to simplify the most common tasks, we do also provide convenient functions that wrap some iGraph functions. get_interactions(string_ids) # returns the interactions in between the input proteins get_neighbors(string_ids) # Get the neighborhoods of a protein (or of a vector of proteins) that is given in input. get_subnetwork(string_ids) # returns a subgraph from the given input proteins

get_homologs_besthits *get_homologs_besthits*

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

```
## 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_interaction_partners
get_interaction_partners

Description

Returns the interaction partners of the input proteins using the locally loaded STRING graph. The returned data frame preserves the edge attributes available for the current `link_data` setting. For a regulatory network, `from` and `to` preserve the regulator-to-target direction.

Usage

```
## S4 method for signature 'STRINGdb'
get_interaction_partners(string_ids, required_score=NULL, limit=NULL)
```

Arguments

string_ids	a vector of STRING identifiers
required_score	optional minimum combined score for returned partners. This value cannot be below the `score_threshold` used to load the local graph.
limit	optional maximum number of partner rows returned per query protein

Value

Data frame containing one row per query-partner interaction. The leading columns are `from`, `to`, `combined_score`, `from_name` and `to_name`, followed by any additional edge attributes loaded for the current `link_data` mode. For functional and physical networks, `from` is the queried protein and `to` is its interaction partner. For regulatory networks, they are the original regulator and target. If `limit` is used, rows are ordered by `combined_score`.

Author(s)

Damian Szklarczyk

get_interactions *get_interactions*

Description

Shows the interactions in between the proteins that are given in input. For a regulatory network, `from` is the regulator and `to` is the regulated protein.

Usage

```
## 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. The leading columns are `from`, `to`, `combined_score`, `from_name` and `to_name`, followed by any additional loaded edge attributes.

Author(s)

Damian Szklarczyk

get_link *get_link*

Description

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

Usage

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

Arguments

string_ids	a vector of STRING identifiers. Can be omitted when network_term_id is provided.
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 "typed". The typed view is available only when network_type="full" (or its "functional" alias). "actions" is accepted as a compatibility alias for "typed".
payload_id	an identifier of payload data on the STRING server (see method post_payload for additional informations).
network_term_id	functional term identifier used by STRING instead of explicit protein identifiers.
hide_node_labels	hides all protein names from the picture. Accepts TRUE/FALSE or 0/1.
hide_disconnected_nodes	hides proteins that are not connected to any other protein in the network. Accepts TRUE/FALSE or 0/1.
block_structure_pics_in_bubbles	disables structure pictures inside the bubbles. Accepts TRUE/FALSE or 0/1.
flat_node_design	disables 3D bubble design. Accepts TRUE/FALSE or 0/1. Default is TRUE.
center_node_labels	centers protein names on nodes. Accepts TRUE/FALSE or 0/1.
custom_label_font_size	changes the font size of the protein names (from 5 to 50).
caller_identity	caller identifier sent to STRING.
typed_physical_edges	include physical edges in the typed network view. Available only with network_flavor="typed". Accepts TRUE/FALSE or 0/1.
typed_regulatory_edges	include regulatory edges in the typed network view. Available only with network_flavor="typed". Accepts TRUE/FALSE or 0/1.
show_regulatory_signs	show positive/negative regulatory signs in the typed network view. Requires typed_regulatory_edges=TRUE. Accepts TRUE/FALSE or 0/1.

Value

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

Author(s)

Damian Szklarczyk

get_neighbors *get_neighbors*

Description

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

Usage

```
## S4 method for signature 'STRINGdb'
get_neighbors(string_ids, mode="all")
```

Arguments

string_ids	a vector of STRING identifiers
mode	direction to traverse in a regulatory network: "all" (default), "out" for regulated proteins, or "in" for regulators.

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

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

Arguments

string_ids	a vector of STRING identifiers.

Value

Data frame containing the within-species homology hits of the input identifiers, with columns ncbiTaxonId_A, stringId_A, ncbiTaxonId_B, stringId_B, and bitscore.

Author(s)

Andrea Franceschini

---

get_png *get_png*

Description

Returns a STRING network image for the given identifiers or for a STRING functional term.

Usage

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

Arguments

string_ids	a vector of STRING identifiers. Can be omitted when network_term_id is provided.
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 "typed". The typed view is available only when network_type="full" (or its "functional" alias). "actions" is accepted as a compatibility alias for "typed".
file	file where to save the image output.
payload_id	identifier of the payload.
output_format	STRING image output format: "image", "highres_image" or "svg".
network_term_id	functional term identifier used by STRING instead of explicit protein identifiers.
hide_node_labels	hides all protein names from the picture. Accepts TRUE/FALSE or 0/1.
hide_disconnected_nodes	hides proteins that are not connected to any other protein in the network. Accepts TRUE/FALSE or 0/1.
block_structure_pics_in_bubbles	disables structure pictures inside the bubbles. Accepts TRUE/FALSE or 0/1.
flat_node_design	disables 3D bubble design. Accepts TRUE/FALSE or 0/1. Default is TRUE.
center_node_labels	centers protein names on nodes. Accepts TRUE/FALSE or 0/1.
custom_label_font_size	changes the font size of the protein names (from 5 to 50).
caller_identity	caller identifier sent to STRING.

typed_physical_edges
: include physical edges in the typed network view. Available only with network_flavor="typed". Accepts TRUE/FALSE or 0/1.

typed_regulatory_edges
: include regulatory edges in the typed network view. Available only with network_flavor="typed". Accepts TRUE/FALSE or 0/1.

show_regulatory_signs
: show positive/negative regulatory signs in the typed network view. Requires typed_regulatory_edges=TRUE. Accepts TRUE/FALSE or 0/1.

network_clustering_algorithm
: clusters the network and colours the proteins by cluster, as the Clusters tab of the STRING website: "kmeans", "mcl" or "leiden" (in any case). The default NULL draws the network without clusters; the other clustering options need an algorithm.

network_clustering_parameter
: the parameter of the clustering algorithm: for k-means the number of clusters (a whole number of at least 1, default 3), for MCL the inflation (from 1.1 to 10, default 3), and for Leiden the resolution (from 0.1 to 10, default 1).

cluster_aware_layout
: lays the network out by its clusters. Accepts TRUE/FALSE or 0/1.

network_clustering_edge_visibility
: how the edges between clusters are drawn: "faded" (STRING's default), "dotted", "solid" or "noshow".

show_cluster_names
: writes the names of the clusters on the picture. Accepts TRUE/FALSE or 0/1; by default the names are not shown.

cluster_name_size
: the size of the cluster names: 1, 1.5, 2, 2.5, 3, 3.5, 4, 5 or 6. By default STRING chooses the size that suits the picture.

Details

A network of more than 1,000 proteins is drawn simplified, as on the STRING website: with confidence edges (whatever network_flavor says) and without structure pictures in the nodes.

STRING draws a PNG (output_format="image" or "highres_image") of at most 150,000 interactions. For a larger network, and for any other request that STRING refuses, the method stops with STRING's explanation; for the PNG, ask for the SVG (output_format="svg") or use a higher required_score.

The clustering options need STRING version 12.5 or later. An older version, such as version="12.0", would ignore them, so they are not used there, with a warning.

Value

For output_format="image" and output_format="highres_image", returns a PNG image array. For output_format="svg", returns the SVG markup as text.

Author(s)

Damian Szklarczyk

`get_ppi_enrichment` *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). The calculation uses the object's selected functional, physical, or regulatory network type.

Usage

```
## 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` *get_proteins*

Description

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

Usage

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

Value

STRING proteins data frame.

Author(s)

Andrea Franceschini

`get_subnetwork` *get_subnetwork*

Description

Returns the subgraph generated by the given input proteins. For a regulatory network, regulator-to-target edge directions are preserved.

Usage

```
## 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` *get_summary*

Description

Returns a summary of the STRING sub-network containing the identifiers provided in input. The summary and its interaction-enrichment p-value use the object's selected network type.

Usage

```
## 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` *get_term_proteins*

Description

Returns the proteins annotated to belong to a given term.

Usage

```
## 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, resulta our cooccurrence algorithm (SVD_Phy)

Usage

```
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

load *load*

Description

Downloads and returns the STRING network (the network is set also in the graph variable of the STRING_db object). When possible, the download uses the threshold-specific streamed network file matching the current `score_threshold`. For `network_type="regulatory"`, the returned `igraph` object is directed: an edge from `protein1` to `protein2` represents regulation of the latter by the former.

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

```
## 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 and Damian Szklarczyk

load_all *load_all*

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

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

Author(s)

Andrea Franceschini

map *map*

Description

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

Usage

```
## 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 contatining 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.

usePreferredSources when takeFirst=TRUE, prioritize aliases using the preferred source order.

Value

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

Author(s)

Andrea Franceschini

---

mp *mp*

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.

Usage

```
## S4 method for signature 'STRINGdb'
mp(protein_aliases)
```

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 *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 *plot_network*

Description

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

Usage

```
## S4 method for signature 'STRINGdb'
plot_network(string_ids, payload_id=NULL, required_score=NULL, add_link=FALSE, network_flavor="evider
```

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. The default is FALSE; disabling this option avoids generating and storing many short URLs when plotting in a loop.
network_flavor	style of network edges: "evidence", "confidence", or "typed". The typed view is available only with network_type="full" (or "functional"). "actions" is accepted as a compatibility alias for "typed".
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.
typed_physical_edges	include physical edges in the typed network view. Available only with network_flavor="typed". Accepts TRUE/FALSE or 0/1.
typed_regulatory_edges	include regulatory edges in the typed network view. Available only with network_flavor="typed". Accepts TRUE/FALSE or 0/1.
show_regulatory_signs	show positive/negative regulatory signs in the typed network view. Requires typed_regulatory_edges=TRUE. Accepts TRUE/FALSE or 0/1.

Details

A network of more than 1,000 proteins is drawn simplified, as on the STRING website: with confidence edges (whatever network_flavor says) and without structure pictures in the nodes.

The picture is a PNG, which STRING draws for at most 150,000 interactions. For a larger network the method stops with STRING's explanation: use a higher required_score, or get_png with output_format="svg".

Author(s)

Andrea Franceschini

---

post_payload *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

```
## S4 method for signature 'STRINGdb'
post_payload(stringIds, colors=NULL, comments=NULL, links=NULL, iframe_urls=NULL, logo_imgF=NULL, leg
```

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

---

remove_homologous_interactions

remove_homologous_interactions

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 contaning 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

Value

interactions data frame where the homologous pairs have been removed, from the input interactions' data frame

Author(s)

Andrea Franceschini

`renameColDf` *renameColDf*

Description

Rename a column of a data frame

Usage

```
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` *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. If you already created a `STRINGdb` object, calling `set_background` is sufficient and you do not need to instantiate a new object again.

Usage

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

Arguments

`background_vector`
: vector of STRING protein identifiers

Author(s)

Andrea Franceschini

STRINGdb-class *Class* "STRINGdb"

Description

The R package STRINGdb provides a convenient interface to STRING, a protein-protein interaction database and functional enrichment analysis tool, for R/Bioconductor users. Please look at the manual/vignette for 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 different evidence channels. STRING v12.5 contains information on 59.3 million proteins from 12,535 organisms and more than 28 billion interactions. It provides functional, physical, and directed regulatory networks. The STRING web interface is freely accessible at: https://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`. The STRING version, e.g. "12.5"; a version without a dot gets ".0" ("11" means "11.0"). For the versions known to the package (11.0, 11.0b, 11.5, 12.0 and 12.5) the object is created without a request to STRING, so it can also be created offline. Without a version, STRING is asked for its current version; when STRING cannot be reached within a few seconds, the newest known version is used and a warning is printed. Any other version is looked up on STRING, and creating the object stops with an error when that version is not available or STRING cannot be reached. ~~

network_type: Object of class character. One of "full", "physical", or "regulatory". The regulatory network requires STRING v12.5 or later. ~~

link_data: Object of class character. One of "combined_only", "detailed", or "full". ~~

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, minVa ~~

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, growin ~~

get_ppi_enrichment(string_ids): ~~

get_proteins(): ~~

get_enrichment_figure(string_ids, category="Process", file, output_format="image", group_by_similarit ~~

get_png(string_ids, required_score, network_flavor, file, payload_id, output_format, network_term_id, h ~~

get_neighbors(string_ids, mode="all"): ~~

get_interaction_partners(string_ids, required_score, limit): ~~

get_link(string_ids, required_score, network_flavor, payload_id, network_term_id, hide_node_labels, hi ~~

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, minScore, max_fdr, min_gene_count, min_strength, mi ~~

get_geneset_description(string_ids, caller_identity="STRINGdb-package"): ~~

`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 and Damian Szklarczyk

References

Szklarczyk D, Nastou K, Koutrouli M, Kirsch R, Mehryary F, Hachilif R, Hu D, Peluso ME, Huang Q, Fang T, Doncheva NT, Pyysalo S, Bork P, Jensen LJ, von Mering C. The STRING database in 2025: protein networks with directionality of regulation. Nucleic Acids Res. 2025 Jan 6;53(D1):D730-D737. doi: 10.1093/nar/gkae1113.

Examples

```
showClass("STRINGdb")
```

Index

∗ **classes**
    STRINGdb-class, 29
∗ **datasets**
    diff_exp_example1, 5
    interactions_example, 22

add_diff_exp_color, 2
add_proteins_description, 3

coeffOfvar, 4

delColDf, 4
diff_exp_example1, 5
downloadAbsentFile, 5
downloadAbsentFileSTRING, 6

envRefClass, *29*

get_aliases, 6
get_annotations, 7
get_annotations_desc, 7
get_bioc_graph, 8
get_clusters, 8
get_enrichment, 9
get_enrichment_figure, 10
get_geneset_description, 11
get_graph, 12
get_homologs_besthits, 12
get_interaction_partners, 13
get_interactions, 14
get_link, 14
get_neighbors, 16
get_paralogs, 16
get_png, 17
get_ppi_enrichment, 19
get_proteins, 19
get_subnetwork, 20
get_summary, 20
get_term_proteins, 21

interactions_example, 22

load, 22
load_all, 23

map, 23
mp, 24
multi_map_df, 25

plot_network, 25
post_payload, 26

remove_homologous_interactions, 27
renameColDf, 28

set_background, 28
STRINGdb *(*STRINGdb-class*)*, 29
STRINGdb-class, 29