

# Package ‘STRINGdb’

October 12, 2016

**Type** Package

**Title** STRINGdb (Search Tool for the Retrieval of Interacting proteins database)

**Version** 1.12.0

**Date** 2016-03-31

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

**Maintainer** Alexander Roth <alexander.roth@isb-sib.ch>

**Description** The STRINGdb package provides a R interface to the STRING protein-protein interactions database (<http://www.string-db.org>).

**License** GPL-2

**Depends** R (>= 2.14.0)

**Imports** png, sqldf, plyr, igraph, RCurl, methods, RColorBrewer, gplots, hash, plotrix

**Suggests** RUnit, BiocGenerics

**biocViews** Network

**NeedsCompilation** no

## R topics documented:

add_diff_exp_color . . . . .	3
add_proteins_description . . . . .	3
benchmark_ppi . . . . .	4
benchmark_ppi_pathway_view . . . . .	5
coeffOfvar . . . . .	6
delColDf . . . . .	6
diff_exp_example1 . . . . .	7
downloadAbsentFile . . . . .	7
downloadAbsentFileSTRING . . . . .	8
get_aliases . . . . .	8
get_annotations . . . . .	9
get_annotations_desc . . . . .	9
get_bioc_graph . . . . .	10

get_clusters . . . . .	10
get_enrichment . . . . .	11
get_graph . . . . .	11
get_homologs . . . . .	12
get_homologs_besthits . . . . .	13
get_interactions . . . . .	13
get_link . . . . .	14
get_neighbors . . . . .	15
get_png . . . . .	15
get_ppi_enrichment . . . . .	16
get_ppi_enrichment_full . . . . .	16
get_proteins . . . . .	17
get_pubmed . . . . .	18
get_pubmed_interaction . . . . .	18
get_STRING_species . . . . .	19
get_subnetwork . . . . .	19
get_summary . . . . .	20
get_term_proteins . . . . .	20
interactions_example . . . . .	21
load . . . . .	22
load_all . . . . .	22
map . . . . .	23
mp . . . . .	23
multi_map_df . . . . .	24
plot_network . . . . .	25
plot_ppi_enrichment . . . . .	25
plot_ppi_enrichment_graph . . . . .	26
post_payload . . . . .	27
ppie.compLambda . . . . .	28
ppie.compLambdaL1L2 . . . . .	29
ppie.compPij . . . . .	29
ppie.getNumEdgesBetween . . . . .	30
ppi_enrichment . . . . .	31
ppi_enrichment_full . . . . .	31
remove_homologous_interactions . . . . .	33
renameColDf . . . . .	33
set_background . . . . .	34
STRINGdb-class . . . . .	35
<b>Index</b>	<b>37</b>

---

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 column 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 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

```
## 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

---

benchmark_ppi	<i>benchmark_ppi</i>
---------------	----------------------

---

**Description**

benchmark a list of protein-protein interactions using pathways (e.g. KEGG). The function outputs a table where the interactions are mapped to KEGG and the number of TPs and FPs are counted.

**Usage**

```
## S4 method for signature 'STRINGdb'
benchmark_ppi(interactions_dataframe, pathwayType = "KEGG", max_homology_bitscore = 60, precision_wi
```

**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

pathwayType      category to use to benchmark the interactions (default KEGG)

max\_homology\_bitscore  
filter out pairs of homologous proteins, having a similarity bitscore higher than this parameter

precision\_window  
define the size of the precision window (i.e. the window used to scan the sorted interactions data frame in order to compute the benchmark precision). At the beginning and at the end of the sorted interactions data frame, the window is automatically enlarged (at the beginning) and reduced (at the end)

exclude\_pathways  
Exclude the terms that should not be used for benchmarking. If this parameter is set to "blacklist", a black list to be used maintained by our group is automatically downloaded from our servers (otherwise it is possible to specify a vector with the terms that have to be excluded).

**Value**

interactions data frame where the interactions are mapped to KEGG and the number of TPs and FPs are counted.

**Author(s)**

Andrea Franceschini

---

benchmark\_ppi\_pathway\_view  
*benchmark\_ppi\_pathway\_view*

---

**Description**

Takes in input the results of the benchmark\_ppi function, and constructs a new table that provides a view at the pathway level (i.e. it list all the pathways to which the interactions belong)

**Usage**

```
## S4 method for signature 'STRINGdb'
benchmark_ppi_pathway_view(benchmark_ppi_data_frame, precision_threshold=0.2, pathwayType = "KEGG")
```

**Arguments**

benchmark\_ppi\_data\_frame  
 data frame that comes out from the benchmark\_ppi function.

precision\_threshold  
 threshold that specify where to stop taking considering the interactions in the sorted input data frame (the list is scanned until the precision goes below this value).

pathwayType  
 the pathway category to use (KEGG by default)

**Value**

data frame containing the pathways (i.e. terms) of the input proteins. Several parameters are reported: pathway coverage (i.e. number of interactions in the list belonging to the pathway / maximum number of interactions between the pathay proteins (i.e.  $\text{proteins} * (\text{proteins}-1)/2$  ) ) total\_representation (i.e. number of interactions in the list belonging to the pathway / size of the input data frame )

**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**

<code>urlStr</code>	url from which to download the file
<code>oD</code>	directory where to store the file

**Author(s)**

Andrea Franceschini

---

`get_aliases`*get\_aliases*

---

**Description**

Loads and returns the STRING alias table.

**Usage**

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

**Value**

a data frame containing the STRING alias table

**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	<i>get_bioc_graph</i>
----------------	-----------------------

---

**Description**

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

**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	<i>get_clusters</i>
--------------	---------------------

---

**Description**

Returns a list of clusters of interacting proteins. 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", "walk-trap", "spinglass" and "edge.betweenness").

**Value**

list of clusters of interacting proteins.

**Author(s)**

Andrea Franceschini

---

get_enrichment	<i>get_enrichment</i>
----------------	-----------------------

---

**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 = "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	<i>get_graph</i>
-----------	------------------

---

**Description**

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

**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

*get\_homologs*

---

**Description**

Returns the homologs of the given input identifiers that are present in the given `target_species_id`.

**Usage**

```
## S4 method for signature 'STRINGdb'  
get_homologs(string_ids, target_species_id, bitscore_threshold=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)  
bitscore_threshold  
                threshold on the bitscore of the blast alignment.
```

**Value**

Data frame containing the homologs of the given input identifiers and that are present in the given `target_species_id`.

**Author(s)**

Andrea Franceschini

---

get\_homologs\_besthits *get\_homologs\_besthits*

---

**Description**

Returns the best blast hits x species of the given input identifiers.

**Usage**

```
## S4 method for signature 'STRINGdb'  
get_homologs_besthits(string_ids, symbets = FALSE, target_species_id = NULL, bitscore_threshold=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)  
`bitscore_threshold` threshold on the bitscore of the blast alignment.  
`symbets` specify whether you want only symmetrical best hits

**Value**

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

**Author(s)**

Andrea Franceschini

---

get\_interactions *get\_interactions*

---

**Description**

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

**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.

**Author(s)**

Andrea Franceschini

---

<i>get_link</i>	<i>get_link</i>
-----------------	-----------------

---

**Description**

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

**Usage**

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

**Arguments**

<code>string_ids</code>	a vector of STRING identifiers.
<code>required_score</code>	minimum STRING combined score of the interactions (if left NULL we get the combined score of the object, which is 400 by default).
<code>network_flavor</code>	specify the flavor of the network ("evidence", "confidence" or "actions". default "evidence").
<code>payload_id</code>	an identifier of payload data on the STRING server (see method <code>post_payload</code> 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	<i>get_neighbors</i>
---------------	----------------------

---

**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)
```

**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_png	<i>get_png</i>
---------	----------------

---

**Description**

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

**Usage**

```
## 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`      *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**

```
## 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_ppi_enrichment_full`  
*get\_ppi\_enrichment\_full*

---

**Description**

Returns a vector showing the enrichment in protein interactions in various positions of the list of genes in input. In practice, a list of 3 vectors is returned: 1) enrichment (i.e. enrichment computed in the window from 1 to x) 2) enrichmentWindow (i.e. enrichment computed in a sliding window of size determined by the "edgeWindow" parameters and the sliding steps determined by the "sliceWindow" parameter) 3) enrichmentWindowExtended (i.e. like the enrichmentWindow, but it also includes an initial window of size "windowExtendedReferenceThreshold" with respect to which to compute the enrichment )



**Usage**

```
## S4 method for signature 'STRINGdb'
get_ppi_enrichment_full(string_ids, sliceWindow = 20, edgeWindow = 140, windowExtendedReferenceThres
```

**Arguments**

`string_ids` a vector of STRING identifiers

`sliceWindow` defines the interval in proteins after which to compute the enrichment, scanning the list (i.e. the resolution)

`edgeWindow` size of the window that we use to compute the enrichment (i.e. the window pvalue is computed using the proteins inside this "edgeWindow")

`windowExtendedReferenceThreshold` defines the size of a window at the beginning of the list. The enrichment will be computed always including the proteins in this window

`growingWindowLimit` threshold where to stop the computation of the enrichment

**Value**

Returns a vector showing the enrichment in protein interactions in various positions of the list of genes in input.

**Author(s)**

Andrea Franceschini

---

get_proteins	<i>get_proteins</i>
--------------	---------------------

---

**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

---

<code>get_pubmed</code>	<i>get_pubmed</i>
-------------------------	-------------------

---

**Description**

Returns vector with the PUBMED IDs of the publications that contain the names of the proteins in the input vector.

**Usage**

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

**Arguments**

`string_ids` a vector of STRING identifiers

**Value**

vector with the PUBMED IDs of the publications that contain the names of the proteins in the input vector.

**Author(s)**

Andrea Franceschini

---

<code>get_pubmed_interaction</code>	<i>get_pubmed_interaction</i>
-------------------------------------	-------------------------------

---

**Description**

Returns vector with the PUBMED IDs of the publications that contain the names of both the input proteins.

**Usage**

```
## S4 method for signature 'STRINGdb'  
get_pubmed_interaction(STRING_id_a, STRING_id_b )
```

**Arguments**

`STRING_id_a` STRING identifier  
`STRING_id_b` STRING identifier

**Value**

vector with the PUBMED IDs of the publications that contain the names of both the input proteins.

**Author(s)**

Andrea Franceschini

---

*get\_STRING\_species*      *get\_STRING\_species*

---

**Description**

Returns a data frame with the species (i.e. organisms) that are present in STRING.

**Usage**

```
get_STRING_species(version=NULL, species_name=NULL)
```

**Arguments**

version            STRING version  
species\_name      name of the species that you are searching

**Value**

data frame with the species (i.e. organisms) that are present in STRING.

**Author(s)**

Andrea Franceschini

---

*get\_subnetwork*            *get\_subnetwork*

---

**Description**

Returns the subgraph generated by the given input proteins.

**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

---

<code>get_summary</code>	<code>get_summary</code>
--------------------------	--------------------------

---

**Description**

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

**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

---

<code>get_term_proteins</code>	<code>get_term_proteins</code>
--------------------------------	--------------------------------

---

**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, results 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	<i>load</i>
------	-------------

---

**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**

```
## 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

---

load_all	<i>load_all</i>
----------	-----------------

---

**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	<i>map</i>
-----	------------

---

### 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

<code>my_data_frame</code>	data frame provided as input.
<code>my_data_frame_id_col_names</code>	vector containing the names of the columns of "my_data_frame" that have to be used for the mapping.
<code>takeFirst</code>	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)
<code>removeUnmappedRows</code>	remove the rows that cannot be mapped to STRING (by default those lines are left and their STRING_id is set to NA).
<code>quiet</code>	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

---

mp	<i>mp</i>
----	-----------

---

### 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	<i>multi_map_df</i>
--------------	---------------------

---

**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=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 deactivate 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

---

plot\_ppi\_enrichment                      *plot\_ppi\_enrichment*

---

**Description**

Plots a graph showing the enrichment in protein interactions in various positions of the list of genes in input.

**Usage**

```
## S4 method for signature 'STRINGdb'
plot_ppi_enrichment(string_ids, file=NULL, sliceWindow = 20, edgeWindow = 140,
                    windowExtendedReferenceThreshold = 260, minVal=0.000000001, title="",
```

**Arguments**

<code>string_ids</code>	a vector of STRING identifiers
<code>file</code>	file where to save the graph as an image
<code>sliceWindow</code>	defines the interval in proteins after which to compute the enrichment, scanning the list (i.e. the resolution)
<code>edgeWindow</code>	size of the window that we use to compute the enrichment (i.e. the window pvalue is computed using the proteins inside this "edgeWindow")
<code>windowExtendedReferenceThreshold</code>	defines the size of a window at the beginning of the list. The enrichment will be computed always including the proteins in this window
<code>title</code>	title of the graph.
<code>minVal</code>	minimum value that the pvalue can assume in the log-scale graph. If the p-value is lower, we convert the value to this minimum value
<code>quiet</code>	if set to TRUE the method runs in quiet mode (turning off any output message)

**Author(s)**

Andrea Franceschini

---

`plot_ppi_enrichment_graph`  
*plot\_ppi\_enrichment\_graph*

---

**Description**

Plots a graph showing the enrichment in protein interactions in various positions of the list of genes in input.

**Usage**

```
plot_ppi_enrichment_graph(proteins, ppi_network, file, sliceWindow, edgeWindow, windowExtendedRefere
```

**Arguments**

<code>proteins</code>	a vector of protein identifiers
<code>ppi_network</code>	an igraph object containing the protein-protein interactions' graph.
<code>file</code>	file where to save the graph as an image

sliceWindow	defines the interval in proteins after which to compute the enrichment, scanning the list (i.e. the resolution)
edgeWindow	size of the window that we use to compute the enrichment (i.e. the window pvalue is computed using the proteins inside this "edgeWindow")
windowExtendedReferenceThreshold	defines the size of a window at the beginning of the list. The enrichment will be computed always including the proteins in this window
title	title of the graph.
minVal	minimum value that the pvalue can assume in the log-scale graph. If the p-value is lower, we convert the value to this minimum value
quiet	if set to TRUE the method runs in quiet mode (turning off any output message)

**Author(s)**

Andrea Franceschini

**References**

Pradines JR, Farutin V, Rowley S, Dancik V. Analyzing protein lists with large networks: edge-count probabilities in random graphs with given expected degrees. *J. Comput. Biol.* 2005;12:113-128. Franceschini, A et al. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. In: *Nucleic Acids Res.* 2013 Jan;41(Database issue)

---

post_payload	<i>post_payload</i>
--------------	---------------------

---

**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)

<code>iframe_urls</code>	vector containing the urls of the iframes to use for every STRING identifier ( the order of the elements must match those in the <code>string_ids</code> vector).
<code>logo_imgF</code>	path to a file containing the logo image to be display in the STRING website
<code>legend_imgF</code>	path to a file containing a legend image to be display in the STRING website

**Value**

identifier of the payload.

**Author(s)**

Andrea Franceschini

---

`ppie.compLambda`      *ppie.compLambda*

---

**Description**

compute the number of expected interactions between a set of proteins.

**Usage**

```
ppie.compLambda(degrees, edgeNum)
```

**Arguments**

<code>degrees</code>	vector containing the degrees of the nodes of a set of proteins
<code>edgeNum</code>	total number of edges of the entire graph

**Value**

number of expected interactions

**Author(s)**

Andrea Franceschini

**References**

Pradines JR, Farutin V, Rowley S, Dancik V. Analyzing protein lists with large networks: edge-count probabilities in random graphs with given expected degrees. *J. Comput. Biol.* 2005;12:113-128. Franceschini, A et al. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. In: *Nucleic Acids Res.* 2013 Jan;41(Database issue)

---

*ppie.compLambdaL1L2*     *ppie.compLambdaL1L2*

---

**Description**

Compute lambda L1L2

**Usage**

`ppie.compLambdaL1L2(degreesI, degreesJ, edgeNum)`

**Arguments**

<code>degreesI</code>	vector containing the degrees of the nodes present in nodeSet I
<code>degreesJ</code>	vector containing the degrees of the nodes present in nodeSet J
<code>edgeNum</code>	total number of edges of the entire graph

**Details**

compute the number of expected interactions between two sets of nodes

**Author(s)**

Andrea Franceschini

---

*ppie.compPij*     *ppie.compPij*

---

**Description**

Compute the pvalue of protein I to interact with protein J (look at the Pradines paper referenced below)

**Usage**

`ppie.compPij(degI, degJ, edgeNum)`

**Arguments**

<code>degI</code>	degree of protein I
<code>degJ</code>	degree of protein J
<code>edgeNum</code>	total number of edges of the entire graph

**Value**

return the pvalue of protein I to interact with protein J (look at the Pradines paper referenced below)

**Author(s)**

Andrea Franceschini

**References**

Pradines JR, Farutin V, Rowley S, Dancik V. Analyzing protein lists with large networks: edge-count probabilities in random graphs with given expected degrees. *J. Comput. Biol.* 2005;12:113-128. Franceschini, A et al. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. In: *Nucleic Acids Res.* 2013 Jan;41(Database issue)

---

`ppie.getNumEdgesBetween`

*ppie.getNumEdgesBetween*

---

**Description**

find the number of interactions between two sets of nodes

**Usage**

```
ppie.getNumEdgesBetween(graph, nodesFrom, nodesTo)
```

**Arguments**

graph	igraph object
nodesFrom	list of nodes
nodesTo	list of nodes

**Value**

return the number of interactions between two sets of nodes

**Author(s)**

Andrea Franceschini

---

ppi\_enrichment      *ppi\_enrichment*

---

**Description**

Computes the enrichment in protein-protein interactions.

**Usage**

```
ppi_enrichment(hitList, ppi_network)
```

**Arguments**

hitList	sorted list of proteins (from the most significant to the least significant)
ppi_network	an igraph object containing the graph of the protein-protein interaction's network.

**Value**

enrichment	pvalue that describes the probability to get such a number of interactions by chance
lambda	expected number of interactions

**Author(s)**

Andrea Franceschini

**References**

Pradines JR, Farutin V, Rowley S, Dancik V. Analyzing protein lists with large networks: edge-count probabilities in random graphs with given expected degrees. *J. Comput. Biol.* 2005;12:113-128. Franceschini, A et al. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. In: *Nucleic Acids Res.* 2013 Jan;41(Database issue)

---

ppi\_enrichment\_full      *ppi\_enrichment\_full*

---

**Description**

Compute the enrichment in protein-protein interactions of a sorted list of proteins. The computation is repeated at different positions in the list.

**Usage**

```
ppi_enrichment_full(hitList, ppi_network, sliceWindow, edgeWindow, windowExtendedReferenceThreshold,
```

**Arguments**

hitList	sorted list of proteins (from the most significant to the least significant)
ppi_network	an igraph object containing the protein-protein interactions' graph.
sliceWindow	defines the interval in proteins after which to compute the enrichment, scanning the list (i.e. the resolution)
edgeWindow	size of the window that we use to compute the enrichment (i.e. the window pvalue is computed using the proteins inside this "edgeWindow")
windowExtendedReferenceThreshold	when we compute the "windowExtended" pvalue we are computing the pvalue that considers the following interactions: 1) the interactions inside the edgeWindow (as we do with the edgeWindow pvalue) 2) the interactions that connects the proteins in the edgeWindow with the proteins in another window at the beginning of the list (i.e. the windowExtendedReference). windowExtendedReferenceThreshold defines the size of this windowExtendedReference window. In this way we can compute, in a reliable way, the enrichment of a sorted list of proteins, in various positions of the list.
growingWindowLimit	stop to compute the enrichment (from position 1 to position n) after growingWindowLimit proteins in the sorted list. (this limit speeds up the computation of the 2 other types of enrichment)
quiet	if set to TRUE the method runs in quiet mode (turning off any output message)

**Value**

enrichment	vector containing the enrichments in protein-protein interactions of the input list of genes (the length of the vector is $\text{length}(\text{hitList})/\text{sliceWindow}$ ).
enrichmentWindow	vector containing the enrichments in protein-protein interactions of the input list of genes (the length of the vector is $\text{length}(\text{hitList})/\text{sliceWindow}$ ). The enrichment is computed considering only the proteins inside the sliding window
enrichmentWindowExtended	vector containing the enrichments in protein-protein interactions of the input list of genes (the length of the vector is $\text{length}(\text{hitList})/\text{sliceWindow}$ ). Look at the description of the windowExtendedReferenceThreshold variable

**Author(s)**

Andrea Franceschini

**References**

Pradines JR, Farutin V, Rowley S, Dancik V. Analyzing protein lists with large networks: edge-count probabilities in random graphs with given expected degrees. *J. Comput. Biol.* 2005;12:113-128. Franceschini, A et al. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. In: *Nucleic Acids Res.* 2013 Jan;41(Database issue)



---

```
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 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
```

### 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	<i>set_background</i>
----------------	-----------------------

---

**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**

```
## 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 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, mi):  
 ~~  
 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, grow
~~
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**

Franceschini, A (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. In: 'Nucleic Acids Res. 2013 Jan;41(Database issue):D808-15. doi: 10.1093/nar/gks1094. Epub 2012 Nov 29'.

**See Also**

<http://stitch-db.org>

**Examples**

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

# Index

## \*Topic **classes**

STRINGdb-class, [35](#)

## \*Topic **datasets**

diff\_exp\_example1, [7](#)

interactions\_example, [21](#)

add\_diff\_exp\_color, [3](#)

add\_proteins\_description, [3](#)

benchmark\_ppi, [4](#)

benchmark\_ppi\_pathway\_view, [5](#)

coeff0fvar, [6](#)

delColDf, [6](#)

diff\_exp\_example1, [7](#)

downloadAbsentFile, [7](#)

downloadAbsentFileSTRING, [8](#)

envRefClass, [35](#)

get\_aliases, [8](#)

get\_annotations, [9](#)

get\_annotations\_desc, [9](#)

get\_bioc\_graph, [10](#)

get\_clusters, [10](#)

get\_enrichment, [11](#)

get\_graph, [11](#)

get\_homologs, [12](#)

get\_homologs\_besthits, [13](#)

get\_interactions, [13](#)

get\_link, [14](#)

get\_neighbors, [15](#)

get\_png, [15](#)

get\_ppi\_enrichment, [16](#)

get\_ppi\_enrichment\_full, [16](#)

get\_proteins, [17](#)

get\_pubmed, [18](#)

get\_pubmed\_interaction, [18](#)

get\_STRING\_species, [19](#)

get\_subnetwork, [19](#)

get\_summary, [20](#)

get\_term\_proteins, [20](#)

interactions\_example, [21](#)

load, [22](#)

load\_all, [22](#)

map, [23](#)

mp, [23](#)

multi\_map\_df, [24](#)

plot\_network, [25](#)

plot\_ppi\_enrichment, [25](#)

plot\_ppi\_enrichment\_graph, [26](#)

post\_payload, [27](#)

ppi\_enrichment, [31](#)

ppi\_enrichment\_full, [31](#)

ppie.compLambda, [28](#)

ppie.compLambdaL1L2, [29](#)

ppie.compPij, [29](#)

ppie.getNumEdgesBetween, [30](#)

remove\_homologous\_interactions, [33](#)

renameColDf, [33](#)

set\_background, [34](#)

STRINGdb (STRINGdb-class), [35](#)

STRINGdb-class, [35](#)